Challenges in Information Representation with Augmented Reality for Procedural Task Support

Dissertation

in Partial Fulfillment of the Requirements for the Degree of

Doctor rerum naturalium

by
Tobias Müller

submitted to:
Universität Paderborn
Fakultät für Elektrotechnik, Informatik und Mathematik

Paderborn, January 2023

Abstract

Supporting manually executed procedural tasks, such as assembly, repair or maintenance, is a very promising area for the application of augmented reality. One reason is its ability to display information spatially integrated into the workspace such that it is available directly at the required location in conjunction with the tools and work pieces. However, this approach is not automatically beneficial, but can also be problematic. Unnecessary sensory and cognitive effort may be a consequence, i.e., effort that is not caused by the actual task but by the use of an augmented reality system. Possible reasons may be, for example, that relevant parts of the physical world are obscured by virtual graphics, or that virtual objects are positioned in such a way that they are difficult to see. For augmented reality to reach its full potential despite such issues, it is necessary to develop a better understanding of how it can be used to appropriately represent information such as instructions.

As one approach to deepening this understanding, the Information Representation Modeling with Augmented Reality (IRMAR) framework and, based on it, the Representation Challenges are developed in this thesis. Potential sources of unnecessary sensory and cognitive effort caused by the use of augmented reality to support procedural tasks are identified and possible measures to mitigate these are proposed.

For this purpose, in a first step, an approach for modeling information presented with augmented reality is developed. As its central element, the concept of information objects is introduced. These are physical or virtual objects that convey information to a user relevant for the completion of one or more work steps in a procedural task. Examples are a physical screw that has to be loosened, an animation that shows the necessary movements, or a text with relevant instructions. A fundamental difference to pre-existing modeling approaches, e.g., for classical user interfaces but also for augmented reality, is the inclusion of the connection that information objects have with the physical world and the information that a user can gain from it.

Based on this, in the second step, characteristics of information objects are identified, which are either valid for a single object or a combination of several objects. Examples are that information objects always are in spatial relationship to each other due to their placement in the physical world (spatial relationship) or that there are often semantic connections between information objects that convey relevant information to a user (connectedness). To be able to use them as a basis for the following steps, only such characteristics are considered, which apply either to all or at least a majority of the information objects.

In the third step, the sources of unnecessary sensory or cognitive effort are systematically derived by combining the characteristics. For example, the combination of the two previously mentioned characteristics spatial relationship and connectedness results in such a source. This one is the potential difficulty for a user to recognize the semantic connections of information objects. The reason is that due to the spatial representation inherent to augmented reality, it can easily happen that the individual connections are not appropriately or unambiguously visible. To facilitate practical use, these sources are formulated as representation challenges that need to be solved when designing representation of information with augmented reality. Overall, the six challenges clarity, consistency, visibility, orientation, motion cue and information linking are identified. For each of them, it is described what constitutes the challenge, how it can be derived from the characteristics, if and where it has been previously described in literature and which measures exist that help solving it. Their descriptions are given in a way that supports designers of augmented reality systems to design information representation in such a way that it generates little unnecessary sensory and cognitive effort for a user.

In a further step, the information linking challenge, which is that semantic connections between an information object and a place or another object must be easily recognizable, is examined more closely. In order to structure this topic area, a taxonomy is created that describes which fundamental possibilities exist for visualizing information linking in augmented reality. By classifying visualizations described in the existing literature into its dimensions, it becomes apparent that there is a lack of measures for this challenge. Therefore, two empirical studies with human subjects are conducted. In these it is compared, how different types of visual connections impact human task performance. The results show a clear order of preference concerning the type, leading to the derivation of new measures for solving the information linking challenge.

As a last step, the application of IRMAR, including the challenges and the measures to solve them, is shown for two use cases. The first is the integration into the ISO norm 9241-210 Human-Centered Design. It is illustrated with the design of an example augmented reality system that supports the removal of a vehicle airbag. The second application is the design of the Augmented Reality Procedural Task Modeling Language (ARPML) that allows authoring augmented reality support for procedural tasks. Again, the airbag removal task is used as an example to demonstrate the use of ARPML.

Zusammenfassung

Ein vielversprechender Anwendungsbereich für Augmented Reality ist die Unterstützung von manuell durchgeführten prozeduralen Aufgaben, wie z.B. Montage-, Reparatur- oder Wartungsarbeiten. Ein Grund dafür ist die Möglichkeit, Informationen räumlich in den Arbeitsbereich integriert anzuzeigen, so dass diese in Verbindung mit den Arbeitsgegenständen direkt am benötigten Ort verfügbar sind. Allerdings ist ihr Einsatz nicht per se hilfreich, sondern kann auch problematisch sein und unnötigen sensorischen und kognitiven Aufwand beim Benutzer verursachen. Dieser zeichnet sich dadurch aus, dass er nicht durch die eigentliche Aufgabe, sondern viel mehr durch den Einsatz eines Augmented-Reality-Systems verursacht wird. Mögliche Gründe dafür sind zum Beispiel, dass relevante Teile der physischen Welt durch virtuelle Grafiken verdeckt werden oder dass virtuelle Objekte aufgrund ihrer Positionierung im Raum schlecht erkennbar sind. Damit Augmented Reality trotz derartiger Problemquellen ihr volles Potenzial entfalten kann, ist es notwendig, ein besseres Verständnis dafür zu entwickeln, wie Informationen, z.B. Arbeitsanweisungen, mit ihr angemessen dargestellt werden können.

Als ein Ansatz zur Vertiefung des Verständnisses werden in dieser Arbeit das Information Representation Modeling for Augmented Reality (IRMAR) Framework und, auf diesem aufbauend, Darstellungsherausforderungen (engl. Representation Challenges) entwickelt. Das Ziel dabei ist, mögliche Quellen von unnötigem sensorischen und kognitiven Aufwand (engl. Sources of Unnecessary Sensory or Cognitive Effort) zu identifizieren, die durch die Nutzung von Augmented Reality zur Unterstützung prozeduraler Aufgaben verursacht werden, und mögliche Maßnahmen zur Lösung aufzuzeigen.

Dazu wird in einem ersten Schritt ein Ansatz zur Modellierung von mit Augmented Reality dargestellten Informationen entworfen. Als dessen zentrales Element werden dabei die Informationsobjekte (engl. Information Objects) herausgearbeitet. Diese sind physische oder virtuelle Objekte, die einem Benutzer relevante Informationen zur Erledigung von einem oder mehreren Arbeitsschritten einer prozeduralen Aufgabe vermitteln. Darunter fallen z.B. eine physische Schraube, die gelöst werden soll, eine Animation, die die notwendigen Bewegungen anzeigt, oder auch ein Text mit relevanten Hinweisen dazu. Ein grundsätzlicher Unterschied zu vorab vorhandenen Modellierungsansätzen, z.B. für klassische Benutzerinterfaces aber auch für Augmented Reality, ist die Einbeziehung der Verbindung, die Informationsobjekte mit der physischen Welt haben, sowie des sich daraus für den Betrachter ergebenden Informationsgehalts.

Darauf aufbauend werden im zweiten Schritt charakteristische Merkmale (engl. Characteristics) der Informationsobjekte bestimmt, die sich entweder auf ein einzelnes

Objekt oder auf eine Kombination von mehreren beziehen. Beispiele dafür sind, dass Informationsobjekte bedingt durch ihre Anordnung in der physischen Welt immer in einer räumlichen Beziehung zueinander stehen (Räumliche Beziehung - engl. Spatial Relationship) oder dass zwischen verschiedensten Informationsobjekten oftmals semantische, also für einen Betrachter sinntragende, Verbindungen bestehen (Verbundenheit - engl. Connectedness). Um sie als Basis für die folgenden Schritte nutzen zu können, werden nur solche Merkmale betrachtet, die entweder für alle oder zumindest ein Großteil der der Informationsobjekte gelten.

Im dritten Schritt werden schließlich die Quellen unnötigen sensorischen oder kognitiven Aufwands systematisch hergeleitet. Das geschieht durch Kombination der charakteristischen Merkmale. So ergeben die beiden vorab genannten Merkmale Räumliche Beziehung und Verbundenheit zusammen die Quelle, dass bedingt durch die Art der räumlichen Darstellung in Augmented Reality, schnell der Fall eintritt, dass vorhandene semantische Verbindungen schwer zu erkennen sind. Um die praktische Anwendung zu erleichtern, werden die ermittelten Quellen als Darstellungsherausforderungen formuliert, die beim Design einer Darstellung von Informationen mit Augmented Reality gelöst werden müssen. Insgesamt werden die sechs Herausforderungen Klarheit (engl. Clarity), Konsistenz (engl. Consistency), Sichtbarkeit (engl. Visibility), Orientierung (engl. Orientation), Bewegungshinweise (engl. Motion Cue) und Informationsverknüpfung (engl. Information Linking) identifiziert. Für jede von ihnen wird beschrieben, was die Herausforderung ausmacht, wie sie aus den Merkmalen hergeleitet werden kann, wo sie ggf. in der Literatur bereits beschrieben wurde und welche Maßnahmen zu ihrer Lösung existieren. Diese Form der Beschreibung wurde gewählt, um Designer von Augmented-Reality-Systemen dabei zu unterstützen, die Informationsdarstellung mit geringem unnötigen sensorischen und kognitiven Aufwand für einen Benutzer zu gestalten.

In einem weiteren, darauf aufbauenden Schritt wird die Herausforderung Informationsverknüpfung näher untersucht. Sie besteht darin, dass semantische Verbindungen zwischen einem Informationsobjekt und einem Ort oder einem anderen Objekt leicht erkennbar sein müssen. Als Ausgangspunkt wird zur Strukturierung des Themenfelds eine Taxonomie erstellt, die beschreibt, welche grundsätzlichen Möglichkeiten zur Visualisierung von Informationsverknüpfungen in Augmented Reality bestehen. Durch die Einordnung der in der vorhandenen Literatur beschriebenen Darstellungen in ihre Dimensionen wird deutlich, dass es an Maßnahmen zur Lösung dieser Herausforderung mangelt. Daher werden zwei empirische Studien mit menschlichen Probanden durchgeführt, in denen untersucht wird, wie sich verschiedene Arten von visuellen Darstellungen auf die menschliche Leistung bei der Erkennung von Verbindungen auswirken. Die Ergebnisse zeigen eine klare Präferenz bezüglich der Darstellung, woraus sich neue Maßnahmen zur Lösung der Herausforderung Informationsverknüpfung ableiten lassen.

Als letzter Schritt wird schließlich die Anwendung von IRMAR, einschließlich der Herausforderungen und der Lösungmaßnahmen, für zwei Anwendungsfälle gezeigt. Der erste ist die Integration in die ISO-Norm 9241-210 Human-Centered

Design. Sie wird anhand des Entwurfs eines Augmented-Reality-Systems veranschaulicht, das den Ausbau eines Fahrzeug-Airbags unterstützt. Die zweite Anwendung ist die Entwicklung der Augmented Reality Procedural Task Modeling Language (ARPML), die zur Erstellung von Augmented-Reality-Unterstützung für prozedurale Aufgaben dient. Auch hier wird der Airbag-Ausbau als Beispiel genutzt, diesmal um die Verwendung von ARPML zu demonstrieren.

Contents

Acknowledgement

I would like to take this opportunity to thank the people who especially helped to bring this thesis to a successful conclusion. Without them, this work would not have been possible.

Most of all I would like to thank my wife Wei and my two children Dorothea and Johannes. They have endured that I spent long hours writing this thesis and through this always supported me.

I would also like to thank my colleagues at Bosch Research in Hildesheim and Renningen. Special thanks go to Gerrit DeBoer, Michael Dambier and Thomas Wittig, who supervised the work at Bosch, and Clemens Günther, who was always willing to help with any questions concerning vehicle repairs.

Many thanks also go to Reinhard Keil, my supervising professor, who time and time again gave me the necessary impulses to develop the concepts presented in this work.

Finally, I would like to thank the students whom I supervised for their internships and final theses and who helped to explore individual topics in more detail.

1 Introduction

For modern cars, there are countless configuration options that allow customers to individually adapt their vehicle to their needs, so that many specific configurations are only built a few times or even only once. One consequence are significantly more complicated assembly, maintenance and repair processes during a vehicle's life cycle. In addition, workers and technicians are likely to have less experience and less knowledge of specific configurations. As a result, even with an increased amount of training, these processes still become more and more error-prone and time-consuming. Other industries also face comparable challenges. For example, similar to vehicles, industrial machines are becoming increasingly complex with new and improved functions, while at the same time their variant diversity is increasing, leading to similar problems as with vehicles.

In many cases, the aforementioned processes still include a significant amount of manual activities. Usually, such work processes are referred to as procedural tasks if they can be subdivided into individual work steps[1] [248, 90]. These are defined as a sequence of individual, self-contained work steps which, when fully completed in the right order, achieve a certain goal [248, 90]. Complexity can range from very simple tasks, such as replacing a bulb in a headlamp, to very complex ones, such as assembling a vehicle from thousands of parts. For each procedural task, the people involved need to have the necessary information available on how to complete its work steps [248]. In many cases, this not a problem because relevant information can be acquired through several means. While for often-repeated tasks this might be training or experience, for simple tasks this might be reading descriptions or trial and error. However, for complex and rarely repeated tasks, acquiring the needed information becomes more problematic.

In the context of automobile production, Tümler [290] describes this information need with a model in which Taylorism, where each individual worker often repeats work steps in the same form, and the Volvo-Udevalla model, where each worker can conduct all possible work steps, are two opposing approaches of how to organize work. Real production is located in between these two extremes. However, the closer it is to the Volvo-Udevalla model, the higher the individual worker's need for information is [290]. Due to the increasing product complexity and the number of variants, there is a trend in this direction. As a possible way to satisfy this need for information, Tümler identifies the use of augmented reality as a supporting technology (see figure 1.1).

[1] Throughout this work sometimes also the phrase manual procedural task is used to emphasize the manual aspect

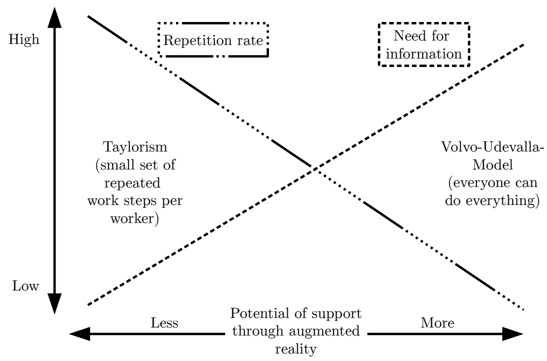


Figure 1.1: Support potential through augmented reality depending on the type of work organisation (Own work based on Tümler [290])

Augmented reality has shown its potential to effectively support the execution of manual procedural tasks as it provides a unified view that combines the task area with further contextual information, like, e.g., work instructions [135]. In comparison with other media such as printed instructions or instructions displayed on a computer monitor, the use of augmented reality can lead to better results in terms of used time and error rate [281, 27]. Even compared to instructions displayed on a head-mounted display without being registered to the physical world, it can lead to to significantly better results [281]. In particular, fewer errors and follow-up errors are made, and the mental load of the conducting persons is lower. The assumed reason for these results is that augmented reality is very well suited to compensate a deficit of spatially located information [287]. Furthermore, by merging the workspace and the relevant information, less attention changes are necessary [281, 205]. This is evident, for example, by fewer head movements that are otherwise necessary to read information relevant to complete the work steps [135]. Also, the support through augmented reality is not only limited to complex tasks but can also help with relatively simple tasks to compensate an information deficit [231]. Even before starting a work step, augmented reality can already help by guiding to the work place [136, 135]. In summary, augmented reality can be a suitable means to support procedural tasks.

However, it is also important not to burden a user unnecessarily. Instead, "the primary task of the user [is not] the interaction with the AR system or the information retrieval from the AR display, but [the] handling of real objects" (Tümler [290], 2009, page 73, own translation). For this reason, adapting augmented reality to human needs is an active field of research. Concerning basic perception, extensive investigations have already been made on how augmented reality must be designed in order to fit optimally to human senses. This includes aspects like color [176, 179], or depth perception [118, 181]. Visualization patterns for recurring problems were also developed and tested. Examples are the virtual window for a better presentation of spatial depth [118] or focus and context visualization for a better integration of virtual objects into a physical environment [155]. In addition, design guidelines for two-dimensional user interfaces and virtual reality interfaces were transferred to augmented reality [89]. For certain use cases like assembly training [301], guidelines have also been developed that give guidance on how augmented reality support must be designed. Thus, the need for augmented reality to be designed in such a way that it is adapted to humans has already been recognized and addressed in research.

Despite these very promising approaches to support the execution of manual procedural tasks and the numerous findings on what needs to be considered to adapt augmented reality to human needs, there are still many open research issues. One, as will be shown in the further course of this work, is how to design the representation of information in augmented reality with the goal to support humans performing procedural tasks. Here, representation does not refer to the actual presentation or visualization of information, but rather how it is embedded in the physical world. Therefore, it can be understood as the treatment of containers for the information to be shown [287].

Inappropriate representation creates the risk of unnecessary[2] sensory and cognitive effort resulting from a non-optimal combination of virtual objects with the physical world. An example of this are font colors which may be appropriate for regular screens but may decrease the legibility of a text in an augmented reality system, when they do not sufficiently stand out from natural backgrounds. Another example are virtual objects that are positioned such that their connections to physical objects are only unambiguous when seen from a certain perspective. Moving away from this perspective pushes the burden to recognize the connections on the user. Even if the cause is different in the two examples, the result is similar: a user is impeded from performing the actual task and has to manage seeing (sensory effort) and understanding (cognitive effort) the information displayed with augmented reality. Thus, when this unnecessary effort is not minimized, completing tasks will take longer and be more error prone than needed all while a user will experience increased cognitive load and faster fatigue [281, 89].

1.1 Research Goal

The main goal of this work is to identify possible sources of unnecessary sensory and cognitive effort caused by embedding virtual information in the physical world using augmented reality and provide approaches to mitigate them. As already discussed, this is barely about the actual visualization of information, which can be very diverse depending on the use case, but more about its embedding into an augmented reality scene. Examples for this have been given in the introduction. To increase usefulness, part of this goal is to formulate these sources as challenges that need to be solved for optimal support of procedural tasks with augmented

[2]The phrase unnecessary is used here to express that the effort is not necessary to complete the task and is only caused by the use of an augmented reality system

reality and identify concrete measures that can be used for solving them. This also includes contributing new measures based on empirical research. Two further goals are derived from the main goal. The first one is to find a suitable way of modeling information representation with augmented reality for procedural task support that facilitates the identification and description of the challenges. The intention behind this goal is to achieve a structured derivation of the challenges based on the specific characteristics of augmented reality. The second is to show the practical use of the developed modeling approach as well as the identified challenges in the creation of augmented reality systems. This is to demonstrate applicability in realistic settings, give guidance on its use and provide an example for how it can be utilized.

In this context, sensory and cognitive effort are not further distinguished, as they are very closely linked to each other and a clear distinction is often very difficult. This is especially true for the visual sense [40]. Even if there are different sensory modalities of augmented reality, only the visual extension of the physical world is considered here, because it is the predominant one in industrial applications [235, 109].

Not part of this work are some other aspects of supporting procedural tasks through augmented reality. In order to clearly define the focus, a few aspects that one might assume to be part of this work are briefly addressed here but will not be explored further.

Other application scenarios beyond the support of manual procedural tasks are not considered. One reason for this is that some basic assumptions, e.g., that sensory and cognitive effort should be minimized, cannot be generalized. For example, it is important to actively generate a cognitive effort in augmented reality training for best results [124, 90]. Generally, it is problematic to generalize design knowledge about augmented reality because it is very context-dependent [89].

Aspects like interaction or user interface design are also not considered, as they are a very large topics on their own. Instead, a well-designed representation of information should decrease the need for interaction. Similarly, and as already mentioned, the concrete visualization of the instructions is not considered in detail and instead, the focus is on embedding them into augmented reality scenes.

Instructions for procedural tasks can either "provide information in the form of procedures (task-oriented instructions), principles (system-oriented instructions), or examples (instance-oriented instructions)" (Eiriksdottir and Catrambone [90], 2011, page 1)[3]. However, these different approaches are only marginally relevant for this work. They have an influence on which information is included and how a visualization has to be designed. They do not affect the fundamental challenges associated with embedding virtual information in the physical world and thus are not further considered.

[3]Throughout this thesis, the page numbers given for citations refer to the page of the respective publication, not to the book, journal, etc. in which it was published. Particularly in view of the fact that articles are often made available as independent PDF files, this scheme should make it easier for a reader to find the citation.

This work is also not confined to a concrete device type, e.g., head-mounted displays, but instead aims to identify general challenges of information representation with augmented reality.

1.2 Approach

The approach used in this work to reach the previously formulated research goal is as follows.

First, the fundamentals of augmented reality are discussed to gain an understanding of the state of the art concerning its possible use as a means to support manual procedural tasks. For this purpose, the technical foundations are presented first, i.e., how augmented reality is defined and what technologies are available to combine virtual images with the physical world. Through this, it becomes clear that augmented reality is an appropriate modality to display information integrated into the physical world and that the technologies required for its implementation are available. Further, existing approaches for using augmented reality to support procedural tasks are discussed. This leads to two particularly relevant findings: firstly, the potential of augmented reality to support procedural tasks is proven by multiple studies, and secondly, the topic of avoiding unnecessary sensory and cognitive effort caused by the use of augmented reality has not yet been researched sufficiently. Furthermore, in order to provide a basis for modeling information representation in augmented reality, existing approaches for similar purposes are discussed. It becomes clear that they are not sufficient, especially since they are missing the integration of virtual and physical objects into one combined consideration. (**Chapter 2: Fundamentals of Augmented Reality for Procedural Task Support**)

Therefore, in the next step, a more refined approach to modeling information representation in augmented reality for the context of procedural task support is developed. The result is the Information Representation Modeling for Augmented Reality (IRMAR) framework. Because information from both physical and virtual sources is relevant for a user to complete a task, the framework goes beyond the mere consideration of virtual objects. Instead, IRMAR combines information from both sources into one modeling approach. This is essential since these sources can only be considered in combination, because in augmented reality both virtual and physical objects share the same space. For this purpose, the concept of information objects is introduced, which are physical or virtual objects that provide information relevant for one or more work steps. Examples of these are a bolt that must be removed, a virtual model of said bolt or a textual descriptions of the bolt's removal. The applicability in augmented reality systems is shown through examples. Another part of the framework is the identification of characteristics of these information objects, which are relevant for the use of augmented reality to represent information for procedural task support. These are used in the further course of this work to identify representation challenges. An example of such a characteristic, called spatial relationship, is that information objects are spatially related to the physical world and located in it. Another example characteristic, called combination, is that information objects as seen from a user's perspective are combined into one view and can interfere, e.g., occlude or distort, with each other in this view. This approach taken for the IRMAR framework is different from previous ones. Particularly, the combined consideration of virtual and physical objects as sources of information and the explicit identification of the resulting characteristics is novel. This creates a basis for a structured examination of avoiding unnecessary sensory and cognitive effort. (**Chapter 3: Information Representation Modeling for Augmented Reality**)

As a further part of the IRMAR framework, it is then explored what can be gained from combining the characteristics of the information objects. The idea is that such combinations can be used to identify possible sources of unnecessary sensory and cognitive effort which do not originate from the task at hand, but from representing information with augmented reality. Coupling the two characteristics spatial relationship and combination, for example, implies that the information objects may be seen from an unfavorable perspective and in front of a background which may make them hard to recognize. As a result, a user must spend sensory and cognitive effort to see and interpret them. With this approach, six relevant sources comparable to this example are identified, and then formulated as representation challenges that must be solved when creating an augmented reality system to support executing procedural tasks. The previous example is formulated as the clarity challenge. Existing work is considered in two ways: first, it is determined where and how such a source or challenge has already been previously identified, and secondly, it is collected what is generally known about the challenge and what ways have already been described to solve it. For this, published studies and general literature references, which provide information on the respective challenges, are evaluated. This way of describing a challenge is intended to help designers of augmented reality systems achieve adequate information representation which avoids creating unnecessary sensory and cognitive effort. While some of these challenges have already been described previously, at least partially, this is not the case for all of them. The clarity challenge, for example, is one which has already been described before. Nevertheless, this description is not only useful for the new but also for previously described challenges. The reason is that this way they are clearly defined, it is shown why they arise from the inherent characteristics of augmented reality and what measures exist to solve them. Finally, possible design conflicts are also considered and discussed. These are situations in which not all challenges can be solved to the same extent because the solutions are fundamentally contradictory. This approach differs from existing ones in that it does not try to analyze and cluster solutions from existing systems, examine how human perception is affected by augmented reality on head-worn displays or similar devices or find solutions for known visualization problems. Instead, the previously identified specific characteristics of information objects are used to determine sources of unnecessary sensory and cognitive effort, which arise directly from the nature of augmented reality. (**Chapter 4: Representation Challenges**)

One representation challenge, the information linking challenge, is examined in more detail. It addresses the need of a user to be able to recognize connections between objects, e.g., between an annotated object and the annotation, with minimal sensory and cognitive effort. The reason for selecting this challenge is that the information available in the literature is sparse, even though it is clearly relevant for information provision during procedural task support. In order to structure the subject area, a taxonomy for possible ways to visually show connections between information objects or between information objects and places is developed. It includes three dimensions, which characterize how a visual connection is established, how the reference systems of the connected objects relate to each other and how information objects are integrated with each other and the physical world via a graphical context. It thus describes the design space for solving the information linking challenge. The dimension visual connection is further examined via two empirical studies in order to better understand how a connection of virtual to physical information objects should be graphically presented. Reason for selecting this dimension is that the previously available research findings are not sufficient to give indications on how to solve this aspect of the information linking challenge. Result of the studies are clear recommendations on the type of graphical connection that should be used to solve the information linking challenge. (**Chapter 5: Information Linking**)

Besides developing the IRMAR framework including the representation challenges, a further goal of this thesis is showing the framework's applicability and giving guidance concerning its use. Thus, a first application shown here is the integration into the ISO norm 9241-210 human-centered design [80]. For this purpose, an approach is developed in which steps to identify relevant information objects and constraints on how to solve the challenges are embedded in the activities described by the norm. Through this, an understanding of how the information representation must be designed for each work step of a task is built up iteratively. The integration is kept sufficiently general so that it can potentially serve as a reference for the application of IRMAR in other design process frameworks. To facilitate understanding, the integration is illustrated through the exemplary design of an augmented reality system that supports the removal of a vehicle airbag. (**Chapter 6: Using IRMAR in Human-Centred Design**)

The second application included here is the the Augmented Reality Procedural Task Modeling Language (ARPML), which an approach to authoring augmented reality support for procedural tasks based on the IRMAR framework. With it, individual work steps are described via a combination of information objects. In addition, aspects are built into ARPML that support solving the representation challenges automatically during the executions of a procedural task. While it is not limited to automotive repairs, it was created specifically for this application. Therefore, ARPML was designed to describe many different variants of a task with as little overhead as possible due to the many variants a type of vehicle can have. This modeling language shows that the concept of the IRMAR framework is very versatile and not limited to the design of user interfaces. The airbag removal task is again used as an example. (**Chapter 7: The Augmented Reality Procedural Task Modeling Language**)

Finally, this approach and the results achieved are discussed. It is reviewed what has been accomplished with this work and in how far the previously stated research goals have been met. A look is also taken at what is still open for future research and in what direction the IRMAR framework including the representation challenges could be developed. (**Chapter 8: Conclusion and Outlook**)

2 Fundamentals of Augmented Reality for Procedural Task Support

This chapter provides an introduction to augmented reality for procedural task support. In particular, it has two purposes: one is to lay the foundation for understanding this work, and the other is to highlight the gaps in the state of the art that will be addressed in the further course of this work.

There is an almost incomprehensible amount of literature on augmented reality that covers a wide range of aspects. These include, for example, descriptions of system architectures, algorithms for combining virtual graphics with the physical worlds or human perception in augmented reality systems. Nevertheless, not all of them are important for the purpose of understanding this work. To provide an overview, the fundamentals relevant to this work are therefore presented in this chapter.

First, it is introduced what augmented reality is and how it has been defined in the literature. Next, the technologies that can be used to create augmented reality are discussed, albeit only to the extent necessary to understand this work. The focus is on the use of tracking and display technologies, as they are specific to augmented reality, while aspects such as rendering 3D graphics are omitted, as they are very general and contribute little to this work. The main goal in presenting the different technologies is to show that augmented reality is a suitable way to combine virtual information with the physical world.

In the next step, the approaches to support procedural tasks with augmented reality are discussed. After a short explanation of what procedural tasks are, it is examined why augmented reality has the potential to support them. A review of the literature on the possibilities for support that have been investigated so far is discussed, specifically which study results are available and which metrics are used to determine success or failure. It becomes clear that augmented reality has shown its potential for support in practice and that especially complex tasks with spatial reasoning can benefit from this.

Following this review, the state of the art is examined with regard to how information supporting the execution of procedural tasks can be optimally represented in augmented reality. For this purpose, a further literature review is carried out on approaches to designing and modeling information representation in augmented reality. It is shown that there is so far no structured approach that supports avoiding unnecessary sensory and cognitive effort caused by inapt representation. Further, the found approaches can also not be directly applied to challenges in information representation. For example, they consistently lack the combination of virtual and physical information and are mostly analytical.

Finally, it is discussed which conclusions can be drawn from this chapter for the further course of this work. In particular, it is addressed which gaps are still open and why, based on the literature, it is promising to address them with this work.

2.1 Definition of Augmented Reality

Augmented reality is the extension of a human's physical environment with artificial sensory impressions. However, since this definition is very general and perhaps even too general [195], various and more specific definitions have been created that allow a more detailed understanding and investigation of augmented reality systems.

One of the first definitions of augmented reality came from Milgram et al. [195] (see figure 2.1). To examine the concept of mixed reality as a combination of virtual and physical parts in an environment, they describe a continuum, called reality-virtuality continuum, that is located between two extrema. One extremum is the physical environment without any virtual objects, while the other is a fully simulated virtual environment separated from the physical world. The range between them is what they call mixed reality. Augmented reality is defined as the part of the continuum that is near the real environment where some parts are virtual, but most are not. This definition was made in the context of a larger framework that includes different types of systems on the continuum, multiple classes of displays and a taxonomy for mixed reality systems. Through the additional parts, it is further specified what is understood as augmented reality, however, these aspects are not relevant for this work.

Compared to this rather complex definition, a more compact and often used one for augmented reality was presented by Azuma [23]:

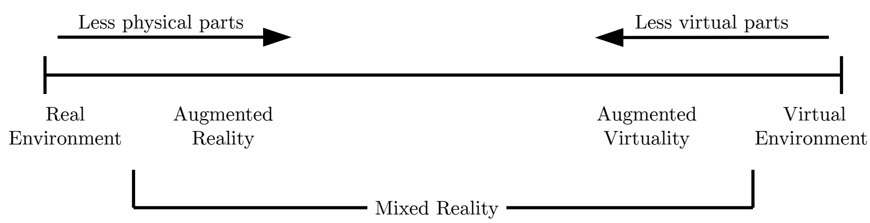


Figure 2.1: Representation of the Reality-Virtuality-Continuum (Own work based on Milgram et al. [195])

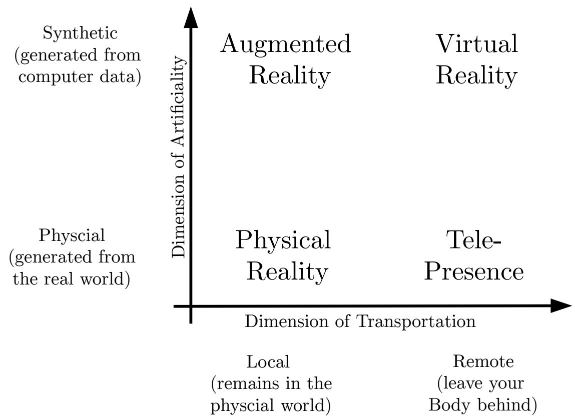


Figure 2.2: Definition of augmented reality based on the combination of two spaces (Own work based on [38])

- Combines real and virtual
- Interactive in real time
- Registered in 3-D

Their concern was to define what augmented reality is with high selectivity. Unlike Milgram et al., however, they were not interested in developing a more in-depth concept. Even if these three criteria as given above could also be applied to other senses, the authors limit their definition to the visual sense.

Another definition comes from Benford et al. [38], who examined mixed reality with regard to the aspects of artificiality of the environment, which ranges from physical to synthetic, and transport, which ranges from local to distant. With transport, the authors are referring to the distance between perception, action and the physical environment. Augmented reality is defined as a combination of synthetic in the dimension artificiality, and local in the dimension transport (see figure 2.2).

Even though more definitions have been published in addition to these three, those highlighted above adequately show the different approaches to the topic of augmented reality and also the different perspectives on it. However, these are all similar in that they always involve the extension of a person's physical environment with virtual objects or information, while the person remains embedded in the physical environment.

For this work, the definition by Azuma [23] is used. While the other two definitions include very interesting ideas, their comparatively higher complexity is not matched by any added value for considerations made here.

2.2 Enabling Technologies

For a system to fulfill the definition from the previous section, it must have at least three specific components, independent of the technologies used to implement it [23, 42] (see figure 2.3). First, the system must be able to recognize and track its relative positions to selected physical objects, which is done with a combination of sensors and tracking algorithms and is called tracker. Secondly, it must be able to create an overlay of a virtual content over the physical world, i.e., virtual 3D graphics that extend the physical environment from a user's perspective. This part is called scene generator. Finally, it must be able to show this virtual overlay in combination with the physical environment to a user. This part is called display, or in some designs, combiner. Other parts, such as interaction, data generation, etc., are needed depending on a system's functionality. A more detailed description of architectures for augmented reality systems can be found in Azuma's [23] or Billinghurst et al.'s work [42].

In the following, the individual parts of an augmented reality system are presented with varying depth and focus. The aim is not to explain the exact principles in detail but rather to give the reader an understanding of the technological possibilities. For example, nowhere in this chapter is the role of filters like the Extended Kalman Filter addressed, although they are clearly core components from a technical point of view. However, they are not relevant to understand the technical possibilities. For trackers, the focus is on which technologies and algorithms can be used to determine the relative position of an augmented reality system and its physical environment. For displays, the operating principles, possible designs and the relation to a human observer are discussed in particular. Rendering 3D graphics, however, has become a regular part of computers and mobile devices, while tracking and creating virtual overlays on a physical environment are very specific to augmented reality. Therefore, a more detailed description of scene generators is omitted in this work, as these render 3D graphics based on the tracker's inputs in such a way that they can be used as virtual overlays on a display. A good introduction into rendering 3D graphics can be, instead, found in the work of Foley et al. [111]. Nevertheless, at several places, requirements for the rendering of 3D graphics in the scene generator will be mentioned.

2.2.1 Tracking For Augmented Reality

Tracking describes the process of following the user's point of view in relation to the environment or to relevant objects in real-time [315]. However, point of view does not necessarily have to be the head or eye position. In many cases, it is instead the position of a camera which provides a video stream viewed by the user. Often, the so-called registration step is also added to the tracking process, during which the positions for tracking are initially determined. Tracking usually covers six degrees of freedom; three for the position and three for the rotation [292]. Often, a high accuracy of the tracking is required, since even smaller errors

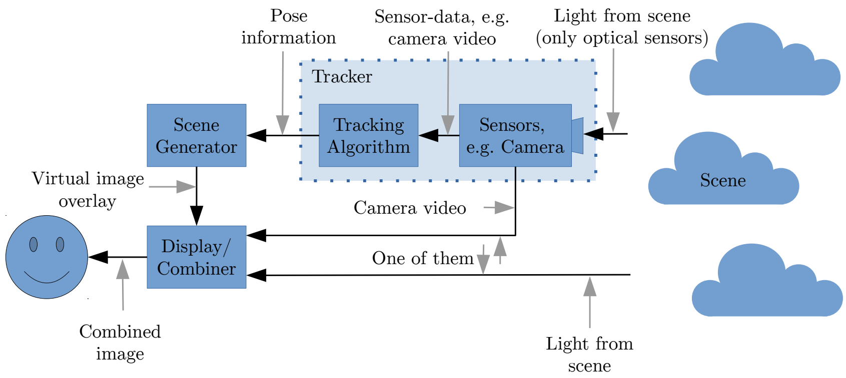


Figure 2.3: Simplified minimal structure of an augmented reality system (source: own work)

have a significant effect on a user's perception of an augmented reality system [23]. The reason is that humans are able to notice even small deviations and quickly perceive them as disturbing. Position errors have a greater effect in the close range around the user, while rotation errors have a greater effect for more distant objects [24].

There are two conceptually different approaches for tracking systems, which are called inside-out and outside-in tracking [218]. With inside-out tracking, the sensors are tied to the combiner or display and thus share its view on the environment. An example of this is an augmented reality application running on a smart phone using the camera on the backside for tracking. Outside-in tracking, on the other hand, monitors an area via external sensors. As soon as an object to be tracked is in the determined area, it is recognized by the sensors, e.g., via a marker, and located. Depending on the design, the combiner or the display can also be an object to be tracked. Not all tracking technologies can be used to implement both approaches; for example, GPS does not allow outside-in tracking.

The first tracking system built for augmented reality, completed by Ivan Sutherland in 1968, was purely mechanical [276]. It used a mechanical arm suspended from the ceiling to measure the rotation and translation of a user's head that it was attached to. Because of its appearance, this system was often referred to as the Sword of Damocles. Although this method has been abandoned for head-mounted displays, similar approaches are still used in a combination of augmented reality with measuring arms [154].

Nowadays, tracking via optical sensors has become very common, whereby there are approaches based on both visible and infrared light [42]. For infrared based systems, retroreflective dots illuminated with infrared light or infrared LEDs are used as markers [84]. The dots are filmed with an infrared camera and extracted from the captured image using computer vision algorithms. Their relative position is then used to estimate the pose, i.e., the cameras position and orientation relative to a reference system. For outside-in tracking, the target, e.g., a head-mounted display, is equipped with infrared LEDs or retroreflective dots, and a camera looks at it from the outside. An example for this design is the system by Ribo et al. [242] that tracks a head-mounted display via wall-mounted infrared cameras. For inside-out tracking, however, an environment is prepared in such a way that an infrared camera can capture enough LED-illuminated dots for typical usage situations and can thus calculate the position of the camera. An example of this is the tracking system presented by Wang et al. [297].

For visible light, so-called markers or fiducials are a way to implement tracking. These are patterns or shapes, which are known to an augmented reality system and are relatively easy to identify via computer vision algorithms [42]. Examples are colored LEDs, colored dots, simple graphical patterns, etc. Once the positions of at least four of these markers are known, a relative pose can be estimated [108]. In order to make the markers more flexible, planar 2D markers were introduced, which are shaped to provide the required four points in one marker, and at the same time allow to differentiate several markers [240].

Since it is not always possible or desirable to prepare an environment to enable tracking, so-called natural feature tracking was developed. This term refers to various methods which all use naturally occurring features of the environment, such as edges, to calculate a pose. In order for a type of feature to be used, it must possible to recognize such features from various distances and angles. For this purpose, multiple algorithms have been developed, which are able to identify corresponding features in multiple camera images, e.g., Scale Invariant Feature Transform (SIFT) [184], Speed-up Robust Feature (SURF) [34] or Fast Retina Keypoint (FREAK) [17]. The estimation of the pose based on the extracted features is then similar to that from marker-based tracking.

Even when using natural features, the environment must be known to the augmented reality system beforehand. In order to enable tracking in previously unknown environments, SLAM methods (simultaneous localisation and mapping) have been adapted for augmented reality [74]. With algorithms implementing this method, the recognized features are not simply compared against known ones, but instead a map of features is created and extended at runtime. The localization relative to this map and addition of newly detected features takes place in parallel. Based on this idea, algorithms specially designed for augmented reality, such as PTAM [161], have been developed for the precision and low latencies required for augmented reality.

Another approach is model-based tracking, in which individual objects are identified on the basis of a digital model. To achieve this, a virtual model is divided into basic primitives like lines, circles, rectangles etc. and these are compared to edges found in a camera image [66, 310]. By adding further identifying factors, e.g., object textures [238] or features of an objects surface [291], tracking can be improved.

In addition to optical methods, other technologies are also used to enable tracking for augmented reality. In the already mentioned work of Sutherland, in addition to the ceiling suspended arm, a first ultrasound system was presented [276]. In this, four localization transmitters are attached to the combiner, e.g., a head-mounted display, which emit ultrasound at different frequencies. External receivers record the emitted sounds and determine the position based on the phase differences. As an alternative setup, it is also possible to use transmitters placed at fixed positions and install the receivers on the combiner.

Another option for tracking is using inertial measurement units (IMU) [112], which do not allow determining absolute positions, but instead register movements. There are sensors that detect rotation rates, as well as those that detect accelerations and gravitation, often combined in one unit. They are characterized by their very low latency and high update rate. By integrating rotations and accelerations based on an initial position known from other sensors, absolute positions can be inferred [171].

World-wide outdoor localization is made possible by global navigation satellite systems (GNSS), such as GPS or Galileo. These use satellites in earth orbit transmitting their current position at regular intervals via radio messages. Their signals can be picked up by receivers and the position of the receivers can then be calculated based on the time differences between the signals and the satellite positions. Especially in augmented reality systems for larger outdoor areas, this is a relevant source of positioning information [100, 283]. Through additional terrestrial transmitters, it is possible to improve the location accuracy significantly [22]. Although GNSS provide a high accuracy, other tracking methods are still needed for a precise overlay of virtual graphics onto the physical world [239].

Due to the strengths and weaknesses of individual sensor types, it is often desirable to combine different types of sensors into a hybrid tracking system. The need for this has already been recognised early during the research on augmented reality [25]. A commonly used combination is optical tracking combined with inertial tracking [313, 312, 171] because the properties of both methods complement each other very well. While optical tracking can be used to determine a relatively precise pose, it has a low update frequency and decreases greatly in quality with fast movements. Inertial sensors are not suitable for determining an absolute position but have very high update frequencies and can also track faster movements. Other combinations, also from more than two sensors if necessary, are used as well. An example is a system developed by Schall et al. [257], which fuses various sensors, including GPS, inertial sensors and a camera for tracking in outdoor environments.

In addition to the methods listed here, many more, such as magnetic tracking [138], can be used. In principle, any technology that allows an object or the environment to be tracked with a sufficiently high accuracy in six degrees of freedom, and for many applications even less, can be used. It is therefore practically impossible to list every possible technology. A more detailed discussion of tracking can be found in the work of Billinghurst et al. [42] and Papagiannakis et al. [218].

The tracking techniques presented here do not only exist in the form of mathematical methods, scientific publications and prototypes, but have also arrived in practice. Various manufacturers offer software libraries with tracking functions optimized for augmented reality. For example, for smartphones, ARKit [1] from Apple and ARCore [7] from Google are available as standard on new devices and provide functions for feature tracking and SLAM. Another example is the Mixed Reality Toolkit from Microsoft [15] running on the HoloLens, which is a device designed specifically for augmented reality. There are also very powerful open Source libraries, of which the ARToolKit is perhaps the best known [3]. Further, an example of a tracking library specifically oriented towards industrial applications like those considered in this thesis is VisionLib [14].

2.2.2 Displays for Augmented Reality

Displays for augmented reality can be implemented using a wide variety of technologies and can come in a number of different designs. As already mentioned, apart from displays for the visual sense, there are also those for other senses, such as the tactile [146, 31] or the auditory [133, 35] sense. However, for this work they are not relevant and thus will not be further considered.

For visual displays, three fundamentally different types can be distinguished [292]. The first types are video based systems, which use video screens like LCDs to show a video stream in which virtual elements are integrated. They are usually used on mobile devices, where the video stream from the built-in camera is shown on the display with the virtual images mixed in. However, it is also possible to use them as head-worn displays, where the screen is placed directly in front of the eyes, or as stationary monitors. When the video display is in the line of sight, i.e., in between the eye and the observed part of the environment, the setup is called a video see-through system. The second type of visual displays are optical see-through systems, which use semi-transparent mirrors or similar optical components to project images into a viewer's eye while also keeping the physical environment visible. The source of these images can be, for example, small LCDs that are deflected by optical elements such that they appear to be positioned several meters in front of the eyes. Another example is the use of lasers whose beam is guided to directly write an image onto the viewer's retina. However, with optical see-through displays, it is only possible to add light while removal is very difficult; instead, dark picture elements appear transparent. The optical image distance of these displays is usually not identical to the distance from the eye to the physical display, so the generated images seem to float in front of the viewer. The third type of visual displays are projective displays, which use projectors to generate an image. These add light to a surface so that an observer perceives the reflected light mixed with the light from other sources. In contrast to the other types of display, the positioning of virtual elements is limited to existing physical surfaces.

Due to technical limitations of video and optical see-through systems, virtual graphics can only be displayed in a limited area at any one time. A video display has a certain size, outside of which nothing can be displayed. Optical see-through glasses are limited to a certain aperture angle within which graphics can be displayed, called the field of view. Since their projection area is restricted, this limitation also applies to projector-based displays in a comparable manner.

Unfortunately, the terminology for different augmented reality displays is not always consistent throughout literature. To avoid creating further inconsistencies, the work of Billinghurst et al. [42] is taken as reference here.

Video-Based Augmented Reality Displays

With video-based augmented reality displays, the combination of the physical world with virtual elements is completely digital [42]. This usually involves recording a video, inserting the virtual content, e.g., 3D graphics, into it and showing it a video screen in real-time. This can be, for example, the screen of a mobile phone or a computer monitor. The physical world is, in many cases, recorded by a camera, but other options like 3D scanners are also possible. With this system design, a very high degree of control is possible, as each pixel can be assigned an individual color and brightness within the technical possibilities of the screen [292]. Thus, a very precise superimposition is possible. Since the latency of the display of the video and of the virtual elements is partially controllable, they can be synchronized and presented simultaneously. Video-based displays are the only display variant that under practical conditions also allows the removal or complete replacement of elements from the physical world. This can be used, for example, to achieve correct occlusion of virtual and physical objects.

Video-based displays are often used as so called video see-through configurations [292]. In systems with such a setup, the display is placed in a user's line of sight, which in many cases means that it is attached to the head in the form of a head-mounted display at a short distance from the eyes. Displays which are head-mounted, but not placed in the line of sight, are called video miniature displays. Since a human eye cannot focus on such a short distance, lenses are used to produce a larger optical distance. Other system designs include handheld setups, where the device can be held in a hand, or virtual mirror, where the display faces the user in the form of a mirror [42].

Due to the aperture angle and the position of the camera used to capture the video stream of the environment, the there is usually an offset between the natural perspective of the eye and the one of the video when used as video see-through [292]. This is particularly noticeable in video see-through systems and can lead to disorientation, misinterpretation of object positions and incorrect depth estimates. Countermeasures include adjusting the displayed video based on eye tracking [285], or placing cameras such that they have a similar perspective as the eyes [280].

Video-based augmented reality displays have reached the real world and are no longer limited to research laboratories. The perhaps the most well-known example are smartphones. Due to their combination of computing power, built-in camera and high-resolution display, they are used for running a multitude of different augmented reality applications ranging from entertainment to industrial use [63].

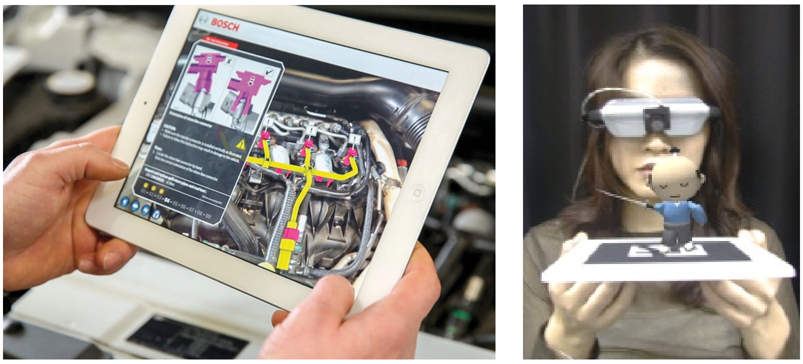


Figure 2.4: Examples for a handheld and a head mounted video see-through augmented reality display (source: Robert Bosch GmbH / Billinghurst et al. [44])

They all share the feature that they display the video from the built-in camera enriched with virtual elements on the phone's screen. Google and Apple have also recognized this and support the use with ARKit [1] and ARCore [7], respectively. An example of an industrial application running on video-based augmented reality displays is the Bosch Common AR Platform [11].

Examples for video based augmented reality displays can be found in figure 2.4.

Optical See-Through Augmented Reality Displays

On optical see-through displays, only the virtual elements are shown while the physical world remains visible. This is achieved via a combiner, an optical element that overlays the view on the physical environment with computer-generated graphics. Different from video displays the physical world is not reproduced but instead, light coming from it passes through the combiner into the viewer's eyes. For this reason, no other designs are possible with optical displays than optical see-through setups.

Traditional designs are based on the reflection of an image generated by LCDs or similar displays into the line of sight of a user [249, 60]. There are multiple ways to combine the light path from physical and virtual objects. One possibility to achieve this is via a semi-transparent mirror that allows external light to path through while reflecting light from a display into the eye [210]. Another way is to guide incoming light and light from a display with a prism into a combined light path [160, 60]. Because light originating from the display seems to come from the same direction as the environment behind the combiner as seen from a user's perspective, the virtual graphics appear to be located in the physical environment. Because the image-generating unit is usually very close to the eye, the optical path must be extended or otherwise the virtual images would appear too close to focus on. This is achieved using concave mirrors or lenses, with lenses being the preferred choice in most cases. This allows head-mounted displays to show the image produced by LCDs a few centimeters away from the eyes at an optical image distance of several meters.

Combiners built following the waveguide principle are similar to those with classical optics [156, 60]. The image is also generated with an LCD or a comparable display and may be enlarged and redirected by lenses and mirrors. However, the light is then directed into an optically dense body, the waveguide, where it is transmitted via frustrated total internal reflection (FTIR), and the angle of refraction is always too large for the light to leave the waveguide and is instead reflected inside of it. A semi-transparent mirror is inserted into the waveguide at the angular position where the image shall be seen, which reflects the light in the direction of the user's eye. At the same time, light from behind the mirror can shine through the waveguide and the mirror because its direction of propagation does not lead to internal reflection. The reflection within the waveguide alone does not extend the light path sufficiently to simulate a corresponding distance from the display, so that appropriate optics are required as well.

As another alternative, holographical optical elements (HOE) can be used. These are optical components with the special characteristic that they can change the propagation direction of light that has a specific frequency and is coming from a specific direction [18, 156]. When irradiated with appropriate light, they can redirect it such that it behaves as if it was coming from a certain distance and direction. These specific characteristics can be set during production. Therefore, if HOEs are positioned in front of an eye and then illuminated appropriately, it is possible to display virtual graphics as inside the physical environment. A positive effect of reflecting exactly the desired wavelengths while other wavelengths are much less affected is a lower energy consumption on lighting and being less noticeable than a regular mirror [156].

Virtual Retinal Displays (VRD), also called Retinal Scanning Displays, differ fundamentally from other display types [229]. While the ones presented so far have an image-generating unit, e.g., an LCD, VRDs write the image directly on the retina. For this, one or more light beams are, via one or more MEMS mirrors and a semi-transparent mirror, directed directly into the eye. The impression of an entire image is created by moving the beam over the photoreceptors in such a way that they are stimulated according the desired pattern, e.g. a specific picture. With a sufficiently fast movement a human is no longer able to temporally resolve any time differences and the impression of a steady image is created. Lasers or LEDs can be used to generate the beams. By using several different colored beams whose paths are combined by MEMS mirrors before entering the eye, multi-colored images can be produced. When a laser is used as a light source, it is technically impossible for an eye to either focus or to not focus because there is no focal plane, as all the photons have the same light path and are narrowly bundled. Thus, the virtual image is always sharp. However, since, in typical designs, the laser beams are fanned out and refocused, this effect does not occur [229].

Since optical see-through systems do not reproduce an image or video of the environment, they do not have problems with an offset between camera and eye position [292]. However, the virtual objects must be presented in a way that matches the user's point of view and focal length. Since this requires an individual

Figure 2.5: The Microsoft HoloLens and the Lumus DK-32 are two examples for optical see-through augmented reality displays (source: sown work)

adaptation to the eyes of each user, complicated calibrations can be the result [289]. Because the composition of physical world and virtual content takes place only in the eye of the viewer, any latency above the human perception threshold of about 30ms is noticeable immediately [228, 139]. With longer lasting movements, even smaller latencies can be seen via a spatial offset [206]. Thus, optical see-through displays require very fast tracking and very fast display of the virtual graphics. Another problem of optical see-through displays are their translucency and low contrasts. Since these displays can only add light to specific locations, but cannot not take it away, a bright environment can bleed through the virtual graphics or cause a low contrast. A further disadvantage is that it is not possible to display any dark images, as they simply appear transparent. There are, however, approaches being investigated to solve this with special optics [160].

Like video-based augmented reality displays, optical see-through displays have reached the needed maturity to be offered commercially. Microsoft's HoloLens, for example, is, at the time of writing, available with its second generation [9]. It allows for the display of virtual 3D objects spatially registered around the wearer while the physical environment is directly visible through the device. The HoloLens is also used for industrial applications such as the Bosch Common AR Platform [4]. There are, however, also optical see-through devices from other suppliers, such as the DK-52 from Lumus [5], or the Moverio BT-350 from Epson [10] with comparable displays.

Examples for optical see-through augmented reality displays can be found in figure 2.5.

Projection Based Augmented Reality Displays

With projection based augmented reality displays, virtual graphics are projected directly onto the physical environment. The surfaces used can be physical models, which gain a texture through the projection [234], or larger areas which are extended with contextual information [233]. However, generally any surface that sufficiently reflects light can be used. With this method, the augmentation is equally visible for all present persons, and no individual devices like head-mounted

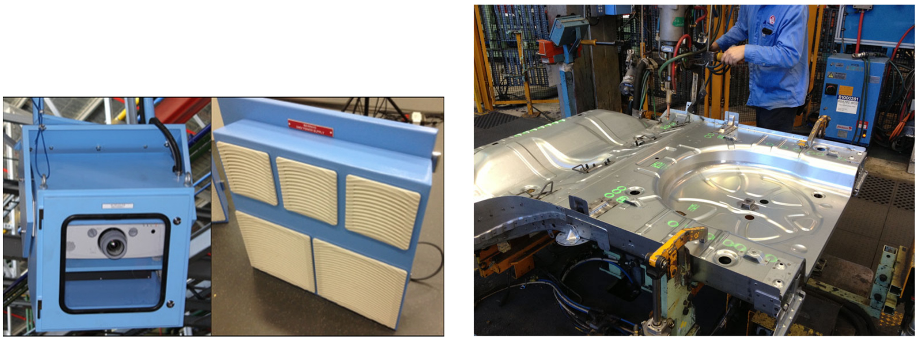


Figure 2.6: Example for an augmented reality projector and the projection created by it (source: Doshi et al. [85])

displays are needed. The use of projection-based displays is also called spatial augmented reality [45].

Video projectors based on LCDs or DLPs are usually used for projection, but laser projectors and laser-based video projectors are also possible [264]. These have the advantage compared to conventional projectors in that they do not have a focal distance and do not get blurry when projection surfaces are outside the focal plane. Thus, they can also produce a clear augmentation when on non flat surfaces. Fully covering strong textures by projection with virtual graphics is almost impossible, as they will still be visible to a certain degree. Furthermore, not every background or surface is equally suitable for video projection [42]. Projection surfaces which do not reflect light very well or are further away require very strong projectors. To achieve even and highly visible virtual graphics, surfaces can be covered with reflective foil, but this is only useful to a limited extent, as they have to be prepared beforehand [169]. Obstacles can also act as shielding, and it may be difficult to project on every surface.

Smaller projectors are not limited to being mounted at a fixed position, but can also be handheld [261] or mounted on other body parts like the head [169, 197]. This makes them easy to move and allows users to orient them where they want to have virtual information projected.

Systems with this type of display are, at the time of writing, commercially available. One example are the Werklicht projectors, which have variants with both video and laser projection [6]. They are used to extend workpieces with virtual information by projecting additional information spatially registered onto them. Their main area of application is industrial, where they are used to support manual work processes. Examples are use cases such as target/actual comparison or spot welding.

Examples for projected augmented reality displays can be found in figure 2.6.

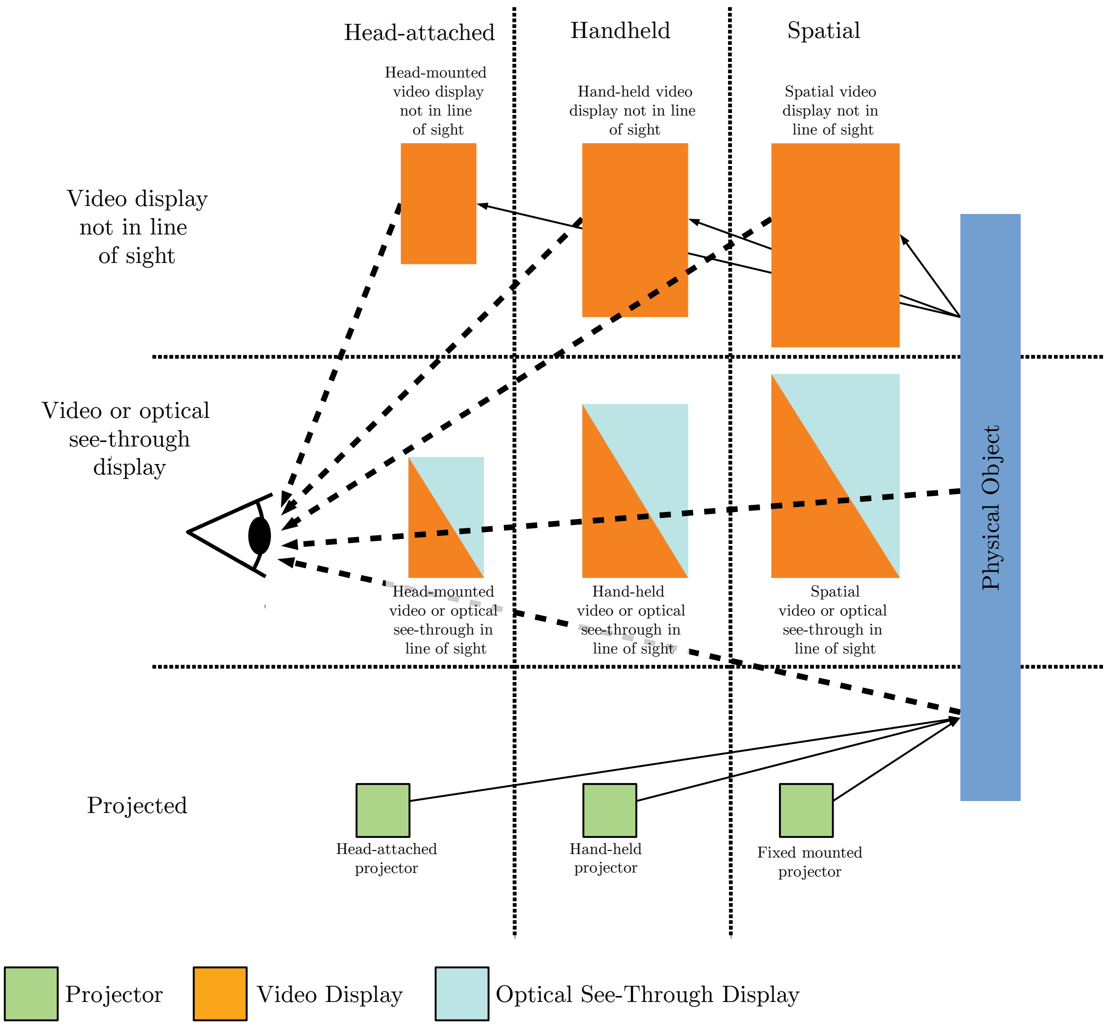


Figure 2.7: Possible positions of augmented reality displays in relation to a user's eye. The lines indicate the path of light, with thick dashed strokes representing where a viewer perceives the light to come from. (source: own work based on [46])

Relationship of Human and Display

Based on the already presented tracking and display technologies, there are various possibilities to set up an augmented reality system. Three categories can be distinguished concerning the display locations [46]. First, head-attached displays are directly attached to the user's head and often placed in front of the eyes into the line of sight. They move along with a user's head turns and follow their every move. The relative position of the eyes and the display thus remains almost unchanged. Second, handheld displays are either held by the user with a hand or are attached to an arm. Therefore, the display is no longer tied to the user's gaze direction or head orientation, but can be moved independently, to a certain degree, by the user. And finally, spatially-positioned displays are detached from the user and are instead located at fixed spatial positions. All three locations can be used with different types of displays. For example, a head-attached display can be a video see-through as well as an optical see-through display. In the same way, a spatial installation can be made with projectors or optical see-through displays. Video-based displays can also be moved to a location where they do not cover the area of the physical world, which they augment with virtual elements. Figure 2.7 gives an overview of positions for augmented reality displays in relation to the user.

2.3 Augmented Reality for Manual Procedural Tasks

Due to the potentials that augmented reality offers, there is a growing interest to use it for industrial applications such as design, manufacturing or commissioning [109]. Supporting processes such as maintenance, repair or assembly is seen by many authors as an especially promising application of augmented reality (e.g. [62, 102, 23, 211, 292, 205, 299]). In this context, the ability of augmented reality to display information in a spatial relationship with workpieces is particularly noteworthy, as it may relieve users from the cognitive task of having to transfer information from other media, such as paper manuals, into the physical environment themselves [205, 281]. In addition, augmented reality makes the relevant information directly available in the workplace and reduces activities to obtain it [205]. Different types of device can be used for such use cases, with research focusing on portable, handheld devices, head-mounted displays and fixed projectors [58]. Examples for examined system setups are are see-through head-mounted displays for repair applications [137], handheld devices for mounting instructions [130] or projectors for assembly instructions in production [57].

These potentials were already recognized relatively early-on in the development of augmented reality. Some of the first systems targeted industrial applications, e.g., by supporting the creation of cable harnesses [62], maintenance of printers [102] or assembly of vehicle components [237]. Over the last decade or two, research in this area has increased considerably [299]. This also led to many conducted user studies, so there is a substantial body of research by now investigating the benefits of augmented reality as a means of supporting procedural tasks in practice.

Not to be confused with the support of procedural tasks is their training. The aim of this is to impart skills that can still be retrieved and applied after using the training system. This requires that a certain cognitive load must be created, which is necessary for learning processes [301]. The support of procedural tasks, on the other hand, is centered around completing them as fast and error-free as possible. Whether a repetition without using the support system afterwards shows any improvement is, however, not relevant. Instead, an important factor is a minimization of the cognitive load [281]. This has major impact on how an augmented reality system must be designed.

In the following, a general overview of the state of the art concerning supporting manually performed procedural tasks with augmented reality is presented. In particular, the possibilities of augmented reality, which were also demonstrated in practical user studies, are discussed, and an overview of the current literature is given. The details of what needs to be considered when integrating information to support such tasks is discussed in more detail in the following chapters.

2.3.1 Potential of Augmented Reality

Tasks that consist of a sequence of distinct steps, which when executed in the correct order reaches a defined goal, are called procedural tasks [248, 90]. This includes, for example, assembly work in which individual parts are to be assembled one after the other to form a complete construction, or repair work in which an error is to be found and resolved through the sequence of steps. Each individual step can consist of two types of phases, or two skills required to complete it [248, 137, 205]. One type are the workpiece activities, also called psychomotor phase, for which motor skills are required. During such a phase, the attention of a person is on the workpieces that are examined or manipulated. Typical activities include inspecting or actuating devices. The other type are the informational activities, also called cognitive phase, for which cognitive skills are required. They comprise the finding and understanding of information. Typical activities include reading, comprehending or interpreting text, or transposing information from a manual to a workpiece. These two types of activities are carried out sequentially and parallelization does usually not happen [205]. Typical criteria for evaluating the performance of a procedural task are the time achieved and the amount of errors made [248]. The term procedural task was first used in the context of augmented reality by Henderson et al. [137], although Neumann and Majoros applied a similar concept before without calling it such [205].

The reasons why it is often assumed that augmented reality can support procedural tasks so well stem from the possibility of presenting virtual information connected to the physical world. It relieves a user from the burden to gain information from a different place. This facilitates low mental and physical costs accessing information, and makes it more likely that a user of such a system will actually use it. Otherwise, users are prone to disregard information which requires too much effort to attain [205]. When instructions are outside the working area, they must be viewed and remembered repeatedly, which is an additional effort that should not be underestimated. While head-mounted displays or projection systems without spatial registration of information already allow providing it at the work place, only spatially registering the information removes a user from the burden to mentally relate the presented information to the spatial surrounding. This means that augmented reality can reduce the effort to associate information to physical objects and lessen location ambiguity [205, 281], and also supports associative information processing and recalling information [205]. For use cases focusing on supporting procedural tasks, usually information is displayed that is closely linked to a location and has a high semantic relevance [306]. Further, augmented reality using head-mounted displays allows to keep the hands free for interaction with the workpieces, since no interaction with physical documentation media is necessary [281].

Neumann and Majoros examined performance when conducting manual procedural tasks with augmented reality support [205]. They found that the time spent is distributed approximately equally over both information and workpiece activities, although with slight variations depending on the task. While there are individual differences for information activities, the same workpiece activities are usually completed by different people in similar time. They also found that repeated changes between the two types of activities have a negative effect on work performance. Augmented reality can help reduce the number of such transitions to the information activities through the possibility of spatially attaching information to the workpieces. This supports a person to stay longer within the workpiece activities before changing activities becomes necessary [205]. But also the execution of workpiece activities offers potential for support through augmented reality [137, 281]. It allows displaying information about necessary interactions directly at the work site, which is especially relevant in complex situations that are not yet known to the user. This includes, for example, instructions on non-trivial movements or highlighting areas relevant for interaction. Augmented reality can also be helpful before a procedural task is even started by pointing to the work location, e.g., by arrows [136].

Industrial research has also recognised these potentials. Partners from industry and academia are intensively involved in research and prototypical development of augmented reality systems to support industrial work processes [109]. including research on maintenance and assembly processes [216, 299]. In Germany, various publicly funded collaborative projects have also been carried out, e.g. ARVIKA [115], ARTESAS [165], AVILUS [262] and ARVIDA [30], in which research was carried out on the use of augmented reality for industrial applications. It is notable that in all these projects, at least a part of the considered use cases center around supporting procedural tasks.

2.3.2 Studies on Supporting Procedural Task with Augmented Reality

Various studies have been carried out to clarify whether the potential benefits of augmented reality can be realized in practical applications. Many of these studies target production processes, but there are also studies in the field of maintenance and repair. Perhaps the most straightforward way to compare augmented reality systems with other media in terms of their suitability for supporting procedural tasks is the time in which a task is completed by test persons. However, other measures like made errors, subjective assessments, cognitive load and physiological factors have also been reported. Typical tasks for evaluations are assembly tasks, where, for example, single building blocks must be mounted to each other to achieve a specific construction. However, other study setups were also used. In the following, studies published in the scientific literature are presented, whereby the parameters just listed are used as categories to show which study results are available and which metrics are used to determine success or failure.

Completion Times

One approach to evaluate the suitability of augmented reality to support procedural tasks is to measure the time taken to complete a given task under conditions with and without augmented reality. The basic assumption for this metric is that completing a task with less time is preferable, and that augmented reality is a way to achieve a faster completion time. Various studies have found that augmented reality indeed helps to minimize completion times. However, there are others that do not achieve this result, although their authors provide some interesting justifications.

Baird and Barfield [27] compared how augmented reality instructions on an optical see-through and a video-see-through head-mounted display compared to paper-based instructions. They found that both augmented reality conditions were completed faster than the paper-based condition. Similarly, Boud at al. [51] compared how augmented reality on a head-mounted display performed in an assembly task compared to traditional media and virtual reality. Overall, augmented reality resulted in the lowest completion times. Nakanishi et al. [203] carried out a similar study in which they compared augmented reality instructions presented via a retinal scanning display and an optical see-through display with paper instructions. Again, they found that the augmented reality conditions were completed faster than the paper-based instructions. Baldassi et al. [28] compared two kinds of augmented reality with a paper-based instruction. The two augmented reality conditions differed in that one contained 3D animations and the other did not. While paper-based instructions lead to lower completion times than the augmented reality condition without animations, the condition with animations was the fastest one. The authors interpret these results that augmented reality needs to include appropriate motion cues to in order to utilize its potential. Similarly, Sääski et al. [252] also compared augmented reality including animations with paper-based instructions for an assembly task. As in the previous studies, the lowest completion times were achieved with augmented reality. The authors partly attribute this to the ability of augmented reality to play animations, which show how parts can be installed. Henderson and Feiner [136] found that for a maintenance tasks, the time to localize the task location could be reduced with an augmented reality head-mounted display compared to computer-aided instructions. The actual work steps were not completed significantly faster. However, they argue that their test persons were highly skilled mechanics and they would have been surprised if they could have been sped up with augmented reality, highlighting a limitation in the study design. Robertson [246] used a block placement task to compare multiple ways of displaying instructions on an head-mounted display. In their case, the condition, which used registered augmented reality, outperformed the others. Similar results for projected augmented reality were reached by Funk et al. [117] who compared projected augmented reality with other conditions where instructions were given via paper, a monitor and a (not augmented reality) head-mounted display. The augmented reality condition showed the shortest times for searching parts but was not faster concerning picking or mounting parts. For handheld devices, Polvi et al. [225] found that augmented reality allows faster completion times compared to the same instructions on a static image. Marner et al. [192] compared instructions given via projected augmented reality with instructions on a monitor. They intentionally used the projected augmented reality because their test setup allowed a very precise use of augmented reality, and specifically wanted to test how augmented reality performed without technical problems such as registration errors. Their results were that when supporting with augmented reality, the tasks were completed faster compared to giving instructions on a monitor, supporting the other studies highlighted in this chapter.

However, not all studies positively report that supporting procedural tasks with augmented reality leads to faster completion times than other media. When Haniff and Baber [132] compared monitor-based augmented reality assembly instructions with assembly instructions printed on paper, the users where much faster using paper. The authors attribute this, at least partially, to technical weaknesses of the used system. Similarly, Syberfeldt et al. [279] also found that their augmented reality system resulted in longer completion times compared to paper manuals. Again, the authors argue that at least part of this can be attributed to technological shortcomings of their systems. Radkowski et al. [232] compared two kinds of augmented reality interfaces for procedural task support with information written on paper but could not find significant evidence for augmented reality to be generally faster. Büttner et al. [57] made a comparison of paper instructions, instructions given via projected augmented reality and instructions via augmented reality on a head-mounted display. While they could not find a significant difference in completion time between paper instructions and instructions via projected augmented reality, the condition with the head-mounted display performed significantly worse. However, augmented reality was only used for parts picking; all other instructions were not spatially registered in the workspace. The authors explain the negative results of the head-mounted display by the lack of technical maturity of the glasses, but never elaborate whether they expected the projected augmented reality to be faster than the paper baseline. Tang et al. [281] compared augmented reality instructions on a head mounted display with paper based documentation and computer assisted instruction on a monitor and on a head mounted display. They found that users completed the augmented reality condition significantly faster than the paper condition, but not compared to the other two conditions. The authors suspect a bad display quality of the head-mounted display as a significant factor for these results. When Henderson and

Feiner [135, 136] tested augmented reality for repair and maintenance, they found that augmented reality supported locating the tasks, but a positive influence on the actual repair task compared to traditional computer aided instructions was not measurable. They conducted their tests with professionals, therefore they argued that augmented reality could not help to increase speed because their test persons were already very familiar with the tasks.

Error Rate

Another approach to evaluate the suitability of augmented reality to support procedural tasks is the consideration of (not) made mistakes. This is based on the assumption that good support ensures people having the needed information, therefore concentrating more on the task and working in a less error-prone way. Since augmented reality allows to spatially register virtual graphics and provides information at the place of work while also reducing location ambiguity, its use should relieve a user from the burden to attain the information and cognitively associate it with the work pieces. As with completion time, many studies do find a positive effect but not all concluded that using augmented reality leads to less errors.

Some studies found a reduction in the error rate compared to other forms of work instructions. Tang et al. [281] were able to find a significant reduction in errors compared to instructions presented on paper or a computer monitor, especially in dependent errors, which are a consequence of previously made errors. Similarly Polvi et al. [225] found that test persons produced less errors when they were presented repair instructions via handheld augmented reality compared to instructional images shown on the same device. Funk et al. [117] and Marner et al. [192] obtained similar results when comparing projected augmented reality at the work space with instructions shown on paper, a tablet display or a head-mounted display. Robertson [246] demonstrated that for a block placement task, registered augmented reality resulted in less errors than other methods of displaying the instructions on a head-mounted display, including a non-registered augmented reality version in which the information was shown when looking to the side. De Crescenzio et al. [75] found a very low error rate when showing maintenance instructions via augmented reality. However, they did not compare their system with a baseline like paper-based instructions.

Other studies were not able to find this positive effect, although this concerns a smaller number of studies. Büttner et al. [57] compared augmented reality instructions shown on a head-mounted display or projected onto the work area with a paper baseline. While the head-mounted display caused significantly more errors than the paper baseline, no significant difference was found between projected augmented reality and the paper baseline in terms of error rate. Henderson and Feiner [136] also did not find a positive influence on made errors when comparing the augmented reality on a head-mounted display to the standard repair instructions on a computer screen.

Subjective Assessment by Test Subjects

Augmented reality support for procedural tasks can also be evaluated based on user assessment of the provided support. Here the assumption is that good support will also be well received by test persons in studies. Again, while many find a positive a positive effect, not all studies concluded in favor of the augmented reality condition.

When comparing augmented reality instructions on a computer monitor with a paper baseline, Haniff and Baber [132] also evaluated user satisfaction. They found that users were generally more satisfied with the augmented reality instructions even though they took more time to complete the task. Similarly, when Robertson [246] compared different types of visualision of instructions on a head-mounted display for block placement, the augmented reality condition in which the virtual blocks were registered to their target position created the highest user satisfaction. Marner et al. [192] also achieved similar results when comparing projected augmented reality with conventional instructions on a computer display. When supporting the psychomotor phase of an assembly task, Henderson and Feiner [137] found that users rated the augmented reality instructions as the easiest, most satisfying and most intuitive to use compared to instructions shown on a computer monitor. De Crescenzio et al. [75] also found to users were generally satisfied with their augmented reality system for aircraft maintenance, but did not compare it to other types of instructional media.

However, other studies have also found non-favourable results for augmented reality. Henderson and Feiner conducted another study [135, 136] where they also compared augmented reality instruction with instructions shown on a computer monitor, but this time for the repair of an armoured carrier turret. The test persons expressed that augmented reality was not the most preferred condition compared to the computer monitor. Büttner et al. [57] similarly found that neither augmented reality on a head-mounted display nor projected augmented reality was more preferred than the paper based instructions, except for the factor joyfulness where projected augmented reality was the highest rated one. In Syberfeldt's et al. [279] study on object assembly with augmented reality, paper was the preferred instruction medium by the test subjects compared to augmented reality. Similarly also for an assembly task, Baldassi et al. [28] did not find any significant differences in the assessment by the test subjects of two augmented reality conditions and one paper condition concerning the factors effectiveness, helpfulness and speed.

Cognitive Load

Concerning cognitive load, the situation is rather clear-cut. There seems to be consensus among different publications that using augmented reality does in fact lower the cognitive load, or at least does not increase it. The assumption here is that a low cognitive load is desirable while working on a procedural task because this leads to positive effects like less fatigue, less made errors etc.

Nakanishi et al. [202] compared augmented reality on a head-mounted display to a paper baseline for a task which required the participants to plug multiple plugs into sockets. They found that the workload in the augmented reality condition was not higher and probably even lower compared to the paper condition. Tang et al [281] also found that using head-mounted display based augmented reality instructions for an assembly task reduced mental workload compared to instructions on paper, on a computer screen or presented non registered on a head-mounted display. Robertson [246] also found that for various display methods on an head-up display the registered augmented reality condition caused the least mental load. Haniff and Baber [132] also found a similar effect when comparing augmented reality instructions on a computer screen to a condition with paper-based instructions. However, in a comparison of registered augmented reality instructions shown on a head mounted video see-through display and paper-based instructions for an assembly task, Syberfeldt et al. [279] could not find a difference in cognitive load. Funk et al. [117] also compared projected augmented reality instructions for an assembly task with instructions on paper, a computer screen and a non-augmented reality head-mounted display. They found a trend that the projected instructions caused the least mental workload, but except for the comparison with the head-mounted display, the results were not significant. For search tasks, Biocca et al. [48] showed that augmented reality on a head-mounted display can decrease the mental workload compared to verbal instructions. De Crescenzio et al. [75] found a generally low cognitive load using augmented reality for aircraft maintenance but did not make a comparison with a baseline. However, using head-mounted displays to support procedural tasks with augmented reality seems, at least with the currently available technology, still impose a cognitive load on a user and thereby cancel out the other advantages, at least in part [33].

Physiological Factors

Another common consensus in the field seems to be that augmented reality results in less head-rotations and less eye-movements compared to traditional instructional methods. This metric is based on the assumption that doing less movements to access to instructions or documentation is preferable when working on procedural tasks.

Henderson and Feiner [135, 136] could find this effect when showing augmented reality instructions on a head-mounted display. They recorded the head movements when showing instructions via augmented reality on a head-mounted display and on a computer screen, and compared them to the theoretically optimal head movements in relation to the task area. With the computer screen condition, they found that many head-movements towards the screen occurred, while in the augmented reality conditions, the movements were close to the optimal ones. Marner et al. [192] found the same effect when comparing instructions given via projected augmented reality compared to instructions given on a computer monitor. Polvi et al. [225] were also able to find this effect, but in their case, for augmented reality on a handheld device compared to instructional images shown on the same device.

2.3.3 Usefulness of Augmented Reality for Supporting Procedural Tasks

For the aspects of completion time, error rate and user assessment, the studies published thus far have not been fully conclusive. However, overall literature on augmented reality support for procedural tasks provides reasonable evidence to suggest that augmented reality is a technology well-suited to support execution of manual procedural tasks. It is not yet fully understood how this can be accomplished or in which situation it is most successful. There seem to be dependencies on multiple factors, such as which technology is used to present the instructions [192], what knowledge or training a person using the system has [116] or how the information is presented [28, 232]. Augmented reality technology still has not reached a level of maturity where interferences caused by insufficient technology can be excluded. For example Haniff and Baber [132], Syberfeldt et al. [279] or Tang et al. [281] argue that at least part of their results, which did not confirm augmented reality as a preferable medium for presenting instructions, was caused by insufficient displays or bad tracking. This is further supported by an experiment conducted by Marner et al. [192] who used projected augmented reality with a fixed mounted projector that was perfectly registered via calibration and thus excluded any possible tracking errors. In this setting, test persons had to complete a task which required complex spatial reasoning to be correctly solved. They found that without technical obstacles, the augmented reality condition was clearly preferable concerning used time, made errors, eye-movement and subjective assessment by test persons compared to other instructional media. However, this argument of technology limitations cannot account for all findings, e.g. those from Funk et al. [117], who also used a comparable projection-based augmented reality display but did not find a significant advantage over paper-based instructions. Other studies also found that not the simple tasks benefit from augmented reality, but instead the more complex ones do, especially those which require complex spatial reasoning [279, 305]. Through this form of support, it is also possible to reduce the difference between experts and beginners by providing appropriate information [205, 116]. There are also hints that augmented reality can hinder experienced workers more than it helps because it introduces an overhead that is not justified by the gain of information [116, 135].

In summary, despite the open points and partly negative study results for augmented reality, there is enough evidence to assume that it is well-suited to support procedural tasks. This conclusion can be partly based on theoretical considerations regarding the potential of augmented reality, as well as many positive results from studies. In particular, the ability to integrate virtual information directly into the workspace and link it to workpieces, thus supporting both the cognitive and psychomotoric phases, plays an important role.

2.4 Information Representation with Augmented Reality

With augmented reality, virtual elements are inserted into the physical world. This unique combination of virtual and physical parts leads to new challenges, which relate to a variety of aspects, such as how human perception is affected by this combination. This differs significantly from other technologies that have been previously used to present information, so that new approaches to modeling and design have to be developed. One aspect of this is how to represent information with augmented reality in an appropriate way. As already described, this is about the embedding of information-carrying containers into an augmented reality environment, whereby the concrete visualization is not relevant [287]. Instead, it is considered how these containers are integrated into the physical world and what their relation to other virtual or physical objects is.

2.4.1 Information Representation in Augmented Reality User Interfaces

For some application areas of augmented reality, comprehensive approaches in the form of guidelines, frameworks etc. already exist. For example, guidelines that support the creation of user interfaces have been established for training applications [87, 88, 68]. Some are even focused on learning the execution procedural tasks [301, 304, 124]. However, these guidelines usually focus on specific use cases and can only be transferred to other application scenarios a very limited extend. Therefore, it is very difficult to apply them to the case of information representation for procedural tasks. Similarly, general guidelines exist that support the creation of augmented reality user interfaces [43, 259, 163]. The issue with these, however, is that they are generic and do not focus on procedural tasks, therefore only supporting the creation of user interfaces in general. Especially with regard to information representation, they only provide limited help. To the best of the author's knowledge, there are no guidelines for creating interfaces supporting procedural tasks or for information representation with augmented reality.

When displaying information using augmented reality, their perception by a human is of great importance and thus it is worthwhile to take a closer look at research results from this area. Kruijff et al. [168] suggested a classification of perceptual issues which occur in augmented reality. They identified three categories of occurring problems based on the nature of the problems that could be found, including *Scene distortions and abstraction*, *Depth distortions and object ordering* and *Visibility*. For these problems, there may be one of the five sources, called issues, which are *Environment*, *Capturing*, *Augmentation*, *Display Device* and *User*.

The selection of these sources is based on the processing pipeline in an augmented reality system, also called perceptual pipeline by the authors. While, for example, the environment is at the very beginning of the processing chain before it is even recorded, e.g., by a camera for further processing, the user is at the very end of the chain, as the results of the previous steps come together in their eyes. For each combination of problem category and source, they list what problems may occur and what their root cause is. With their approach, they try to cover as many different problems as possible, which also leads to the inclusion of aspects such as device properties. Their focus is especially on describing the problems and providing a comprehensive overview, but less on defining measures to solve them. Information representation is not one of these problems, but their findings provide relevant hints on how information needs to be represented. Causing perceptual problems will increase sensory and cognitive effort and thus, they must be avoided. A similar approach is pursued by Drascic and Milgram [86] by providing a comprehensive overview of perceptual issues in augmented reality, which they clustered into the categories *implementation errors*, *current technological limitations* and *hard problems*. Each of these categories also has further subcategories. Why these categories were established in this way is not specifically explained by the authors. While they expected the issues in the first two categories to be solved by better applications and technological progress, they expected that the problems in the third category need new fundamental developments in order to be solved. Where possible, they give further hints on how to handle these issues. Similarly, Livingston at al. [179] also created an overview of perception in head worn displays, which they structured based on aspects of human vision, such as visual acuity or contrast sensitivity. For each aspect, they give in-depth information on how human perception functions in augmented reality settings. As for Kruijff et al., information representation was neither in Drascic and Milgram's nor in Livingston's at al. focus, and was therefore not examined. Nevertheless, through the perceptual effects listed, they specify basic requirements that must be considered when representing information in order to avoid unnecessary effort for a user.

In addition to these comprehensive works on human perception in augmented reality, many studies on individual aspects have been published. One example for such an aspect is depth perception and related depth cues. Research has been conducted ranging from basic perceptual phenomena, e.g., Furmanski et al. [118] or Swan et al. [277], to concrete visualizations, e.g., Livingston et al. [181] or Avery at al. [19]. Other examples are color perception, e.g., Livingston at al. [179] or Gabbard et al. [121] and text legibility, e.g., Debernardis et al. [78] or Gabbard et al.[119, 122]. These works also contain solutions for some of problems relevant to the representation of information. Such findings are relevant because they give cues on how virtual elements showing information must be represented in an augmented reality system. The single studies are included together with the challenges they are relevant for in chapter 4 Representation Challenges.

2.4.2 Principles from Other Types of User Interfaces

Other types of user interfaces have already been used for some time to represent information. Thus, it is reasonable to attempt transferring the existing knowledge from these to augmented reality. In this section, existing types of user interfaces are considered and it is discussed whether the existing knowledge is transferable, and if not, which obstacles prevent it.

Classic 2D user interfaces have already been around for a longer time and have found a much wider use than augmented reality. Thus, it seems obvious to adapt established rules or even complete guidelines. However, augmented reality interfaces are, unlike those of traditional computers or mobile devices, not limited to a flat screen with a fixed size. Instead, they are integrated into the physical world whose spatial structure they have to adopt. They are also not fixed on a screen, but are viewed through a small field of view, e.g., with a head-mounted display, while their spatial dimensions extend beyond it. Thus, the transferred rules will be rather general and are not well suited for the specific characteristics of augmented reality [89]. Therefore, some aspects are transferable, but it does not appear to be useful to transfer existing modeling frameworks, design guidelines etc. to augmented reality. The characteristics of the various technologies are simply too different to achieve an adequate adaption [89]. The transferable parts are addressed in a later chapter of this work as part of the representation challenges.

Virtual reality is already more similar to augmented reality than 2D interfaces. In both cases, a user moves inside a 3D world, one simulated and one physical, and both of their interfaces have a spatial structure. In addition, many aspects of virtual reality like depth perception, have already been scientifically investigated [53, 26]. However, there is still a significant difference in the basic functionality. While the core of augmented reality is the combination of the physical and virtual world into one perceptual space, virtual reality completely diminishes the physical part [195]. While in virtual reality everything is fully controllable, e.g., every object can be moved or changed, this is not the case for augmented reality. Most physical objects cannot be manipulated through an augmented reality system, making it difficult to directly transfer findings from virtual reality to augmented reality aside from some limited aspects. Again, the transferable findings are addressed in a later chapter of this work as part of the representation challenges.

Quite close to augmented reality is augmented virtuality, which is a class of interfaces that are located on the reality-virtuality continuum between virtual reality and augmented reality [195]. While their physical environment has largely been replaced by a virtual one, individual components of the physical world are still included. So far, they have not been researched as extensively as augmented or virtual reality [260]. However, research results in this area can, in many cases, be directly transferred to augmented reality, especially if they include the interaction of virtual and physical objects [86]. In fact, most research results are not even specifically focused on it but are instead considered to be a part of augmented reality research [86]. Thus, no specific findings can be transferred from this area as well.

2.4.3 Approaches to Modeling Information Representation

While overview works and individual studies on perception in augmented reality give insights on how information must be represented in specific situations, they do not support a structured, uniform approach. Instead, conceptual considerations for information representation in augmented reality can be helpful for this purpose. Therefore, existing approaches for modeling augmented reality user interfaces are discussed in detail in the following.

One approach, a taxonomy, was created by Wither et al. [306], which they use to classify the way annotation are presented in augmented reality. Even though the focus of their work is on outdoor scenarios, the developed concepts are kept very general and can also be applied to other situations. It is relevant here because the annotations considered by them are a form of represented information. According to their definition, annotations are any virtual object that adds information to a physical object or place, with a complexity that can range from simple labels to complex virtual representations of physical structures. Every annotation has a spatially-dependent component that is bound to the object that is annotated, and a spatially-independent component, which contains the actual information. They further classify annotations by the following six dimensions:

- *Location complexity*: the complexity of the location the augmentation is attached to, which can be e.g., a 3D point, a 6DoF point or a 2D or 3D region
- *Location movement*: if the spatially independent component of an annotation moves and how free it is to move relatively to its spatially dependent component
- *Semantic relevance*: how close the semantic relation of the annotation to the annotated physical object or place is
- *Content complexity*: how complex in terms of information complexity or visual complexity the content of one annotation is
- *Interactivity*: how interactive in terms of manipulability by a user an annotation is
- *Annotation permanence*: how permanently visible an annotation is based on the source that controls this permanence, e.g., the user or an information filtering system

The main focus of their work is to explore what kind of annotations exist in different application contexts, what they have in common and where they differ.

For this reason, the individual dimensions are defined in such a way that the differences and common parts between existing systems can be described as well as possible. Information representation is only partially covered by this taxonomy. Objects that are not annotations are not considered, even if they are relevant in the context of information representation. But even with regard to annotations, their analytical approach does not contain design knowledge necessary to inform user interface design. It describes how annotations were included in existing augmented reality systems, but does not give hints on not how they should be included. However, they provide a conceptual basis for the modeling of information representation in augmented reality. Especially the dimensions provide valuable information on the possible characteristics of virtual objects, which must be taken into account when modeling the representation of information.

Concerning conceptualization of information representation in augmented reality, Tönnis et al. [287] created a taxonomy which can be used to analyse and classify the way information is represented in an augmented reality user interface. Their goal was to facilitate the analysis of augmented reality systems, support investigation of human understanding of presented information and identify gaps in augmented reality search. The dimensions were created on the basis of a targeted search for factors influencing information representation, as well as an analysis of existing systems. As part of this work, they also defined the term *information representation* in this context. Their taxonomy allows classification of information represented with augmented reality by the following five dimensions:

- *Temporality*: the lifetime of objects, which can be either continuous, when it permanently exist throughout the runtime of the system, or discrete, when it only exists for a limited amount of time
- *Dimensionality*: the spatial dimensions of objects, which can be either 2D for flat objects including texts or 3D for objects with spatial expansion
- *Viewpoint reference frame*: the point of view from which virtual objects are being presented that can be either egocentric, egomotion or exocentric
- *Mounting*: the mounting point of virtual objects which can be either human, environment, world or multiple mountings
- *Type of reference*: the extend to which virtual objects refer to a physical object or location which can be either visible objects inside the field of view (direct overlays and references), invisible objects inside the field of view (indirect overlays and references) or objects outside of the field of view (pure references)

In addition to this taxonomy, the authors also collected and analysed design challenges that arise for individual dimensions. This includes considerations such as what constitutes egocentric representation or which problems can occur in the

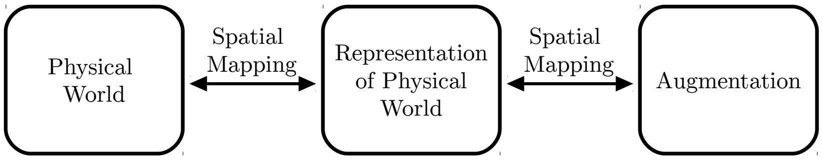


Figure 2.8: Graphical representation of the modeling of augmented reality as three categories connected by spatial mappings (source: own work based on Vincent et al. [293])

representation of spatially-registered objects. Due to their focus, however, they do not go into depth concerning human factors, but instead describe challenges related to the analysis of existing applications. For example, the discussion about egocentric representation concentrates on distinguishing them from other representations. They address open representation problems only concerning the aspect of including spatially registered objects, but do not offer solutions. Even though their approach is an important basis to model information representation in augmented reality, it is far too analysis-oriented to be directly applied to the problem of designing information representation which avoids unnecessary effort on the user's side.

Another approach to modeling of information shown in augmented reality relevant to this work was presented by Vincent et al. [293, 294]. They describe augmented reality as a combination of the categories physical world, representation of the physical world, and digital augmentation. Thereby, it includes both physical and virtual objects in a common modeling. A physical bolt, for example, would be classified in the category *physical world*, its representation in a video in the category *representation of the physical world* and a virtual model of the bolt in the category *digital augmentation*. These categories are connected by *spatial mappings*, which may have different forms depending on their strength. The strongest mapping, when a precise overlay is used, is called *conformal*. The mappings in which a spatial connection is still recognizable, but is no longer achieved by a direct overlay, are called *relaxed mappings*. There can also be no mapping at all, when no spatial overlay is used. Due to technical problems such as inaccurate tracking or problems with user interaction, it may not be appropriate to maintain the conformal mapping but allow the spatial relationships to degrade to match the particular use case and environment. The authors focused on the topic of user interaction, so that the applicability of the model to information representation was not considered. Nevertheless, the classification into physical world, representation of the physical world and augmentation is an important building block for modeling information representation. A graphical presentation can be found in figure 2.8.

2.5 Implications for this Work

In conclusion to the presentation of the fundamentals of augmented reality for procedural task support, the question arises what implications they have for this work.

Regarding the technology used to enable augmented reality, it was shown that the necessary technological basis is available and ready for use. With a general system architecture, it was established how the needed components tracker, renderer and display respectively combiner relate to each other. In addition, for the two augmented reality specific components, tracker and display, different technologies were presented which can be used to implement them. As proof of technical maturity, devices and software libraries available on the market were included as examples. From the availability of these technologies, it can be concluded that the considerations on information representation made in this work are not only theoretical but can be used for actual augmented reality systems. Further, the wide range of available technologies, especially concerning displays, also implies that the representation challenges considered must be kept sufficiently general. Otherwise there is a risk that only limited aspects off augmented reality are addressed and the desired breadth is not achieved.

Regarding the support of executing procedural tasks, it was presented that there is indeed reason to assume that using augmented reality in this context offers added value over other traditional media, such as digital and printed manuals. Making the virtual information available directly in the workspace reduces the effort for access, and embedding it directly into the physical environment results in lower effort for the transfer of information into the working environment. However, this also implies that showing information directly in the workspace, i.e., in superimposition with the parts to be manipulated during a task, is a prerequisite for enabling the potentials of augmented reality. It must therefore be taken into account when designing information representation. Various studies have been able to demonstrate this benefit in practice using metrics such as error rate, time used, cognitive effort or physiological factors. However, these advantage have not always been found. Reasons may be technical or design shortcomings or the use of augmented reality for inappropriate tasks. Nevertheless, in summary it can be assumed that augmented reality has proven to be suitable for supporting the execution of procedural tasks. Therefore, it seems reasonable to further pursue the support of procedural tasks by augmented reality. The use of metrics on physiological load, e.g., eye movements, and cognitive load in some of the presented studies is also relevant in this context. This is an indication that these metrics are relevant factors when designing information representation with augmented reality and negative impacts on them should be minimized as already defined in the goals for this work.

After determining that the medium augmented reality has the potential to improve the execution of procedural tasks, it was investigated what is known about information representation in augmented reality. In the first step, it was examined if design knowledge about augmented reality, classical, virtual reality or augmented virtuality user interfaces can be transferred to information representation. However, no structured description such as a design guideline could be found that is applicable. They either cannot be transferred to augmented reality because of fundamentally different user interfaces, are intended for other application areas, are about examining perceptual issues in augmented reality or describe visualizations that solve a narrowly defined problem. This, however, does not imply that knowledge is not already available. Instead, the information is scattered over many publications and not in an applicable or structured form. It is also unclear where gaps in this existing knowledge are. This means that for the further procedure an approach should be chosen that helps to systematically uncover these gaps. In a second step, it was also determined which approaches exist for modeling information representation in augmented reality. Here, approaches can also be found, but again they are not directly related to the problem considered in this work. There are taxonomies for different aspects of information representation in augmented reality, but they are more analytical and help to classify what already exists. None of the taxonomies provide significant indications of how to design information representation in augmented reality. In fact, they are largely mostly to the context of supporting procedural tasks. Furthermore, none of these approaches, with one exception, combines physical and virtual objects in a common modeling approach. The exception, however, is only intended for interaction design, and an application to information representation is hardly possible. Thus, not only structured design knowledge for information representation is missing, but also a modeling approach, which could help to find a structure or to identify still existing gaps. Therefore, such an approach must be created in the further course of the work.

3 Information Representation Modeling for Augmented Reality[1]

In this chapter, an approach to modeling information represented with augmented reality is developed and presented. Its purpose is to provide the level of conceptualization of augmented reality in the context of procedural task support that is needed for identifying and examining representation challenges in a following step. The aim is to achieve the goal defined in the motivation, to find a suitable way of modeling information representation with augmented reality for procedural task support that facilitates the identification and description of the challenges.

When working on a procedural task, humans interact with physical objects and modify them to achieve a certain goal [248, 90]. For example, such interactions may be assembling a motor, repairing a machine or picking parts. Object manipulation is a relevant aspect of such a task, and additional interaction with documentation can only assist, but cannot replace it [290]. Physical objects do not only have to be manipulated; they also provide information that is relevant to successfully complete a procedural task. Instructions that are printed on work pieces or status indicators like power LEDs may be the most obvious examples. However, this is not limited to explicitly given information. Instead, every object that is visible also provides information on its position, orientation, size, surface and other characteristics. An example of this is a bolt mounted on a machine. Just by looking at it, one can see the mounting position, what condition it is in or what kind of wrench is needed to loosen it. When information like this is not visible for a human, it negatively interferes with the ability to perform the task quickly and as error-free as possible. By using augmented reality, new objects are added in a human's environment that provide further information. An example may be a label giving further information on the bolt from the previous example or an animation that is showing how to remove it. Regardless of the possible advantages, this may make it more complex for a user of such an augmented reality system to see and understand the relevant information.

To address the resulting challenges, however, it is necessary to first find a suit-

[1]This chapter is based on results which I previously published in my papers *Towards a framework for information presentation in augmented reality for the support of procedural tasks* [198] and *Challenges in representing information with augmented reality to support manual procedural tasks* [199]

able modeling approach that supports their identification and description. For this reason, the Information Representation Modeling in Augmented Reality (IRMAR) framework is created and presented in this chapter. It is intended for modeling information represented in augmented reality to support the execution of procedural tasks. The focus is on the examination of those objects that provide a user with relevant information during the work steps of such a task. For this purpose, these objects are first defined, including the environment in which they are located and a vocabulary for their description is created. A way of categorizing such objects based on their relationship to the physical world is also presented. To facilitate the understanding of IRMAR and demonstrating applicability, an example classification of existing augmented reality systems is also included. Following this, special characteristics of the aforementioned objects, which result from this modeling approach, are identified and discussed. In the next chapter, these characteristics will then be used to derive representation challenges that need to be addressed for augmented reality applications which support performing procedural tasks. As a conclusion, this modeling approach is compared with those already presented in the fundamentals chapter in order to clarify both the differences and the common aspects and why a new one was developed. As already mentioned in the motivation, this approach focuses on information containers in an augmented reality scene without examining the semantic or graphic content more closely. To emphasize this, the term *information representation* is used to distinguish it from concrete presentation or visualization [287].

3.1 Basic Concepts and Terminology

The first step to develop the IRMAR framework is to determine an adequate conceptualization of information representation in augmented reality. The concern here is not the informational content or information value that is presented to a user, but instead the way containers holding pieces of information are included in an augmented reality user interface. Therefore, the concept of *information objects* is introduced. These are objects that present information to a user, which is relevant to complete one or more work steps of a procedural task. In the further course of this this work, they are the central instrument that will be used in to investigate information representation in augmented reality. How their information is visualized or even what the information is, is not relevant in the context of IRMAR. This definition without consideration of the informational content or the visualization was deliberately chosen in order to focus on the representation. Also, both physical and virtual objects are included in this definition to combine these two sources of information in one model. However, at least one physical and one virtual object must exist to meet the basic definition for augmented reality [23]. It is intentionally not defined how an object is identified as an information object, how their granularity can be determined and how an information object is separated from its environment. These aspects depend too much on the single use cases and the requirements for individual augmented reality systems. Therefore, a general definition is not possible. Instead, this hast to be a part of the domain knowledge of a designer who models information representation.

Information objects can have connections with each other, with areas or with points. This does not only include spatial connections, where, e.g., one or more objects are close to each other. Instead, they can also have semantic connections that convey relevant information to a user and are important to recognize. An example is the connection between a virtual label and the physical object that it provides information on. Such objects, points or areas to which an information object is connected are referred to in IRMAR as its *anchors*. These connections should not be confused with complex semantic relations as they are, for example, modeled in ontologies. Instead, the IRMAR framework is only concerned with those that shall be represented.

Due to the spatial nature of augmented reality, each object also has a specific position where it is located. Therefore, each information object also has a *mounting*. This is a combination of the pose, i.e., location and orientation, and the reference system in which it is located. It may depend on the anchor, but it does not have to. This will be discussed in more detail in the next section.

There will always be substantial parts of the physical world that are not relevant for completing a work step, meaning that their specific structure could be different without any impact on the task. These parts are referred to in IRMAR as the *environment*. It is necessary to include them, since augmented reality merges the virtual with the physical world, so that they have an influence on the representation. A combination of a set of information objects with an environment is referred to as a *scene*. During the completion of a procedural task, the shown scene may change multiple times, e.g., when a work step is completed and a new one begins. A user views a scene from a certain position, which is referred to as the *viewpoint*. To describe the impression that one gets from the viewpoint, the phrase *combined view* is used. In it, the environment, the physical and the virtual information objects are combined as they are seen through the used augmented reality system. It may be constrained in size by a limited field of view, also called view frustum, outside of which no virtual or, in some systems, also no physical objects or no environment can be seen.

3.2 Five Classes of Information Objects

From the previous example with the physical bolt, the virtual model of the bolt and the label, it becomes clear that information objects can be very diverse. A key factor is how tightly coupled they are with the physical world. While the physical bolt is an inseparable part of it, the label merely gives further information on it in a textual way and is not subject to any physical laws. The virtual model of the screw is somewhere between these two, since it is already matched to the physical world by its shape and size, but at the same time, like the label, is not bound by its laws. So one characteristic of information objects is that they do not all share the same connection to the physical world. These differences will become relevant later when considering the characteristics of this modeling approach and when examining the representation challenges. Based on the gradation of the spatial connection that information objects can have, the following categorization was created as part of this work:

- *Direct physical information objects*: These are information objects which are physical objects and can be directly viewed by a user without any mediation from an augmented reality display. This contains every physical object that is relevant to complete a work step, such as tools, machine parts, operating controls and indicators, switches, wires etc. Viewing an object through the combiner of an optical see-through display is still regarded as a direct view, as this may only results in a distortion, but no active processing with a possibility to modify it takes place.

- *Indirect physical information objects*: These are information objects that are reproductions of physical objects on a screen or a similar display. Their appearance can be changed, such as enlarged or colored, compared to their physical counterpart, or can be true to original. The objects of which they are a reproduction of can be information objects but do not necessarily have to be. This is the case if the physical object is of no interest and only a changed reproduction makes it relevant for the work step. For example, there may be cases where only the infrared image of an object is needed. The physical object as seen by the human eye is then not relevant and therefore not an information object. Examples for this class of objects are a video image of a tool shown on a mobile device, or images of physical landmarks in a video that are needed to easily associate the augmented video with the physical world. Because they are derived from the physical environment, the position of the underlying physical object usually indicates their mounting. However, this does not always have to be the case, for example, when an interaction with a physical object shall be indicated by moving its corresponding indirect physical information object. Anchors of this class of objects are always the physical objects from which they are derived.

- *Spatial virtual information objects*: These are information objects that are virtual and have at least one combination of size and pose in the scene that gives more information to the a user than an arbitrary combination would. By changing these two factors, even if the information object and its connections to the anchors remains well recognizable, information about spatial relations is lost. For the IRMAR framework, this combination is also called a meaningful spatial context. Virtual here means that all information necessary for the presentation of the object is available in digital form, and that no physical object has to be recorded and processed at runtime. Information objects of this category can be either 1D, 2D or 3D, whereas most spatial information can be conveyed via 3D objects. Examples are 3D models of machine parts that indicate installation positions of their occluded physical counterparts, a 2D map of wires overlaid on a wall with the real wires matching the map, or an animation that shows how to use a tool by overlaying it on a physical work item. Just because an object can be positioned in a meaningful spatial context does not mean that it will be realized this way in an actual augmented reality system. What is relevant for categorization, however, is the possibility.

- *Spatially referenced virtual information objects*: These are information objects that are virtual and have at least one point, area or another information object as anchor in the scene. Other than spatial, virtual information objects they do not have a concrete pose and size that conveys information to a user through a meaningful spatial context. Changing their mounting or scaling and rotating them is, therefore, possible without information loss. Instead, the relevant spatial information is only present in the form of one or more connections to anchors. Examples are text labels with instructions on how to manipulate an object, icon markers for points of interest or graphs from sensor readings.
- *Detached virtual information objects*: These are information objects which do not have a spatial connection, i.e., no spatial pose and size that creates a meaningful spatial context and no objects from the other categories, points or areas as anchors in a scene. Independent of this, they have a mounting. Examples are general instructions, progress indicators etc. Although they can be positioned in the scene, the representation of this type of information object does not benefit from the spatial positioning that augmented reality provides. However, they are included as a category because they also provide information for a user of an augmented reality system working on a procedural task.

The classification into these categories cannot be made independently from the context of use. An example of this is a model of a bolt that is used in a work step to show how an identical physical bolt needs to be fastened. Clearly, the model is classified as a spatial virtual information object because it has a certain size (the one of the physical bolt) and a certain pose (along the way of being fastened), which create a meaningful spatial context. However, in another work step, the very same model might be used as a generic icon for bolts with multiple bolts as an anchor. This time, there is neither a specific size nor a specific pose which creates a meaningful spatial context. Thus, it is classified as spatially referenced virtual information object. In a third work step, it might not be needed at all and thus would not even be an information object.

Direct and indirect physical information objects do not necessarily have to be explicitly integrated into an augmented reality system. For example, there is no need to define placeholder objects for them in a digital representation of a scene. Instead, it is relevant that they are considered when modeling the information representation. Nevertheless, in many cases it may be helpful to include these objects, e.g., to facilitate screen management. What is necessary, however, is that individual information objects can be identified so that a decomposition of an augmented reality scene into individual objects is possible. This is necessary for all types of information objects. If this is not the case and they can, for example, only be modeled as an inseparable unit or without a clear distinction between the objects, the approach presented here is not practicable. Adequate documentation models, which provide the needed granularity for information objects, have already been described [162, 201].

The distinction between the different classes must not be confused with a subdivision according to presentation quality. Just because an object is a spatial virtual information object, no certain graphical quality is implied. Instead, the information that can be provided to a user is the differentiating factor. The form of visualization can also cause confusion when it gives the impression that a spatially referenced virtual information object has a meaningful spatial context. This can happen when, for example, a visualization is chosen that allows only one exact position relative to an anchor. However, for categorization purposes, it is relevant if a combination of pose and size exists that conveys information through a meaningful spatial context.

As mentioned earlier, indirect physical information objects, spatial virtual information objects and spatially referenced virtual information objects can have one or multiple anchors. This is similar to the idea of using augmented reality to create annotations [306]. The anchor of an indirect physical information object is the physical object, if it is an information object, or otherwise the area which they are a reproduction of. For the other two classes of information objects, anchors can be points, areas or other information objects. Direct information objects do not have anchors because modeling relations in the physical world is beyond the scope of this framework. Detached virtual information objects do not have anchors, either because this would imply that they had a connection to the physical world. The connection to an anchor can be shown in various ways, such as the position of an information object or in the form of graphical connections. An anchor should also not be confused with the mounting of an object. While an anchor describes what an object is semantically connected to, the mounting describes at which position in which reference system an object is placed [287]. For example, a detached virtual information object can have a mounting relative to a direct physical information object without having any anchor.

Table 3.1 shows which classes of information objects are available on which display types.

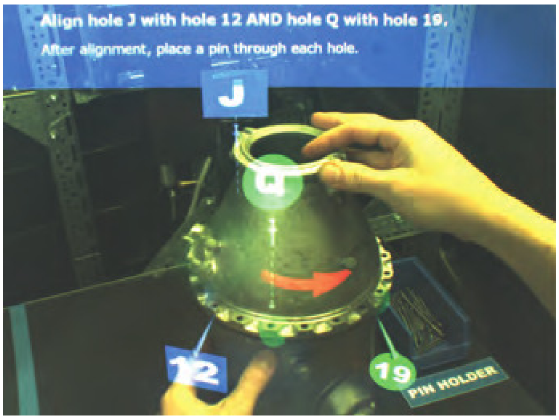


Figure 3.1: Instruction to turn a part of combustion chamber from a jet engine (source: Henderson and Feiner [137])

3.3 Example Classifications

To demonstrate the definition and the use of information objects and their classes, an example classification is made in the following section. For this purpose, systems are selected which support manual procedural tasks with augmented reality and have been reported in scientific publications. While they were chosen to give a wide range of applications and situations, the selection does not follow any particular methodology. The images are taken as depicted in the publications and interpreted based on the available information, e.g., what kind of work step is displayed, what parts are included etc. However, due to limited descriptions, it is not always possible to unambiguously describe every part or to include every information object, e.g., for Henderson and Feiner [137] or for Webel et al. [302].

While some authors conduct a statistical analysis based on their classifications, e.g., Tönnis et al. [287] or Wither et al. [306], this is not done here because the information given in the publications on the nature of represented information is not sufficient to create a meaningful evaluation. For example, a certain distribution of the different classes of information objects would not allow any specific conclusion beyond which illustrations the authors have chosen for their respective publications.

The first example for classifications can be seen in figure 3.1. It shows instructions for a user to align two parts of a combustion chamber by turning the upper part [137]. In the corresponding publication, it is specified that the system uses an optical see-through head-mounted-display. Since the step cannot be completed without moving and thus interacting with the upper part, it is an information object. Due to the display type, a user sees the actual, physical object and not a reproduction of it. Thus, it is classified as a direct physical information object. Other direct physical information objects are the pins on the right side in the pin holder that must be inserted in the holes. Another information object is the arrow which indicates the direction of rotation. Its spatial position on the surface of the

	Direct Physical IOs	Indirect Physical IOs	Spatial Virtual IOs	Spatially Referenced Virtual IOs	Detached Virtual IOs
Head-attached					
- Video See-Through	-[1]	x	x	x	x
- Video Miniature Display	x	x	x	x	x
- Optical See-Through	x	o[2]	x	x	x
- Projected	x	o[2]	o[3]	x	x
Hand-held					
- Video-Display	x	x	x	x	x
- Optical See-Through	x	o[2]	x	x	x
- Projected	x	o[2]	o[3]	x	x
Spatial					
- Video-Display	x/-[1]	x	x	x	x
- Optical See-Through	x	o[2]	x	x	x
- Projected	x	o[2]	o[3]	x	x

Table 3.1: Possible combinations of display type and class of information object
x: class available when using the display type
o: class only available to a limited extend when using the display type
-: class not available when using the display type
[1]when the video display is placed in the line of sight, the direct physical information objects are impossible to see
[2]to preserve its spatial context, an indirect physical information object will have to be superimposed on its also visible direct physical counterpart
[3]projectors can only create graphics on surfaces and thus it is not possible to create a spatial context that is not located on a surface

Figure 3.2: Instruction to remove a sensor (source: Webel et al. [302])

combustion chamber and the orientation towards the right side create a meaningful spatial context from which a user can infer information. If it were moved or rotated significantly, this context would get lost. Therefore, it is a spatial virtual information object. Similarly, the very small virtual models of the pins together with the dotted lines are also spatial virtual information objects. They provide information to a user on how to insert the pins by indicating the movement at a specific spatial position. The labels with the numbers and letters do not provide any specific information through their spatial context. They could be moved or rotated without losing any information for the user. Only the connection to the anchor needs to be shown and therefore, they are classified as spatially referenced virtual information objects. Finally, the description text above is a detached virtual information object. Even if in the text individual information objects from the scene are referenced, it is still sufficiently general and does not refer to them specifically enough to justify using them as anchors. However, this example adequately demonstrates that the boundaries between the individual classes are not always unambiguous and that there are borderline cases. Therefore, a different interpretation would also be possible in this case.

In figure 3.2, two different approaches are used for displaying the same work step to remove a sensor [302]. The used device is a handheld device with a video display. First, the left image is analyzed. The green 3D model of the sensor is a spatial virtual information object with a specific pose and size that conveys information, such as the mounting position, to a user through its spatial context. Interestingly, the description text also indicates that a physical sensor is present but hidden by the virtual model. One could argue that even though its fully occluded by the virtual sensors model, the sensor is still reproduced on the screen and thus is an indirect physical information object. However, from the still image it is not clear whether it provides any information to a user during the work step and can even be included as an information object. Concerning the arrow, the argument is the same as for the arrow in figure 3.1. It indicates how a movement must be made and can only be moved or rotated in a very restricted way before the spatial context is lost, and is therefore a spatial virtual information object. The description and the sample image on the right do not have any direct spatial connection and

Figure 3.3: Indication where to open a car bonnet (source: Stanimirovic et al. [273])

are therefore, depending on the concrete modeling, one or two detached virtual information objects. In the right image, there is another presentation for the same situation. It is very interesting to classify because of the diffuse circle that marks the sensor. One could argue that it is a spatial virtual information object because it cannot be moved without losing its spatial context. On the contrary, it can also be seen as an aspect of the video reproduction of the physical sensor, which is then classified as an indirect physical information object including the ring. The descriptive text is again a detached virtual information object. Because a handheld device is used, the physical sensors can also be seen by a user, even though it is not included in both images. Thus, it is a direct physical information object.

Figure 3.3 shows an augmented reality system running on a handheld device with a video display that instructs a user to open a car's bonnet [273]. One thing that is especially noticeable is the changed color of the bonnet in the video image in comparison to the physical car. This is an example of the change that an indirect physical information object can go through during reproduction compared to a direct physical information object. A tool to open the bonnet is also included in the augmented reality presentation. This is a spatial virtual information object because its spatial context is defined by the size of the physical tool and the pose it must be held in. As far as it can be determined, the other elements that can be seen on the tablet screen do not have any spatial connection and thus are classified as detached virtual information objects.

Figure 3.4 shows a work step during the repair of an armored personnel carrier turret [136]. The device used is a head-mounted video see-through display. The row of screws and washers, which point to their installation position with a dashed line, stands out. These are one or more spatial virtual information objects. The exact granularity depends on the desired division by the designer of the system and cannot be assumed from the given context. Furthermore, the model of the wrench with which the bolt must be screwed in is also a spatial virtual information object. A technician is shown who is using a physical wrench, presumably to fasten

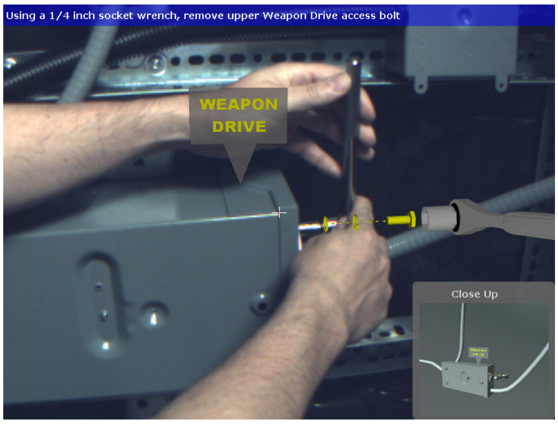


Figure 3.4: Work step from an armored carrier repair (source: Henderson and Feiner [136])

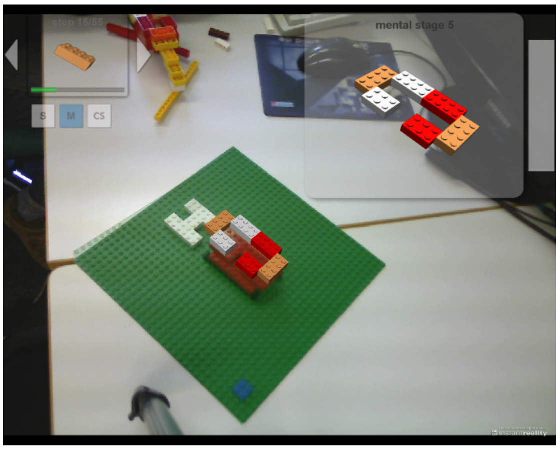


Figure 3.5: Instructions to set Lego bricks (source: Engelke et al. [95])

a bolt, which is an information object as well. Since only a video reproduction of it is visible due to the display, it is an indirect physical information object. The weapon drive label cannot be set in a meaningful spatial context but references a physical object, therefore, it is a spatially referenced virtual information object. Finally, the general description of the work step and the example model of the weapon drive are both detached virtual information objects.

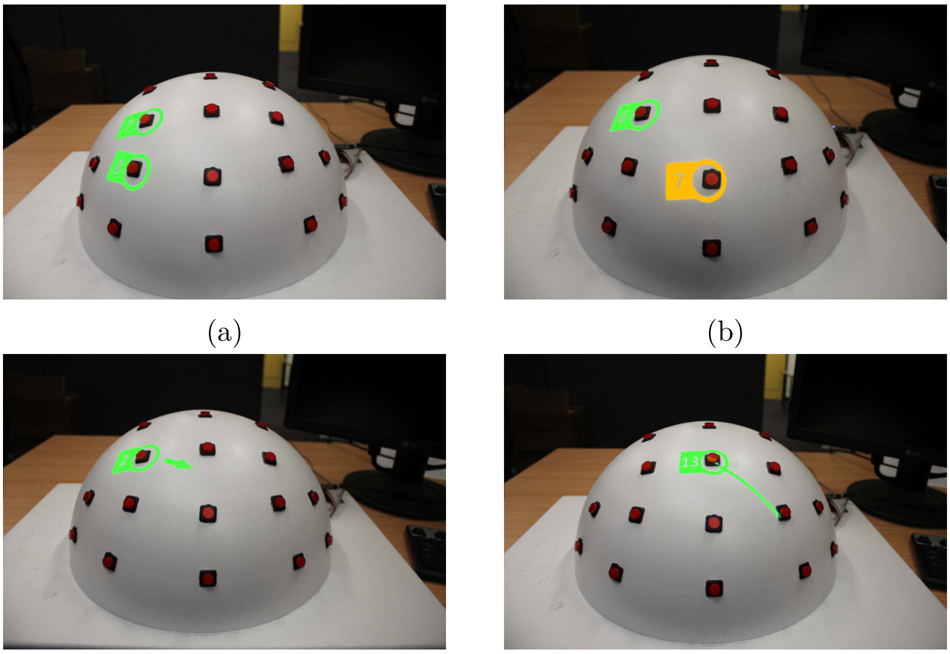
Figure 3.5 shows a scene from a system that is intended to be used for comparing different ways of displaying instructions using toy blocks [95]. The interface used is a fixed video display outside the line of sight. In the shown view, various information objects can be identified. The green base plate and the building blocks, which can be set, are physical information objects. In the figure the reproductions are shown on a computer screen, which makes them indirect physical information objects. The base plate is also superimposed with its virtual model and, additionally, some models of building blocks. Since the place where these can be positioned without a loss of spatial context is clearly defined, they are

Figure 3.6: Instruction to place a picked up block (source: Syberfeldt et al. [279])

spatial virtual information objects. Even though the mental progress bar at the top right consists of virtual models of building blocks, those are still detached virtual information objects because these blocks are not spatially connected to the physical world but instead are a symbolic representation. Similarly, the progress indication at the top left is also a detached virtual information object.

Figure 3.6 shows an instruction to place a block, which has been picked up, at a position with the same number [279]. The device used is a head-mounted video see-through display. The physical block that is being moved (hidden behind the green virtual block) is an information object, and since a video see-through system is used, it is an indirect physical information object. The two virtual models of the block are spatial virtual information objects. However, the classification of the arrow is not completely unambiguous, since its function is not fully described. It is most likely a part of the model superimposed on the physical block and thus not an independent information object at all. The image in the background is not an information object because it does not contribute to the task but instead is a technical artifact needed for the tracking functionality.

Figure 3.7 shows an experimental setup in which the test person has to press a sequence of buttons given by green and orange indications [33]. The device used is a projected augmented reality display. The buttons to be pressed are direct physical information objects, whereby at least the current and the next button belong to the information objects for each step, while the others are probably not relevant and instead are part of the environment. The projections of the step numbers to the buttons are spatially referenced virtual information objects. Due to the selected visualization with the circle around the button, it may appear as if these information objects can hardly be moved without losing a spatial context. However, the number assigned to the button does not convey any information to a user through a meaningful spatial context and could be moved without any loss of information, as long as the connection to the button is still easily recognizable.



(a) (b) (c) (d)

Figure 3.7: Instruction to press buttons and cues to the next button to press (source: Baumeister et al. [33])

Instead, the impression is created by the visualization with a circle around the button, which determines a concrete relative positioning. Thus, a classification as a virtual spatial information object is not justified. In contrast, the arrows from sub-illustrations (c) and (d) are spatial virtual information objects since they convey information through their spatial context by their position on the shortest path from the current to the next button.

Figure 3.8 shows a view from an augmented reality system that supports assembly of machinery. A work step has been reached during which a tube that is held by the user has to be mounted to two connection points [252]. The interface is shown on a video display outside the line of sight. Because they are relevant for a user to see, the physical tube and the two connection points shown on the display are indirect physical information objects. It is not discernible whether more parts of the machine are relevant for the work step and therefore have to be added as an indirect physical information object. These objects are reproductions of physical objects that are relevant for a user to see to complete the work step, but are not shown in the figure. Thus, the corresponding physical objects are classified as direct physical information objects. The model of the pipe is a spatial virtual information object because it matches the physical pipe in size and pose to convey information to the user. The relevance of the wireframe model on the right side is unclear and no description is given. Therefore, a classification cannot be made. The three symbols below seem to have no spatial connection to the scene and are

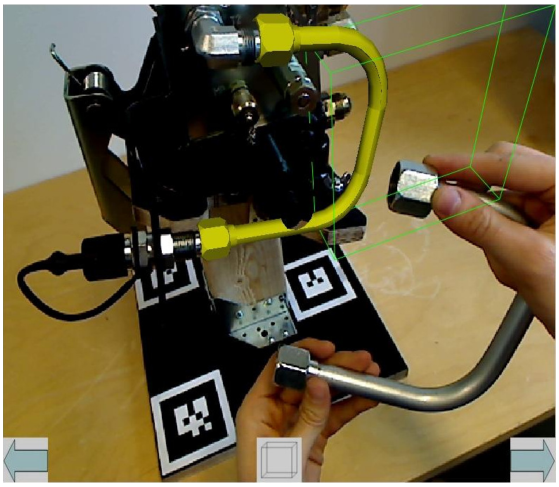


Figure 3.8: Instruction to place a tube at a position indicated by a virtual model (source: Sääski et al. [252])

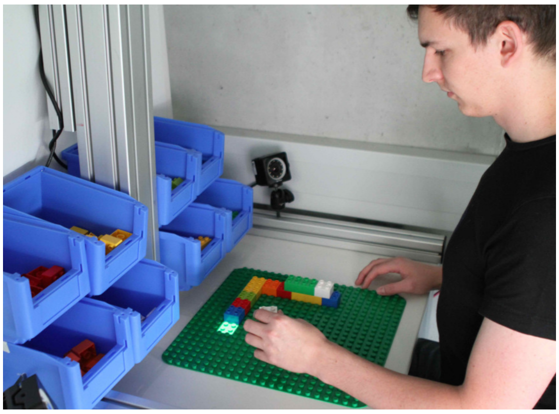


Figure 3.9: Instruction to set a toy block (source: Funk et al. [117])

classified as detached virtual information objects.

Figure 3.9 shows instructions to set a toy block on a base plate, which are shown via a projected augmented reality display [117]. The ground plate and the brick, which has to be set, are direct physical information objects. Whether other blocks also have to be classified as information objects depends on the usage situation, which is not fully described in the publication. If they are relevant for the current work step, they are included as direct physical information objects, but otherwise they are not information objects. However, this view gives no reason to assume they are, since it is not necessary or helpful to know their positions in order to set the current block. So they are instead a part of the environment. The only virtual information object is the projection of the target position for the current block. Since it cannot be moved, rotated or resized without losing the spatial context, it is a spatial virtual information object.

As last example, figure 3.10 shows an instruction to remove a panel inside a

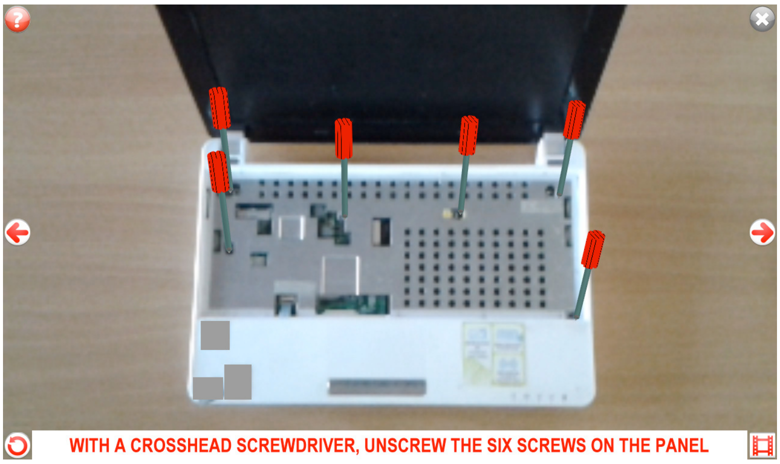


Figure 3.10: Instruction on how to remove a panel from a laptop (source: Sanna et al. [255])

laptop [255]. The interface is displayed on a video display placed at a fixed position outside the line of sight. The screws to be loosened and possibly the metal sheet, depending on the specific procedure, are indirect physical information objects because these are the parts that must be interacted with in the step shown. The physical parts, of which they are reproductions of, are thus classified as direct physical information objects. It does not appear as if the other parts of the laptop are relevant for this work step and therefore they are part of the environment. The screwdrivers are spatial virtual information objects. Their spatial context is defined through the way they indicate how a physical screwdriver must be used to unscrew the screws. As in previous examples, the textual description of the work step is a detached virtual information object. The other virtual elements do not have any spatial connection and thus are classified as detached virtual information objects.

3.4 Characteristics of Information Objects

The way information objects are modeled in the IRMAR framework does not directly lead to the identification of representation challenges. This is because their definition alone does not provide sufficient insight into their nature. What is missing is a consideration of aspects like, e.g., the type of information they carry, how they interact with each other or with other parts of a scene or how they behave as part of an augmented reality system. Thus, it is first necessary to examine their characteristics in this regard in detail. This involves identifying those which, alone or in combination with others, can lead to potential sources of unnecessary sensory or cognitive effort. The starting point is the definition of information objects as information carriers for procedural tasks with anchors and a graded spatial connection, which are embedded in the physical world.

In order to be usable, the characteristics must be described in such a way that the challenges can be derived from them in the later course of the work. One important aspect in this regard is, to provide a sufficiently general description of them that is not only valid a few individual use cases, but allows the challenges to be described with a reasonable degree of general validity. At the same time, however, they must be specific enough to allow the derivation of concrete challenges. For example, the characteristic that virtual information objects are graphics generated by an augmented reality system would certainly be correct and sufficiently general. However, its usefulness for deriving the challenges is questionable due to lack of specificity. Likewise, characteristics may not be concerned with specific technical features of individual systems or technologies, such as tracking precision. This would also limit their usefulness in deriving challenges due to limited generalizability.

Five such characteristics of information objects are identified in the following: spatial relationship, connectedness, discrete and continuous change, manipulability and physical change. While some aspects of them have been previously described by other authors like, e.g., the characteristic spatial relationship is addressed by Neumann and Majoros [205], the here given derivation and explanation are novel. Only those characteristics, which are independent of a concrete technical implementation and contribute to the development of the representation challenges in the next chapter, are included. It should be noted that a concrete assessment of the relevance of a characteristic can only be made afterwards, when it is used to derive a challenge, which is also the basis for this selection. This collection is also not to be understood as comprehensive or complete. In fact, it is very likely that further characteristics can be found that possibly lead to further challenges. Nevertheless, only the characteristics mentioned will be considered here, as a search for more would have exceeded the scope of this work.

3.4.1 Spatial Relationship

Information objects are, through a spatial relationship, connected with and localized in the physical world. This has already been considered for their definition via the mounting property. Since this characteristic applies to all information objects, it also implies that they are in a spatial relationship with each other, allowing them to be in front, behind, left, right etc. of each other. This is especially relevant for physical and virtual information objects, whose spatial relationship is only made possible by augmented reality. However, the form of the spatial relationship is different for the individual classes of information objects. For direct physical information objects, this is trivial, as they are an inherent part of the physical world, with or without augmented reality. Indirect physical information objects are either reproductions of direct physical information objects or parts of the environment, which may be true to original or altered. As a result, they always have a naturally predefined position, which does not have to be assigned to them. For virtual information objects, this characteristic follows directly from

Azuma's definition that they must be located and registered in 3D space [23]. In short, "virtual objects have locations" (Neumann and Majoros [205], 1998, page 3). Regarding spatial virtual information objects, this can and in many cases will be a meaningful spatial context but it does not have to be. Instead, it can also be positioned at a different position or with a different orientation, which leads to a spatial relationship where the meaningful context is lost. Spatially referenced virtual information objects have one or more anchors where the connection must be made apparent to a user. Irrespective of whether this is done through a spatial assignment or through another form of presentation like connection lines, the information objects must be positioned at some locations. Even for detached virtual information objects, which are not spatially connected by definition, it is impossible to not have this characteristic. They have a spatial relationship simply by being positioned at some place.

What also differs in regard to the individual classes of information objects is the significance of the spatial relationship. For example, the positioning of an indirect physical information object may be based on a physical object and thus by mirroring its position also provides information on this object. For detached virtual objects this is completely different. The positioning has no significance besides supporting user interaction. For spatial and spatially referenced virtual objects the position is significant in terms of their meaningful spatial context or anchor. Especially for spatial virtual information objects this means that information is also immediately conveyed to the user from the spatial relationship. While it may seem like this, the characteristic spatial relationship does not imply any specific reference systems. Instead, two or more information objects can be located in the same one as well as in different ones.

As already described in the section 2.1 of the fundamentals chapter, embedding virtual information into the physical world and, thus, creating a spatial relationship is at the core of augmented reality. Without it, there is no augmented reality in the way it is considered here. The definition by Azuma [23] directly incorporates this by defining that for augmented reality, objects must be registered in 3D. But this characteristic is also found in previously published modeling approaches. Tönnis et al. [287] use two dimensions in their taxonomy to include this characteristic: mounting is used to classify the reference system that a virtual object is located in, i.e., what it is attached to, while registration addresses the actual position that is assigned to it. Wither et al. [306] defined an annotation, i.e., a spatial or spatially referenced virtual information object, as a virtual object that must have a spatially dependent component specifying the relationship of the annotation to the physical world. Through their dimension location complexity, they also define different levels of complexity that this relationship can assume ranging from low to high. This relationship is also considered in more detail in the author's earlier work. For this purpose, two types of reference systems for information objects are defined which an be applied independently to position and orientation of an object [200]. These are world or spectator coordinate systems (WCS resp. SCS) that define which type of references an object's mounting can

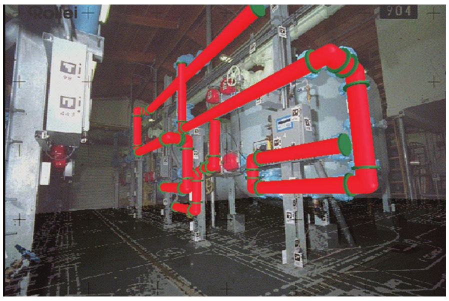


Figure 3.11: Integration of virtual pipes into a physical machine as a form of spatial relationship (source: Navab [204])

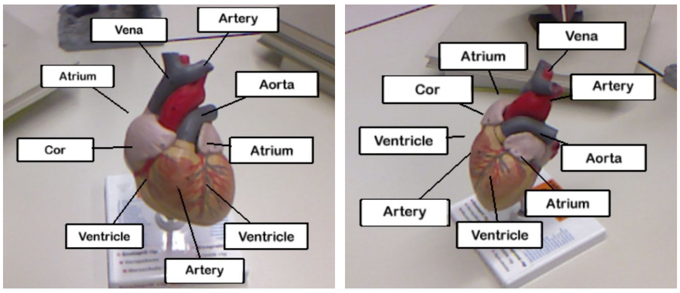


Figure 3.12: Movement of a physical information objects leads to a reorganization of the labels and thus creates a different spatial relationship (source: Madsen et al. [191])

be bound to.

As an example, figure 3.11 shows an integration of virtual pipes into a physical machine, which places them in a clear spatial relationship with the physical world. The pipes are in front of the machine and from the viewpoint, parts of the virtual pipe are at the front while others are further back. A spatial relationship of virtual and physical information objects is also created. The virtual pipes are aligned in such a way that they are placed in a meaningful spatial context. However, it would also be possible to position them somewhere else, which would result in a different spatial relationship with a changed significance. An additional example are the labels in figure 3.12. They are placed in a spatial relationship with each other and with the heart. In this example, the spatial relationship is not primarily used to convey information for a user, which is done via the black connection lines. Instead, the positioning supports the assignment so that the labels are close to their anchors and their layout follows the arrangement of their anchors as seen from the viewpoint. Thus, moving the heart model causes the labels to be reorganized and change their spatial relationship.

3.4.2 Connectedness

Besides the discussed spatial relationship, information objects can also have semantic connections with each other. For instance, this is the case when a label superimposed via augmented reality gives information on how to use a physical button. It can apply to combinations of information objects from either the same class or from different classes [200]. For the definition of information objects, this has already been considered by the possibility of objects having one or more anchors. These can be used to model the semantic connections that one information object has with others. However, not every anchor is an information object, and instead an anchor can also be a point or an area from the physical world. Often, such connections can be found between information objects of different classes when they offer different forms of information on the same matter [306]. Since augmented reality offers the possibility to extend the physical world with virtual information, this applies especially to combinations of virtual and physical information objects.

This characteristic manifests itself in different ways for the individual classes of information objects. Indirect physical information objects are reproductions of a part of the physical world, therefore, if they are created based on a direct physical information object, there is also a semantic connection to that object. Spatial and spatially referenced virtual information objects can also have connections to direct or indirect physical information objects. These are annotations, where the virtual object adds information to the physical object. However, there can also be connections between information objects from the same class. For example, the magic lens [183] adds spatial virtual information objects as annotations to information objects from the same class. Detached virtual information objects only exhibit this characteristic to a very limited extend, because they do not have any connection into the scene, and the only anchors left could be from objects of the same class. Direct physical information objects do not have any anchors themselves but can be an anchor for other information objects.

This characteristic has already been considered by other authors, but only limited to virtual annotations of physical objects. The taxonomy designed by Wither et al. [306] focuses on this relationship between virtual and physical objects. In their modeling of augmented reality annotations, they include the *spatially dependent and spatially independent components* to capture this characteristic. Similarly, Tönnes et al. [287] include this characteristic with the dimension *type of reference* in their taxonomy. The practical relevance of this characteristic is demonstrated by the fact that annotations in augmented reality are often used to support maintenance, inspection, design and assembly tasks [306, 192]. In general, they are a common means of supporting procedural tasks with augmented reality [192].

The examples previously used for spatial relationship can also serve as examples for connectedness. The labels in figure 3.12 show the names of different parts of the heart to which the connection is visualized by the black lines. Thus, each line

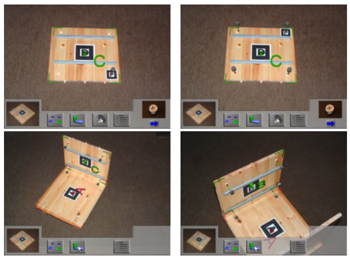
Figure 3.13: A physical information objects and its reproduction on a screen share a semantic connection (source: Webel et al. [302])

indicates such a connection. The connectedness in figure 3.11 is more difficult to model, since there is no obvious graphical connection between two information objects. However, since the virtual pipes are an extension to the physical machine, they also establish a connection. Thus, an indirect physical information object becomes an anchor for a spatial virtual information object. The connection is represented by positioning the virtual pipes in a meaningful spatial context. If the single segments of the virtual pipe are treated as separate information objects, then each of them has its own connection to the respective part of the machine. Another example can be found in figure 3.13. Here, a direct physical information object, the machine, is reproduced on a screen as an indirect physical information object. Thus, these two objects are semantically connected even though there is no visualization or spatial context to this.

3.4.3 Discrete and Continuous Change

Displayed information objects will continuously or discretely change. Existing ones may move, change or disappear and new ones may appear. Continuous means here that a change takes place, but the object is still present. Discrete, in contrast, describes a change in which there is no continuation for an object, i.e., it begins or ceases to exist. The primary reason for this characteristic is the fragmentation of procedural tasks into work steps, each of which requires different information objects [274]. When one is completed, the no-longer needed information objects are removed; those that are still required are possibly rearranged and new ones are added for the following step. This results in temporal sequencing of the displayed information objects leading to transitions between different sets of them with different spatial positions. Whether these changes happen abruptly or animatedly is not relevant for this characteristic. However, changes in the displayed information objects can also occur within a work step. For example, a movement of physical objects relative to the viewpoint (or vice versa) may make it necessary to rearrange the virtual information objects [21, 191]. Availability of new information such as sensor readings may also cause the inclusion of new ones

(a) (source: Encarnação and Stricker [94])



(b) (source: Zauner et al. [314])

Figure 3.14: Two examples for procedural tasks with work steps which each contain a specific set of information objects

during a work step.

This characteristic does not imply that there are no information objects that can exist throughout the duration of a complete procedural task. Other than in simple or unusual cases, however, the composition of the information objects will change over the course of such a task. Direct physical information objects do not have this characteristic because physical laws prevent them from randomly vanishing, appearing or instantaneously changing their appearance or place.

Again, this characteristic of change has been previously described by other authors. With regard to the life span of virtual objects, Tönnis at al. [287] call such behavior *discrete temporality*, while Wither et al. [306] call it *temporal annotation permanence*. The movement of information objects is covered by Wither et al. through the dimension *location movement*, and is further divided into different classes regarding their freedom. Tönnis et al., on the other hand, include only the mounting dimension, which only partially corresponds to this characteristic. Unlike the two taxonomies, there is no division into the two aspects object lifetime and movement in this characteristic but instead it is treated as one because the focus is on the changeability of the information objects and not the influencing factors that lead to it. It is especially found in augmented reality systems supporting construction and assembly tasks [306].

Two examples for a change of information objects between work steps can be found in figure 3.14. The figure shows two sequences of work steps, each with its distinct set of information objects. If one step is completed and the next step begins, at least some previously-shown information objects disappear and others are added. Whether the change is animated or not cannot be determined based on the available information. However, as explained above, this does not affect discrete and continuous change. Instead, an animation would be a measure to a challenge resulting from it. Another example was shown previously in figure 3.12. When the heart model is turned or the viewpoint moves around it, the labels are reorganized to maintain an organized layout. Depending on the application and the arrangement of the labels, this may happen instantaneously or with an animation.

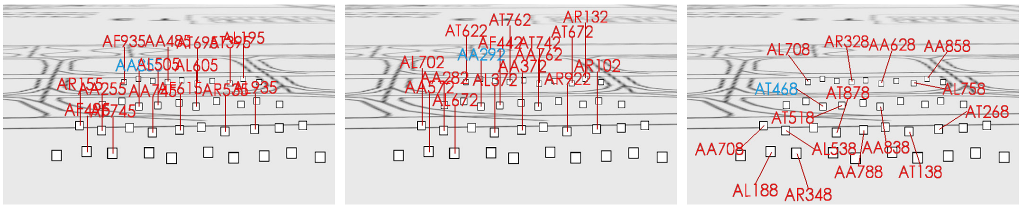


Figure 3.15: Different arrangements for labels which are possible without loss of information due to them being spatially referenced virtual information objects (source: Peterson et al. [221])

3.4.4 Manipulability

The representation of information objects from different classes of spatial connection can be manipulated by an augmented reality system to a varying degree. With decreasing spatial connection depending on the class, the manipulability by an augmented reality system simultaneously increases. For direct physical information objects this is obvious. Changing their representation is impossible in an augmented reality system. A user sees them in a direct way without the light arriving at an eye ever being processed or modified by an augmented reality system. For indirect physical information objects, this is different. To create them, a physical object has to be recorded, processed and then reproduced, making it is possible to change how and where the object is displayed during this processing chain. However, these information objects always depend directly on the physical object being recorded. Spatial virtual information objects are already more easily manipulable. They do not directly depend on physical objects and all their information is completely available without a physical object as a source. Nevertheless, when they are removed from their meaningful spatial context, information is lost for a user. On the contrary, spatially referenced information objects do not contain such a spatial context. Instead, only the connection to the anchor must be maintained such that no information gets lost. Otherwise, it can be manipulated by being e.g., moved, scaled or placed in an arbitrary reference system, without any loss. Detached virtual information objects do not have an anchor to which a connection must be maintained, so manipulation is possible without the restriction of a spatial reference. To the best of the author's knowledge, this characteristic has not been described previously.

An example for manipulability is visual clutter management for augmented reality labels as shown in figure 3.15. The methods typically used are based on the fact that labels, i.e., spatially referenced virtual information objects, can be moved without destroying a spatial context. They make use of the freedom they receive through the form of the spatial connection of labels. For spatial virtual or indirect physical information objects, as shown in figure 3.13 or figure 3.11, such methods could not be applied without losing information on a spatial context. In figure 3.3, it is shown how an indirect physical information object is changed compared to the direct physical information object it is a reproduction of.

3.4.5 Physical Change

To complete a procedural task, a person has to interact with and often also manipulate physical objects [248, 90]. In many cases, these must be moved so that their spatial position changes. An augmented reality system supporting procedural tasks provides necessary or helpful information to conduct these changes. This is conveyed via information objects, whereby at least some of these carry information on how their anchor must be manipulated. This means that, speaking in IRMAR terminology, many information objects describe movements of their physical anchors. This can be accomplished through, e.g., outlining what needs to be changed or by showing the final state as part of the information they convey. In any case, they contain the information on physical changes before these have been made.

For many detached virtual information objects, this characteristic does not apply because they contain other information or describe changes of the physical world only on a very general level that is not related to specific objects. Direct physical information objects often only provide information about themselves and not about other objects. There are exceptions, of course, such as when work instructions are physically printed on assembly parts. Objects of the other three classes can have this characteristic, but do not necessarily need to. For example, an information object might even indicate the opposite, that a certain manipulation must not be carried out. For spatial virtual information objects, the information on a needed change can also be part of the meaningful spatial context.

This characteristic of physical change is implicitly found in in the work of various authors where the basic task involves manipulation of physical objects. For example, Tang et al. [281], Wang et al. [299] and Feiner et al. [99] all consider instructions which include modification of physical objects. Röltgen and Dumitrescu[250] come very close to describing this characteristic in their taxonomy for industrial augmented reality use cases with the dimension effect level. With it they distinguish what the action supported by an augmented reality application affects, the possible classes being real world and virtual world. For any application that falls into the real world class, the information objects used convey the information necessary for this manipulation. However, due to their different focus, they do not directly apply this to the individual objects an augmented reality scene. Relatively close, though not on a conceptual but on a practical level, are also Henderson and Feiner [137], who investigated supporting the psychomotor phase of procedural tasks with augmented reality. To do this, they created virtual elements that were specifically designed to convey movement information. Similarly close are Baldassi et al. [28], who investigated if animations in augmented reality are a viable way to display information on necessary physical movements.

One example can be be found in figure 3.16, where the blue arrow is a spatial virtual information object that gives the instruction to rotate its anchor, the combustion chamber, which is a direct physical information object. Further examples can be found in figure 3.14. In the two subfigures, the movement of spatial virtual

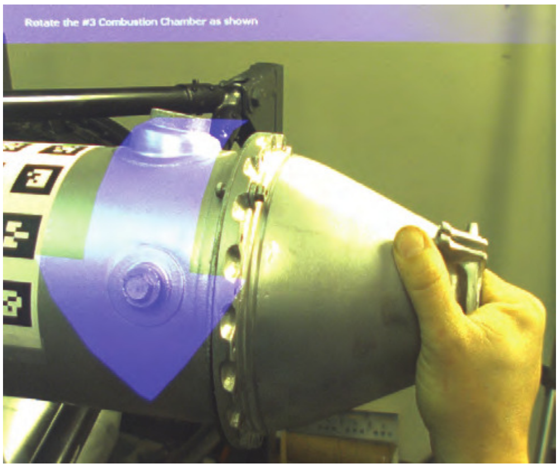


Figure 3.16: Instruction to manipulate a physical object shown with a spatial virtual information object (source: Henderson and Feiner [137])

information objects is used to indicate which physical parts need to be removed and how this movement should take place.

3.5 Characteristics of a Combined View

The information objects and the environment from one scene are incorporated into a combined view, which is seen from the viewpoint. Like the individual information objects, this superposition of parts from a scene also has certain characteristics that alone, or in combination with others, can lead to potential sources of unnecessary sensory or cognitive effort. Therefore, those characteristics that influence how the information represented in a combined view and do not apply to single information objects are included here.

Three such characteristics of a combined view are identified in the following: combination, reference systems and fluctuation. As with the properties of the individual information objects, some aspects have already been described in previous publications by other authors. For example, Azuma [23] already describes the need to combine virtual objects and the physical world into one view as part of his definition of augmented reality, which is an important aspect of combination. However, their derivation and explanation as given here are novel. The further restrictions already listed for the characteristics of the individual objects also apply. This means that only characteristics independent of a concrete technical implementation that contribute to the development of the representation challenges are included. Again, also this collection cannot be comprehensive or complete. Instead, it is very likely that further characteristics may be found that contribute to finding further challenges. However, as a search for more would have exceeded the scope of this work, only the three mentioned ones are considered here.

3.5.1 Combination

In a combined view, the way information objects and the environment are seen from the viewpoint depends on many details and is not always optimal. It can also be very volatile and difficult to predict, as even seemingly small changes of the scene, including environment and information objects, the viewpoint or the view frustum can have a great impact on it. Information objects are, in many cases, not clearly separated and do not stand out from the background. Instead, in a combined view they overlap and blend with each other or with the environment. They may also obscure each other, influencing in how well they can be seen and how much of them is visible.

This behavior is caused by two two reasons in particular. First, when supporting procedural tasks like repair or maintenance with augmented reality, the user's viewpoint can only be controlled to a limited extent by an augmented reality system. Instead, users control from where they look at the scene and select the point of view based on the current activities necessary for the work step [23]. Therefore, the way information objects and the environment are combined is only partially predictable and controllable. Secondly, an environment cannot be sufficiently controlled or be known in advance, except for some cases such as as assembly workstations in factories. This is different from traditional computer interfaces or virtual reality where everything is controllable by software and every part is known perfectly. For augmented reality, a slight change to either the environment, one of the information objects or the viewpoint may lead to a significant change in the combined view.

This characteristic already appears in the definition by Azuma [23], according to which virtual and physical world must be combined in 3D to create augmented reality. While the characteristic spatial relationship is also based on this definition, there are major differences. It is focused on the idea that information objects are spatially related to a physical 3D world and, consequently, one object is also spatially related to others. The characteristic, therefore, always refers to the individual object. Combination, in contrast, is that in a combined view, information objects influence each other's perception. The individual object is only one of many in this regard. Further, this characteristic is limited to a view frustum outside of which only direct physical information objects can be seen. It may span the whole field of view of a user, but current technologies only allow augmentation in a much smaller area [42].

Effects resulting from this characteristic have been widely studied and reported in the literature. Information objects, for example, may overlap from one viewpoint and, thus, occlude each other [181]. Figure 3.15 shows this very well. Each label on its own would be easily legible, but their combination into one combined view creates visual clutter. However, in the same scene, there may also be perspectives from which a view without the overlap is possible. An example for a combination of environment and information objects with different text drawing styles and backgrounds is shown in figure 3.17. While a text with one drawing

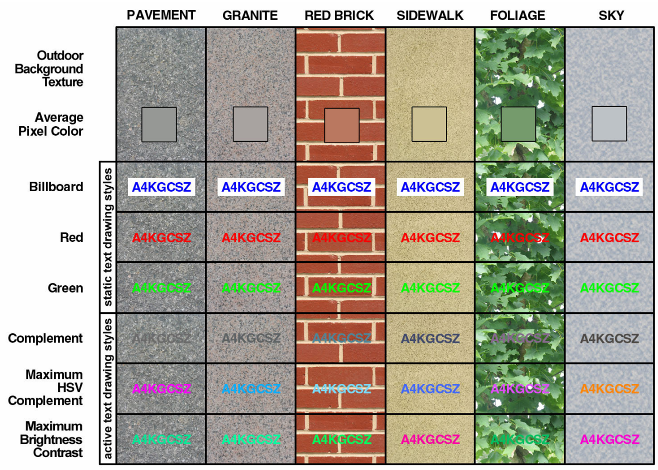


Figure 3.17: The combination of background and text drawing style significantly influences text legibility (source: Gabbard et al. [120])

style may be easy to read on one background, its legibility may be considerably impaired against a different background [122]. But also colors are affected by this effect. For example, the perceived color of an information object may change due to a color change in the background [179]. A combination into one view also implies that when one information object stands out more, e.g., because it is closer to the viewpoint or has an increased in luminance, it may easily diminish other information objects. A poor combination of information objects and environment can even lead to an augmented reality system being no longer usable for its intended use case [168].

3.5.2 Reference Systems

When multiple information objects are combined into one scene, they may be located and oriented in different reference systems, meaning that they can move independently from each other, be fixed to each other or show any behavior in between. Examples for sources of such reference systems are the screen plane of a handheld device, a user's head or a physical object somewhere in a scene. For physical objects, a common reference system is typically easy to see because it always requires a way to transmit a force, i.e., there must be a physical connection. For example, it is clear that a part mounted on a robot arm also moves together with it. For other types of information objects, however, this connection does not have to be visible since virtual objects are not moved by force transmission but by rules defined in software. In addition, such virtual reference systems can be non-linear and thus much more complex to comprehend than physical ones [287]. Placing an object in a reference system can be done independently from the anchor. A spatially referenced virtual information object, for example, could be placed on a screen plane of a handheld device, but have an anchor in the physical world. This was already shown in a previous example in figure 3.12. When the heart is rotated, the labels do not turn together with it but instead stay on the screen plane.

An information object can also be positioned in one reference system while it is oriented in another one. Two classes of reference systems can be distinguished: world coordinate systems (WCS) and spectator coordinate systems (SCS) [200]. The first class includes the reference systems that are not user oriented, meaning that the position or orientation of objects in these reference systems stays the same relative to the scene when the viewpoint is moved. Examples are global reference systems that refer to the scene, or reference systems that are aligned with objects in the scene. The second class includes reference systems that are related to the user and have, for example, her or his body or a device attached to it as reference. Billboards are an example for information objects that are WCS located but SCS oriented. They stay at a fixed position in a scene but are always oriented towards the user's viewpoint. When a spatial virtual information object is placed in a meaningful spatial context, it is WCS positioned and WCS orientated in the scene [200]. However, not every object that is WCS positioned and oriented is also placed in a meaningful spatial context.

The combined view is strongly influenced by these different reference systems because they have the already-mentioned bindings to different origins. For example, a reference system can be directly bound to the viewpoint so that a change of its orientation is sufficient to cause a different superposition of information objects with each other and background. In the same way, the SCS orientation of information objects has the effect that every time the position of the viewpoint changes, the object's orientation changes as well. Thus, the combined view depends on how the individual reference systems behave based on the view position with respect to both positioning and orientation.

In their taxonomy on annotations in outdoor augmented reality Wither et al. [306], include a dimension *location movement* which is similar to this characteristic. They use it to assess how freely the spatially independent component of an annotation can move relative to the spatially dependent component. However, they do not generally cover reference systems. In contrast, Tönnis et al. [287] include them in their taxonomy with the dimension *mounting*, which they divided into the types of reference systems: world, environment and human. All of these could be found in applications they scanned, though the environment was the most used binding for reference systems.

One example for reference systems can be found in figure 3.18. The colored stripes are mounted to the user's head while the sheep is mounted to the hand. Unfortunately, examples in pictures can only provide a limited impression, since relative motion of user, information object and scene all need actual movement to be seen.

Figure 3.18: Two reference systems: the colored stripes are head mounted while the sheep is hand mounted (source: MacWilliams et al. [190])

3.5.3 Fluctuation

The characteristic fluctuation comprises that when merging virtual and physical information objects with the environment into a combined view, hard-to-control fluctuations, i.e., non-perfect overlays of physical world and virtual objects, can occur [20]. These are caused by problems to correctly align different reference systems with each other at a given moment. An example are the so called registration errors, which occur when the detection and tracking of the position in the physical world is faulty. This results in an incorrect estimation of the viewpoint in relation to the physical world. Thus, an incorrect mapping of virtual coordinate systems to the physical world is made [141]. In effect, virtual information objects placed in a reference system bound to the physical world or an object in it are not displayed at the intended position [245]. However, causes for fluctuation are not limited to tracking, but extend to all factors that influence the alignment of virtual and physical reference systems. Examples are insufficient calibration of optical see-through glasses [128], faulty eye tracking [128], large latencies or low update rates [92].

In unprepared scenes, it is not possible to exactly know all the parts in advance [158]. Due to ambiguities, incorrect assumptions on position and orientation of the viewpoint in relation to physical objects can be made based on the collected sensor data, leading to tracking errors and making perfect mapping impossible. Whether this problem can be solved in practice to such an extent that this characteristic becomes irrelevant remains to be seen. However, for the near- and medium-term future, ubiquitous tracking that would allow to perfectly align reference systems is not to be expected [158].

Vincent et al. [293] model these alignment problems by defining two spatial mappings that connect the physical world with representation and the representation with the augmentation. If needed, these spatial mappings can be relaxed, meaning that a less-precise way of aligning physical world, its representation and the augmentation with each other is chosen. One important factor for them to select a more relaxed mapping is a detected decrease in tracking quality.

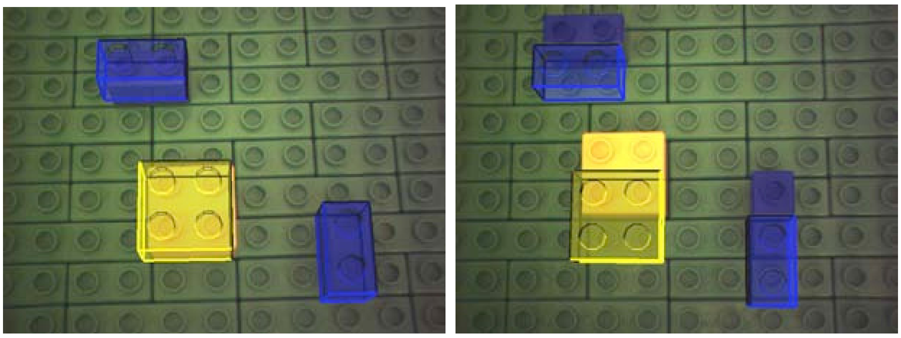


Figure 3.19: A combined view free of fluctuation (left side) and with fluctuation (right side) (source: Robertson et al. [245])

An example for fluctuation can be found in figure 3.19. It shows the same scene, where a toy brick must be placed, two times. While the left subfigure is free of visible fluctuation and the virtual bricks are perfectly aligned with their intended positions, the right subfigure shows significant fluctuation and all virtual bricks a shifted towards the bottom.

3.6 Comparison with other Modeling Approaches

As already discussed in the fundamentals chapter, there have been previous approaches to conceptualize and model information represented with augmented reality. In the following, three relevant approaches are compared in more detail with the approach chosen for this thesis. In particular, their differences to IRMAR are examined in detail, which of their individual concepts are included in IRMAR and why some concepts from these approaches are not included.

The selected approaches are the two taxonomies by Wither et al. [306] and Tönnis et al. [287], because both are approaches to conceptualize the topic of information representation in augmented reality. The third is the model by Vincent et al. [293, 294], because they created a modeling approach, albeit for a different use case, that combines the physical world, a representation of the physical world and virtual objects. The author is not aware of any other concepts related to augmented reality that have such similarity to IRMAR. These three approaches have already been discussed in the fundamentals chapter. While the focus there was on showing which approaches exist and which new developments are still needed, here their relationship to IRMAR is discussed.

The taxonomy for annotations in outdoor augmented reality by Wither et al. [306] is specifically designed to analyse existing augmented reality systems. For this purpose, Wither et al. developed six evaluation dimensions that can cover a wide range of annotations in augmented reality. In addition, they provide a simple and very comprehensive definition of an augmented reality annotation that shares similarities with the definition of information objects. However, by their limitation to annotations, they include neither physical objects nor other virtual objects which do not annotate physical objects. Systems supporting procedural tasks are included as well but these are not the focus and instead one of many system types.

Wither et al. define "an augmented reality annotation [as] virtual information that describes in some way, and is registered to, an existing object" (Wither et al. [306], 2009, page 2). This corresponds to what is described in this work as spatial or spatially referenced virtual information objects with a physical information object as an anchor. The visualization of an anchor thus corresponds to the annotation's spatially dependent component, while the information object corresponds to the spatially independent component. Since other types of information objects like physical objects must also be included in the IRMAR framework, their definition of an annotation cannot be adopted directly for information objects. Many of these may not even have something comparable to a spatially dependent and spatially independent component. This also motivates the much more open definition of anchors compared to the spatially dependent component. Worth mentioning in this context is a similar concept, proposed by Webel et al. [303], which is called pointer. It is defined as a highlight of an anchor which connects visually represented information with a physical location. Here, too, the same reasons apply why the more open concept of anchors is used for IRMAR.

In IRMAR, characteristics are identified and considered while Wither et al. [306] use dimensions. The reason for this difference are the distinct usage scenarios. Dimensions with a division into individual classes are used to find differences and support an analysis. This also explains why no physical objects are considered in their taxonomy. These would simply not show the required diversity to make a meaningful separation into classes possible. With the characteristics as used here, it is different. They are the same for different use cases and augmented reality systems and do not show differences but similarities. This supports investigating sources of unnecessary sensory and cognitive effort in the further course of this work because this way it is possible to work on the basis of defined general characteristics. Otherwise, the issue would arise of how to realize and define challenges in a sufficiently general way. Nevertheless, there is also a certain similarity between dimensions and characteristics. This is due to the fact that dimensions also must have a certain generality in order to be applicable a diverse set of systems. Therefore, it is only logical that they are based on characteristics which can be effectively divided into classes with significant differences.

Overall, Wither et al.'s taxonomy contains six dimensions. While some of them have have correspondences to the here identified characteristics or other parts of the IRMAR framework, this is not the case for every of them. One dimension with a correspondence is *location complexity*, which describes how complex a real-world location is to which an annotation refers. This also includes in how far an annotation's pose is determined by its location, e.g., by the annotated location mandating an orientation of the annotation. At first glance, this may seem similar to the classification of information objects by spatial connection, specifically to the distinction between spatial and spatially referenced virtual information objects. However, location complexity describes an actual use in an augmented reality application while the distinction between spatial and spatially referenced information objects is a difference in the potential to convey information through a spatial context. Instead, it is more comparable to mounting. A simple text, for example, may have a high location complexity but not be a spatial virtual information object. Concrete relationships to points, areas or information objects are, instead, modeled via the anchor property in IRMAR, allowing for a separate consideration of what an object refers too, what information can be conveyed through their spatial context and what their actual pose is.

Wither et al.'s dimension *location movement* describes the behavior of locations and their associated reference systems to move relative to each other. There is no equivalent single characteristic in IRMAR. Instead, the movement of locations is determined by a combination of discrete and continuous change and reference systems. As a basis for a source of unnecessary sensory and cognitive effort, it is discussed in more detail as part of the representation challenge orientation in the chapter 4 Representation Challenges. Another one of their described dimensions is *annotation permanence*. However, in the context of procedural task support, in many cases virtual information objects have a temporally controlled permanence. Except for direct and, in many cases, indirect physical information objects, they are not displayed anymore when a work step is completed and they are no longer needed in the next one. Hence, this dimension is reduced to one specific class. This is also determined by the characteristic discrete and continuous change. The other three dimensions *semantic relevance*, *content complexity* and *interactivity* are not relevant as characteristics. They either address details of the annotations' content or interaction, all of which are deliberately left out of the IRMAR framework.

Tönnis et al. [287] model information representation in augmented reality with a focus on classifying alternative ways to represent information. To achieve this, they define the five already-presented dimensions, each of which covers different aspects. Due to their focus and unlike Wither et al. [306], they are less concerned with the concrete structure of the objects, and instead concentrate on how virtual objects are integrated into an augmented reality scene. Like Wither et al., they do not consider physical objects but they do consider virtual objects that do not have a direct relation to physical objects. Their focus is also on analysis resulting in a necessary structure of the dimensions. The approach by Tönnis et al. covers almost the entire spectrum of augmented reality applications and does not take the specifics of procedural tasks into account.

Especially their dimension *dimensionality* is similar to the way information objects are classified in IRMAR based on their spatial relationship. However, there are some notable differences. Dimensionality only considers virtual objects, while information objects can also be of physical origin. This is, as already explained, not necessary to analyze the differences between existing systems, but for the examination of representation challenges it is. Furthermore, they distinguish only roughly between different gradations of relationship that objects can have to physical space. However, it is exactly this gradation that sets augmented reality apart from other ways to provide information for supporting the execution of procedural tasks. In this dimension, they only differentiate between the two classes 2D and 3D, which are similar to the classes spatial and spatially referenced virtual information objects in IRMAR, but not identical. For example, while a simple 3D icon with a reference to a physical location is in the 3D object class, it is not a spatial virtual information object. Instead it is a spatially referenced information object because it cannot have a meaningful spatial context. At the same time, a 2D map that indicates power cables hidden in a wall would be in the 2D objects class but a spatial virtual information object, since it can be superimposed onto the wall to create a meaningful spatial context.

Their dimension *mounting*, which is used to describe where an object is attached to and in which reference system it is located, is found in several places in the IRMAR framework. The most noticeable point is the mounting of an information object, which comprises exactly the information that is considered by Tönnis et al.'s dimension. Furthermore, comparable concepts can be found in the two characteristics spatial relationship and reference systems. The division into these two aspects is made in IRMAR because this finer subdivision helps in finding sources of unnecessary sensory and cognitive effort as used in the next chapter 4 Representation Challenges.

As was already the case with Wither et al.'s dimension annotation permanence, the dimension *temporality* by Tönnis et al., which classifies the lifetime of an object, is in many cases reduced to one class, called discrete. The reason is that with nearly each change of work steps, virtual objects are removed, and new ones are added. Only in some cases, does an information object exist throughout an entire procedural task. This is covered by the characteristic discrete and continuous change.

Another dimension is *viewpoint reference frame* which Tönnis et al. use to classify how the view of a user compares to that of the used rendering system. This is not directly modeled in the IRMAR framework and is only marginally considered through the viewpoint and the combined view. Similarly, their dimension *type of reference* is also not included in the IRMAR framework. The reason in both cases is that these dimensions respectively corresponding characteristics have not led to sources of unnecessary sensory or cognitive effort, respectively, representation challenges. However, it is also not excluded that a closer look will find a source after all, so this can only be seen as a momentary observation.

Further, the three categories by Vincent et al. [293, 294] show some similarities with the way information objects are classified in the IRMAR framework. Their category *real world* corresponds to a combination of environment and direct physical information objects. Similarly, their category *representation of the physical world* corresponds to a combination of a recorded environment with indirect physical information objects shown a video display. Lastly, their category *augmentation* corresponds to the spatial and spatially referenced virtual information objects.

Since their focus is interaction design, they do not define single objects with specific properties, but rather tread each category as not further divided. For this reason, they do also not model any relationship on a per object basis. Instead, through their spatial mappings, they describe how precise spatial overlays can be realized considering technical constraints and desired interaction. Which information can be conveyed in this way is not relevant for them. However, the division into single objects and identification of their properties is needed to sufficiently model information representation for procedural tasks. What remains, however, is the similarity to the fluctuation characteristic, as this covers the problem of correct mapping. A further similarity is that the physical and virtual world are combined in one model, although this is done in very different ways, based on the particular focus.

4 Representation Challenges[1]

In this chapter, the representation challenges are developed and presented based on the characteristics of information objects as identified in the previous chapter. In addition, their individual aspects are examined in more detail, previous descriptions of them are collected and measures for solving them are developed on the basis of relevant literature. The purpose is to achive the goal defined in the motivation, to find sources of unnecessary sensory and cognitive effort, describe them as challenges, and develop solutions for them.

As already explained, an augmented reality system for procedural task support must minimize sensory and cognitive effort that is generated by its use. Otherwise, completing tasks will take longer and be more error prone than necessary all while a user will experience increased cognitive load and faster fatigue [281, 89]. In cases where the felt additional effort exceeds the benefit of such a system, it will simply not be accepted by users. While the assumption that sensory and cognitive effort must be minimized is not true for all kinds of augmented reality systems, e.g., training systems [301], this goal can be assumed for procedural task support [281]. One contributing factor to such effort is information representation that makes it difficult for a user to see and understand the presented information. Therefore, the way information is represented must be designed such that as little effort as possible is generated on the user's side.

As an approach to this problem, the previously described characteristics of information objects and combined view are used to identify possible sources of unnecessary sensory and cognitive effort. For this purpose, combinations of these are formed and possible effects of their interaction on a user are analyzed. For example, the characteristics that information objects are connected with each other (connectedness), integrated with varying degrees in a spatial structure (spatial relationship) and that the registration of the virtual objects in the scene does not function perfectly (fluctuation) can be combined. This results in source of possible unnecessary sensory and cognitive effort when a user has to recognize the connections in a three-dimensional space while the correct overlay is not fully ensured. The information representation must, therefore, be designed in such a way that

[1] The description of the challenges is based on my previous publication *Challenges in representing information with augmented reality to support manual procedural tasks* [199]. The discussion of the measures for the challenges as well as the challenge motion cues have been added since then.

even considering this combination of characteristics, a user can easily recognize the connections. To facilitate use, such sources are formulated as representation challenges that have to be considered when designing augmented reality systems to support procedural tasks.

What distinguishes this approach from others is that it explains why certain challenges are caused by the use of augmented reality. They are, therefore, not random problems resulting from an inadequate design of user interfaces, but rather arise directly from the nature of augmented reality as a medium that connects virtual information and the physical world in real time. This insight is achieved by using characteristics for derivation and not relying on observations of existing augmented reality systems and published studies. Comparable approaches in the context of augmented reality are not known to the author. Hence, even if a challenge has already been previously described in literature, their derivation is novel in all cases.

In the context of this work, six different challenges could be identified following this approach. These are the clarity challenge, the consistency challenge, the visibility challenge, the orientation challenge, the motion cue challenge and the information linking challenge. As already mentioned, this collection is not complete, and, instead, through further combinations of characteristics or addition of new characteristics, it is possible that new ones are found. However, since six relevant challenges have already been defined, further focus was placed on describing them and identifying measures for solving them, thus stopping any additional search.

For each of the six found challenges, a description is given including what the challenge is, in which way unnecessary sensory or cognitive effort can be caused and how it is derived from the characteristics. Furthermore, it is stated whether a comparable challenge has already been formulated elsewhere in the literature. In order not to limit available references unnecessarily, not only literature on the support of procedural tasks is considered, but augmented reality in general. Lastly, the possible measures to meet the challenge as well as relevant influencing factors from the scientific literature are collected and discussed. Measures are always described with a perspective on information objects, even when this is not the case in the original publications. This is done to make it easier to use them directly on the basis of the IRMAR framework. Publications which do not contribute measures for the challenges but, instead, describe technological approaches for previously described solutions, e.g., algorithms that sort labels in a way that avoids overlapping, are left out. In some cases, however, this distinction is somewhat vague, so that also some more technical publications have found their way into the measures for the challenges. Additionally, there will likely be other publications that contribute measures to solve the challenges, which are not mentioned here. The reason for this is that augmented reality alone is already a very large and sometimes confusing field, and findings from other areas, to which there is no obvious connection, may also contribute. The goal with this structure is to give the reader an understanding of the challenges and to provide helpful information to solve them. Only for the information linking challenge no measures are given this chapter, as it is examined in more detail in the following one.

Some of the discussed measures go beyond addressing how information objects can be included in a scene such that the challenges are solved. This particularly applies to basic recognisability. Instead, the measures change the presentation of the information objects, e.g., by using specific combinations of font colors. The reason to also include such measures is that they also serve to minimize the additional sensory and cognitive effort caused by the way information objects are embedded in a scene. If this is achieved by changing the way an information object shown, it is a possible way to solve the challenge. However, no specific information visualizations are discussed, but instead, boundary conditions are outlined whose fulfillment simplifies solving a challenge. Such measures can be, for example, to visually separate information objects from each other and from a background by changing their color. Some measures also introduce new auxiliary elements such as shadows or direction arrows. These can be extensions or modifications of existing information objects where the presentation is changed to allow for an improved information representation in a scene. Examples are connection lines of information objects to their anchors or outlines that make texts more legible. However, these can also be new information objects that are included to represent information that that emerges from embedding information objects in the scene, such as arrows that point towards information objects when these are outside the field of view.

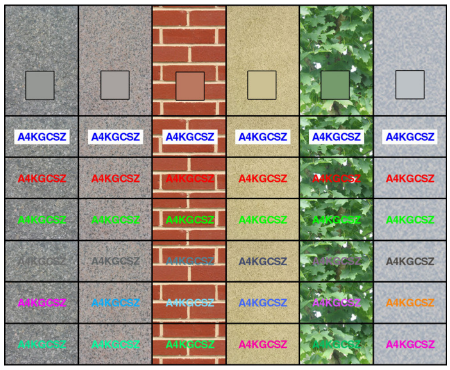
In many cases, challenges could also be solved by introducing interaction mechanisms that help a user to overcome them. An example would be to allow moving and resizing virtual information objects until a user considers the scene appropriate. However, this introduces additional effort which is not directed at solving the procedural task. Thus, interaction should be avoided by selecting an information representation that minimizes the needed interaction. While user interaction techniques are left out of this work, it does not mean that they are not an integral part of an augmented reality system to support procedural tasks. Instead, they should be used complementary with an appropriate information representation, but cannot replace it.

The measures for the challenges are not always without conflicts. On the contrary, solving one challenge may even aggravate another one to the point where it is impossible to solve. For this reason, the presentation of the challenges will be followed by a brief discussion of such conflicts.

4.1 Clarity Challenge

Irrespective of the concrete composition of a scene, the individual information objects must be as easily recognizable as possible. For example, texts must be legible and shapes must be shown in a way that a user is able to identify them. The challenge is, therefore, to integrate information objects into a scene in such a way

(a) Color distortion is a common problem when using optical see-through glasses (source: Itoh et al. [147])



(b) Text legibility depends on background and text drawing style (source: Gabbard et al. [120])

Figure 4.1: Examples of effects covered by the clarity challenge

that they are clearly and concisely represented, and it costs only minimal sensory and cognitive effort to see them and recognize the relevant information. This goes beyond the issues that exist for classical user interfaces, such as contrast, text display, etc. Objects with structures that are easy to recognize in conventional user interfaces can quickly become unrecognizable as part of an augmented reality scene. Reasons may be inappropriate positioning or an unfavorable environment. For example, the perception of colors or the legibility of text changes depending on the immediate background [97, 120]. Because the viewpoint may change based on a user's movements, it is important to ensure that the clarity challenge is not only solved from one specific perspective but from as many as possible. This challenge especially concerns human perception and is, in comparison to the following ones, the one that regards the sensory aspect most.

Only visible information objects are considered for this challenge, but not those which cannot be seen because they are outside the view frustum or occluded by other parts of a scene. The reason is that if an object is invisible to a user, the recognizability of its structures is not relevant. Instead, these cases are covered by other challenges that arise from a different combination of characteristics and stand for other sources of unnecessary effort. Consequently, other measures are needed to solve these challenges.

Two examples for effects that are covered by this challenge can be found in figure 4.1. Subfigure 4.1a shows how color distortion makes virtual information objects harder to recognize. With textured backgrounds this effect even increases. As shown in subfigure 4.1b, text legibility depends on the combination of color, text drawing style and background. From one viewpoint, text on an information object might be easily readable while from another viewpoint with a different background, this might become very hard.

The challenge is divided into three aspects, which are distinguished based on the type of interference that is addressed by the measures. Visual acuity covers what needs to be done to keep structures and textures of information objects discernable, color perception covers how colors can be used to keep information objects distinguishable and text legibility covers how legibility of text from information objects can be ensured.

Derivation. The source of possible unnecessary sensory and cognitive effort that is addressed with this challenge is caused by a combination of three characteristics. The first characteristic is spatial relationship. Due to the spatial arrangement of information objects described by this characteristic, it can happen that information objects have a position and orientation relative to the viewpoint that is unfavorable. For example, they can be positioned in such a way that they are very far away from the viewpoint and therefore appear very small, or at an unfavorable angle. The second characteristic is combination. It implies that when information objects and the environment are seen as the combined view, each information object is positioned in front of a background which may make it difficult to recognize. This background may be hard to control or predict by an augmented reality system. Having a user to change the viewpoint to create a preferable combined view should also be avoided because it creates effort on the user's side. The third characteristic is manipulability. It has the consequence that solving the clarity challenge is not trivial, because simply moving objects into preferable positions is not always possible or may have side effects. As described by this characteristic, changing an object's position or orientation may be very difficult to achieve or even impossible, may disturb its meaningful spatial context or move it away from its anchor.

Previously described by. This challenge has been previously described by Livingston et al. [179] for augmented reality systems in general. However, they do not consider it on a basis of objects that are integrated into the physical environment via augmented reality. Instead, they describe it as a general problem that occurs when using displays for augmented reality. Apart from this specific work, the investigation of perception in augmented reality has been a large focus of research [168]. Therefore, not only this one description but, instead, many aspects of this challenge have been described in many other contexts, even though it has nowhere been outlined as comprehensively as by Livingston et al.

4.1.1 Visual Acuity

Shape, contrast and spatial distance between information objects are all important factors on how well they can be seen. The limiting factor here is the visual acuity of the human eye. Structures whose angular size is too small cannot be individually distinguished, and, instead, blur into each other. Usually being able to resolve structures with the size of 1 arcmin is accepted as full vision while the maximum resolution of the human eye is up to 0.5 arcmin [96]. In the depth direction, humans achieve an acuity of about 2 arcmin when wearing binocular head-mounted displays [177]. However, humans already start making mistakes and take longer to recognize even much larger structures. Livingston et al. [178] reported significant improvements on optical see-though head-mounted displays up until at least sizes of three pixels, equaling 5 to 7 arcmin, under optimal conditions. When sizes became larger, they still found some improvements, but not to the same extent. However, because optical see-through displays do not block out the background, the needed minimal sizes might even be larger depending on the background as seen from the viewpoint. To support easy recognition, horizontal and vertical structures should be used, because they are more easily recognizable than slanted structures [176].

To achieve a full visual acuity, a high contrast between the lowest and highest luminance of a pattern is needed. A decreasing contrast leads to an approximately exponentially increasing recognition time and error rate [178]. However, with a contrast above 0.1, the recognition time and error rate are stable at a low level. While optical see-through systems have the problem that in bright environments, their luminance might not be enough to create a high contrast, video see-through systems may be even more problematic due to the limited contrast ratio that the used camera can provide [179].

Various measures can be used to represent an information object in a manner suitable for human visual acuity. However, their use is restricted by the object's class of spatial connection and the related limited manipulability including a potential loss of spatial context. Indirect physical objects could be enhanced, e.g., by increasing the contrast, by zooming in on them or by placing them in the center of a virtual fisheye lens. However, the author is not aware that this has ever been systematically investigated. Spatial virtual objects allow more manipulation and they could also be rotated, moved, resized, etc. Doing so would have the downside that information can be lost on spatial context. A possible solution would be to split the information objects into two. One could be used to present the spatial context while the other could be used to show the information object with the needed details. Again, the author is not aware that any of the previously mentioned measures has ever been systematically studied. Alternatively, so-called level of detail interfaces [83] or stylized depictions [107] can be used to help keep an object easily recognizable when its displayed at a smaller size. With these display forms, different presentations are defined for different angular sizes which offer different levels of detail. The larger an object is presented, the more detailed information is shown. However, these measures are already very much in the realm of visualization and can no longer be performed without concrete knowledge of how the object is supposed to be depicted. Virtually referenced information objects and detached information object can be manipulated without losing information from a spatial context. This allows to enforce a minimal angular size to keep their structures easily recognizable. Contrast can also be increased by modulating the features of an information object, e.g., by increasing pixel brightness, if the structure of an information object allows for this [193].

4.1.2 Color Perception

Human perception of color is highly volatile and depends on graphical context. The often changing superimposition of virtual graphics and physical world in augmented reality can lead to colors being perceived differently than in a purely physical surrounding [78]. It is non-trivial to predict the resulting color as seen by a user. The visual impression created in the eye depends on the color gamut and the luminance of the display device and the color and luminance of the background. Humans also do not directly associate an electromagnetic wavelength with a specific color. Instead, it is determined based on the graphical context [97]. This leads to the effect that different colors may be affected differently by the same background [121]. Darken and Lennerton [71] observed this effect on video see-through displays. They built a flight simulator which used video see-through head-mounted displays to combine a physical cockpit model with a virtual landscape. When evaluating the system, they found that color identification dropped from nearly 99% in a purely physical environment to about 62% when the colors were seen through a head-mounted video display.

Use of colors should thus be employed carefully, because based on background and illumination, they may no longer be perceived as their original color [78]. Livingston [176] was able to identify nine combinations of color and intensity that were relatively opaque and easily visible for multiple lighting conditions on optical see-through display. These are “three intensities of purple, two of blue, two of yellow, and two of green” (Livingston et al. [179], 2013, page 17). Outside of augmented reality, the colors blue, green, red, yellow, purple, orange, brown, pink are known to be maximally distinguishable colors and can be used when distinctive colors are needed [268]. However, Livingston [176] found that brown, orange and, to a smaller degree, red had a high likelihood of confusion in an augmented reality setting.

On optical see-through displays, the use of colors is further complicated by the missing ability to block out light coming from physical origins. This leads to the background shining through, especially when displaying darker colors. This can alter the perception of a color [179]. Livingston [176] conducted an experiment during which they let test subjects name colors shown on optical see-through displays in front of various backgrounds, and it turned out that the names they gave each color changed based on the background. One color, for example, was perceived to be brown in front of a dark surface, but perceived as orange on a white surface. This provides some key insights into the perception of color combinations on optical see-through displays. Luminance also has a high influence on how saturated colors are perceived; a bright background will likely lead to colors seeming desaturated and close to the white point, while a dark background will lead to colors seeming saturated [121]. Red tones are particularly affected by this. With an increased luminance, colors will seem less blue, likely because blue is perceived relatively poor compared to other colors, and thus the blue color portion fades more than others [121].

Even though the color combinations that a user perceives may be hard to control, knowledge of the background and the surrounding environment can help to mitigate this effect by selecting for maximum discriminability in colors [179]. Possible models to determine this are, for example, hue saturation value or maximum brightness contrast. Applying them on video displays is relatively simple due to the possibility to fully cover parts of the background and the availability of pixel-wise information on the shown colors. With optical see-through displays, this approach is also possible in principle, but will most likely be subject to a certain error based on their calibration. The color selection can be made automatically by a pre-distorting function that selects the colors used for information objects based on the background [179]. Of course, all this presupposes that the color of information objects can be adjusted without loss of essential information. When a display is used that replicates the scene, i.e., any kind of video display, the color of the environment and/or indirect physical information objects can be altered, for example, by lowering the saturation. This can be helpful,when the color of some or all information objects may not be changed. Further, the limited color gamut of currently available devices must be taken into account [147]. Due to this limitation, some colors might need to be altered such that information objects stand out from each other and from the environment [179].

Correct calibration can reduce problems with color perception [147], especially for video see-through displays which suffer color distortion from camera and video screen [179]. A further improvement may also be to adapt the color calibration for single users to their individual perception of colors [110]. However, this cannot solve problems with backgrounds shining through and context dependence of colors [97].

An extensive overview on color perception for augmented reality, especially for optical see-through displays, that goes beyond the scope of this work can be found in the work of Vincent et al. [179].

4.1.3 Text Legibility

Text is often used as a comprehensive way to provide information in augmented reality support for procedural tasks. It must be easy to read in order for a user to get this information with minimal effort. However, various factors in augmented reality can make text that is normally legible, difficult or even impossible to read.

Distortions created by the background behind a text are a significant problem for legibility, especially when the contrast is low [175]. For optical see-through displays, legibility is also influenced by the environment shining through [122, 78]. Video see-through displays seem to be more robust here because the background is visible around the letters but cannot be seen through them [78]. Background structures with a frequency of about 1.5 to 3.0 cycles per letter interfere with legibility the most [175].

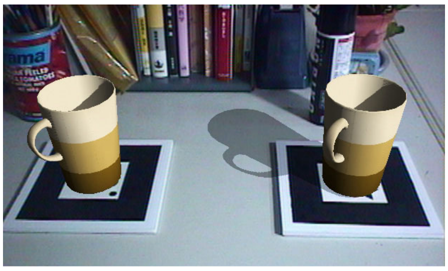
Several drawing styles have been suggested to increase legibility of texts on information objects. One that has been investigated by multiple authors is the billboard style [119, 122, 78, 104]. In this style, the text is backed with a plain colored area that covers the background behind the letters and prevents it from interfering with the legibility. This way, the error rate is comparably low while the reading speed high. Two contrasts are important for billboards: The first one is the contrast from the text to the billboard, and the other from the billboard to the direct environment, i.e., the background from a user's viewpoint [119]. Both should be as high as possible [78, 119]. Debernardis et al. [78] found that white text and blue billboard is a successful combination on optical see-through displays for many different backgrounds. For video see-through displays, they suggest using dark backgrounds because of automatic exposure and white balance settings of cameras that generally speaking cause images to be relatively bright. On the downside, billboards occlude larger parts of the scene which are behind the plain colored areas.

To avoid this problem, drawing styles with drop shadows or outlines can be used, which enclose the letters with a differently colored frame [119, 104]. They create a contrast to the background, but legibility does not reach the level of billboards. Because they do not have the size and the opacity of a billboard, their color selection must be done especially carefully. Gabbard et al. [119] suggest dealing with this limitation by using what they call active text drawing styles. For these, the background behind the text is analyzed and a contrasting color for the outline is computed at runtime. The best results concerning error rate and reading speed were achieved by using a maximum brightness contrast compared with the direct background.

Additionally, the color of the text itself has significant influence on legibility. In multiple studies, red text was found to have a bad legibility, especially in bright environments, while green text had a better legibility [122, 119, 104]. Simple blue text on optical see-through displays is also poorly legible, likely because human eyes have a low sensitivity for this color spectrum [78]. However, if a text has a drawing style that adds a frame around the font, such as a billboard or drop shadow, the influence of the color is reduced [119]. Dynamic drawing styles that account for the background to determine the font color, e.g., by displaying the text in complement colors, are not as legible as the previously presented drawing styles that add some sort of frame around the letters [120, 122].

In addition to drawing style and font color, the size is also critical for text legibility. Based on their empirical study, Renkewitz et al. [241] suggested using font sizes that appear to the user with an angular size between 0.92°and 1.68°. While fonts as small as 0.32°were legible, they found that only these larger sizes allow users to easily read texts without unnecessary effort. The large difference between the lower and upper number is caused by different combinations of color and background which they tested. However, no clear rules are given as to when which size applies. For spatially referenced and detached information objects, enforcing such size can be easily implemented. However, spatial virtual information objects must be moved or transformed for this, which may cause a loss of meaningful spatial context.

(a) The green square is supposed to be behind the wall but no visual cues support this (source: Furmanski et al. [118])



(b) Shadows can help to integrate a virtual information object into a scene (source: Sugano et al. [275])

Figure 4.2: Examples of a) an effect that is covered by consistency challenge and b) a measure for this challenge

4.2 Consistency Challenge

While the physical world is inherently consistent due to physical laws, this does not necessarily apply to an augmented reality scene. There are several factors that can disturb the consistency, which is the impression that all information objects match the physical world in position and orientation in the way they are intended to at all times. Examples are faulty tracking that places virtual information objects at wrong positions [245], incorrectly used depth cues that make objects appear at the wrong distance [168], or latencies that prevent simultaneous and conformal movement of physical and virtual parts of a scene [91]. If a combined view no longer appears consistent, a user must expend additional sensory and cognitive effort to understand the correct spatial relationships. Therefore, a user should be presented with a consistent view at all times. This includes a correct and sufficient use of depth cues, or a minimization of latencies below certain thresholds. In contrast to classical user interfaces or virtual environments, this is made more difficult by the fact that not all parts of a scene can be controlled equally. Instead, it is always necessary to react to changes in the physical world that can take place at any time without prior preparation or a possibility to prevent them. It is interesting to note that the causes for inconsistencies can be very diverse from a technical perspective. Nevertheless, they have the same effect on the presentation of information; namely that the scene is in an inconsistent state which must be mentally compensated by the user.

Two examples of this challenge can be found in figure 4.2. Subfigure 4.2a shows a green square which is supposed to be behind the wall. However, because this is not supported by visual cues, it will be very hard for a user to see and understand this relation and get a consistent impression of the scene. In subfigure 4.2b a drop shadow is added to the right cup but not to the left one, integrating the cup with the shadow better into the physical world and helping a user to perceive this scene as a unity.

This challenge is divided into three aspects, depending on the form of the inconsistency and the necessary measures. One of them, depth perception, is further divided into two sections for the sake of comprehensibility. The first one, depth cues, covers influencing factors on how depth is perceived by humans and which cues can be used to convey it. The second, depth visualization, covers specific visualizations that have been created to convey correct depth of information objects in augmented reality. The aspect visual integration covers techniques that support the visual integration of information objects into a scene. The last aspect, timing and latency, includes timing factors that must be considered to make a scene appear consistent.

Derivation. The source of possible unnecessary sensory and cognitive effort that is addressed with this challenge is caused by a combination of three characteristics. The first characteristic is spatial relationship, which states that the information objects of a scene are positioned in a spatial structure that must be recognizable for a user. If it is not unambiguous where each object is positioned and how each of them is oriented, it is difficult to get the impression of a consistent scene. The second characteristic is combination, which implies that information objects and the environment from a scene cannot be seen individually but only in a combined view that depends on the viewpoint. However, neither changes of the viewpoint, the environment or physical information objects are always controllable or predictable. An augmented reality system can only react, and a consistent state must be recovered after each change of one of them. Additionally, if no unambiguously consistent structure of the scene is recognizable in a combined view, the user must spend extra efforts to understand the scene. Therefore, the structure, including the position of a user within the scene, must be as clearly recognizable as possible from any viewpoint. The third characteristic is fluctuation. When the effects described by it become too strong, e.g., if virtual and physical information objects are not aligned appropriately, an impression of a consistent scene is difficult to achieve.

Previously described by. To the best of the author's knowledge, this challenge has not yet been formulated as a challenge of consistency. Nevertheless, various aspects have already been described, but they have never been combined to the complete challenge, and neither have they been related to the use case of supporting procedural tasks. In particular, depth perception is a topic that has been widely studied for augmented reality, and multiple authors have addressed the challenge to provide sufficient depth cues. The first ones were seemingly Drascic and Milgram, who note that "the difficulty lies in the creation of an accurate sense of depth" (Drascic and Milgram [86], 1996, page 1). Concerning visual integration, Haller [131] recognized its importance and investigated which use cases require how much realism when virtual objects are included into an augmented reality scene and what techniques are needed to achieve this. However, even before that, Fischer et al. [106] described “effects that cause discrepancies between the visual appearance of real and virtual image portions” (Fischer et al. [105], 2015, page 5) that cause keeping immersion high a challenge. For tracking, the challenge is often described as the need to improve sensors and algorithms. However, less attention is given to the problems resulting from not having near-perfect tracking. Nevertheless, there are already authors who have identified the challenge as such, especially Robertson et al. [245] and Vincent et al. [293, 294]. There are also investigations on timing and latency in augmented reality, but the challenge of consistently representing a scene in terms of these two parameters was not clearly addressed. Ellis et al. [92] come closest by examining how latency and delay affect user performance in augmented reality environments.

4.2.1 Depth Cues

Depth perception is an important factor in augmented reality and has been intensively researched. If an augmented reality system fails to provide sufficient depth cues, it becomes difficult or even impossible for a user to develop a correct and consistent understanding of a scene [113]. Humans use multiple cues to estimate the depth of objects. It is not important that all of them are present, but enough cues must be provided and, maybe most importantly, they must not contradict each other [70]. They are generally divided into binocular and monocular depth cues [70], with the former being those that can be perceived because the human eyes are located at a small distance from each other, and the latter being those that can also be perceived with one eye.

Despite extensive research results, it has not been fully clarified how depth cues have to be given in augmented reality in order to always display a scene consistently. As the next section shows, the identified ones are not entirely free of contradictions and some aspects are not yet fully clarified. Nevertheless, using them can help to improve the depth representation in an augmented reality scene.

Multiple compilations of depth cues have been developed. One widely regarded model was created by Cutting and Vishton [70], who worked out which depth cues exist and what their importance is in which distance. It was not created for augmented reality, but their work serves as a solid foundation for what depth cues can be used for what distances. Overall, they identified the following cues:

- *occlusion*: objects are in front of other objects which they occlude
- *relative size*: the relative size of objects becomes smaller compared to their actual size with increasing distance
- *relative density*: stochastically equally distributed objects seem to get more dense with distance
- *height in the visual field*: objects on an imaginary ground pane get increasingly higher in the visual field with distance
- *aerial perspective*: distant objects look more and more hazy

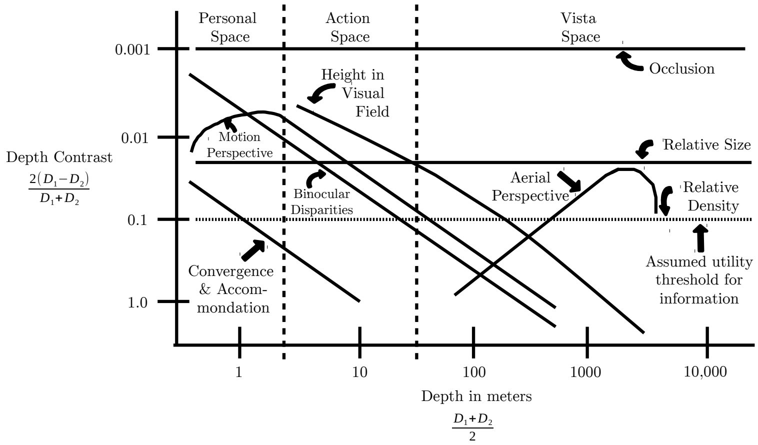


Figure 4.3: Different depth cues according to Cutting and Vishton including their relevance for different depths. D_1 and D_2 are the distance to the two objects whose relative depth shall be estimated (source: own work after Cutting and Vishton [70])

- *binocular disparities*: objects that are seen from two slightly different positions show different perspectives which converge at a greater distance
- *motion parallax*: based on their distance objects seem to change their relative position when a viewer moves
- *convergence*: the angle between the optical axis of the eyes gets smaller with increasing distance of an object
- *accommodation*: the eye lens must be adjusted differently based on the distance of an object for a focused retinal image

Figure 4.3 shows the depth cues by Cutting and Vishton including their relevance depending on the distance to the viewer.

Specifically for augmented reality, Furmanski et al. [118] suggested another set of depth cues:

- *transparency*: translucent surfaces through which other objects can be seen imply a depth order
- *occlusion*: the occlusion of one object by another object conveys a depth order
- *size-scaling gradients & texture*: objects and textures get smaller with distance
- *shading gradients*: contrast of objects relative to their environment decreases when their distance increases
- *cross-referenced depth*: external cues such as shadows or virtually added cues like ground panes help to estimate depth

- *motion parallax*: the relative motion of objects to each other and to the viewpoint when the viewpoint is moved gets less with an object distance
- *structure-from-motion*: reconstructing objects based on the different perspectives while moving give a user a general understanding of a scene's spatial structure
- *binocular cues*: the different perspectives of each eye from an object converge and the angle between the optical axis of the eyes gets smaller at a greater distance.
- *visual-motor cues*: needed accommodation of an eye to get a sharp image depends on object distance

Based on these cues, they developed a set of design solutions that can be used to convey depth in obscured information visualization scenarios. Examples are additive transparency, size scaling of rendered surfaces or temporal distance coding [118]. Together with other visualization the convey distance, they will be discussed in the next section.

Even though many cues are known and have been examined, there is no universal model that in every situation allows determination or prediction of the most relevant depth cues unmistakably [70, 182]. However, with the results from Cutting and Vishton, one can estimate which cues should be primarily used for a given distance. The influence of individual cues for the perception of depth can be modelled as a linear combination of the present ones, weighted by the perceived reliability of each of them [194]. The weighting can be influenced by previous experience of individual users. Using a combination of several depth cues is preferable compared to only using single cues [79].

When estimating the distance of virtual objects on head-mounted displays, users tend to report distances incorrectly [150, 277, 278, 79]. Close objects are estimated as too close, but this decreases with growing distance until, from some distance on, objects are estimated as further away than they are [277, 278, 79]. In one experiment this point was found to be at 23 meters but this value cannot be generalized [277, 278]. However, in a later experiment with different conditions, no large underestimation of distance was found [152]. At least in part, this seems to be due to incorrect verbalization of depth measurements, since walking was used in this later experiment. Perceiving motion parallax, i.e., being able to significantly move the viewpoint, does not improve this depth estimation error in augmented reality [177]. Furthermore, the actual effect seems to depend on the size of the field of view [278].

Occlusion is a very important cue in practice, maybe even the most important one [69]. However, other studies have also found different depth cues to be the most important one for some situations, e.g., for autostereoscopic displays, these were found to be the binocular cues [194]. If an occluder, e.g a billboard, blocks out significant parts of a scene behind a virtual object, the depth estimation gets worse and the mean error of the estimation increases [277, 278]. However, this effect was not present in all studies [79], suggesting that the concrete effect seems to be quite complex. Similarly, if a virtual object becomes too transparent and a user cannot use the occlusion depth cue correctly anymore, depth estimation becomes less accurate [93]. Dimmed targets tend to be estimated as further away than brighter targets [177]. Binocular cues can be used to convey distances in depth between objects larger than two arc minutes [177]. Below this limit, users do not reliably see a difference anymore. A mostly local cue, which especially helps in monocular systems, are virtual shadows cast by virtual objects on the physical environment [275]. In user tests, these shadows showed a significant improvement in correct depth estimation [275, 79]. Interestingly, the correct alignment of physical and virtual light source is, at least for indoor scenes, not that important. Instead, a correct shape of the shadow is the key factor. The cues shading, aerial perspective and texture were not found to be very helpful to improve depth judgement in augmented reality [79]. With all these cues, it is important that a user has references. If the depth of objects in the vicinity of an object, whose depth has to be estimated, is clearly visible, the estimation becomes easier and much more precise [307].

4.2.2 Depth Visualization

In order to avoid having to develop new visualizations that give the needed depth cues when creating augmented reality user interfaces, reusable visualizations have been developed and published by various authors. These allow distance or position of virtual objects to be conveyed for defined situations. Some of them are presented this section.

In augmented reality, virtual information objects often seem to float in front of opaque objects when they are supposed to show hidden structures inside or behind of them. One solution to this is to add a virtual window or over-rendered transparency. For them, a spatially confined area on the surface of a physical object is graphically overlaid in such a way that the impression of a view inside like through a physical window is created [118, 267, 39, 180]. When the viewpoint is moved, the virtual information objects move relative to the virtual window. The parts of them that cannot be seen through the virtual window are cut off to the give the impression of occlusion. To further improve depth perception, the window can also have walls that seem to go into the physical object as if a part was punched out.

Another approach to represent virtual objects in or behind physical objects is called edge overlay visualization. In this approach, edges, i.e., parts with a high local visual dynamic, of the physical object in the foreground are preserved and areas with less features, i.e., uniform parts of the surface, are shown as transparent [19]. In place of the transparent parts, the virtual objects are displayed instead. This form of masking creates a clear order of depth. As an improvement, further feature types can be included which are preserved from the physical object [253]. To also balance between the virtual objects in the background and the physical object in the foreground, a saliency map can be created. Based on features of each object, this map can be used to decide which part of which object must be shown. In a user study, the simple edge overlay was preferred over more complex features [253]. However, the authors of this study argue that the results especially reflect flaws of prototype and test design and not the quality of the underlying method. They also suggest including a tunnel cut, which is similar to a virtual window, with virtual walls forming a tunnel behind the edge overlay. Instead, of recording and processing the actual surface structures for edge overlay, it is also possible to render them as part of the virtual image [174]. The advantage is that properties like color, texture etc. do not need to be extracted at runtime. However, this needs very thorough prior knowledge of the surface of the physical object.

Another way to visualize virtual information objects as inside of physical objects is focus and context visualization [155]. Here parts of the physical object are virtually reproduced as the context and displayed together with the information object. The context is a stylized visualization, which is kept very simple, non-distracting, and is spatially correct to the information object in terms of perspective and overlapping. In contrast to edge overlay, surface structures are not simply reproduced, but are included as an abstracted model, which usually needs to be known in advance. Additionally, it can also contain parts from the inside of the physical object which are not visible from the outside. Giving such a context has proven very useful in connecting virtual with physical objects [245].

One type of visualization has also been developed for the case when multiple overlapping objects are displayed. In such a situation, it is even more difficult for a user to see the relative depth and position of these objects. This is especially true when physical objects are among the overlapping objects [181]. The solution used for the visualization is to greatly simplify the virtual objects to wire frames and flat surfaces [181]. These surfaces are then shown with a certain transparency, decreasing with every consecutive overlay and creating an impression of depth for the objects. This type of transparency is also known as additive transparency [118]. The disadvantage of this method is certainly that the necessary changes to the information objects often lead to a highly changed and simplified appearance.

A number of visualizations are also available for better estimating the depth of objects, i.e., their distance from the viewpoint or from other objects. These are particularly useful for objects that do not directly interact with others and where no direct cues such as occlusion are available. A ground pane, a flat 2D pane that is projected in the scene, can provide a means of reference for object depth and distances [181, 221]. A regular texture on this pane, such as a grid with fixed cell sizes, can further support the perception of distances. The depth estimation can also be supported through shadows cast by the virtual objects on this pane [307]. Similar but less intrusive approaches are virtual rails, which are two parallel lines that span from the user to the horizon like train tracks [29], virtual tape measure that is placed in the scene and is marked with distances indications [196] or distance markers like virtual yard sticks or textual information on the distance [118]. Another way to support depth estimation is the introduction of auxiliary objects in the scene, which are positioned once at a depth that is easy for the user to recognize and once at a point of which a user needs to estimate the depth [307, 170]. The relative size makes it easy to see at what distance the second object is located. A combination of them with the virtual window is possible as well [170]. Further, a mini map of the scene can also be used to support correct depth estimation [307]. While these different types of visualizations are suitable to support estimation of depth, the author is not aware of any comprehensive investigation as to when which of them should be used.

An overview of depth visualizations for augmented reality that goes beyond the scope of this work can be found in the work by Furmanski et al. [118] and Livingston et al. [181].

4.2.3 Visual Integration

It is important to integrate virtual information objects into a scene such that a user has the impression that they coexist with the physical information objects and the environment. For this, it is not necessarily important to create a presentation that is as photorealistic as possible. Instead, it is more important that the integration is believable, which also includes aspects like a plausible behavior of virtual information objects [131]. In many cases, it is very well possible to use stylized visualizations for virtual information objects [106], and may sometimes even provide a better integration than rendering information objects as realistic as possible [131, 107]. It is also known that increased realism does not necessarily increase trust of users in the information presented by augmented reality [311]. Actual rules on how much realism is needed and relevant influence factors are still missing. However, there seems to be a certain need for a minimum of realism [131]. For systems which display a video stream with integrated virtual information objects, it is also important to emulate camera properties like camera noise or motion blur when rendering the objects[105].

Lighting and shadow influence how far virtual objects are perceived by users to be part of a scene. A correct lighting and illumination can support the impression of visual integration [103, 275]. Shadows cast by virtual objects also contribute to this [275]. Even if lighting, illumination and shadows deviate from the actual physical conditions, using them still supports the impression of presence in the scene. For shadow representation, the characteristic shape of the throwing object is especially important [275]. A more extensive overview on the topic of illumination models, albeit with a more technical focus, can be found in the work of Jacobs and Loscos [149].

Concerning spatial registration, two kinds of errors can affect the perception of visual integration: static errors, where the virtual parts of a scene are displaced by a fixed offset compared to the intended position, and dynamic, randomly changing misalignment errors, where the virtual parts of a scene are displaced by an offset that changes randomly. Both affect how well a user perceives a scene as consistent. How large errors may be in order to be acceptable, however, has not yet been clarified. For movements of the viewpoint, a local error of approx. +/-1 mrad (approx. 3.5 arc minutes) is accepted before virtual objects are perceived as not uniformly moving with a background [113]. Depending on the use case, larger errors might be tolerated, but so far, no general rule has been determined. Several measures exist to ameliorate negative effects caused by faulty registration. One is to include a graphical context to virtual objects, which is a stylized visualization of the physical structures around the intended position of a virtual object [244]. It has proven to increase trust in and the understanding of an augmented reality representation [245]. Besides indicating the actual position of a virtual information object, context even helps to increase the visual integration of virtual objects when no registration errors are present [155]. Another approach is to relax the spatial mapping between physical world and virtual objects based on the available precision [293]. This can range from allowing some freedom between the intended mounting of the virtual object and the actual position to placing virtual information objects in the reference system of the display, to using lines or pointers to show the connection. If the registration accuracy can be estimated at runtime, level of error filtering can be applied where an appropriate presentation is selected which corresponds to the needed relaxation of spatial mapping [188].

4.2.4 Timing and Latency

Also temporal aspects can affect how consistent a scene is perceived by a user. A high latency and a slow frame rate both cause the combined view to be inconsistent. This effect appears in both cases after changes when the representation of virtual objects does not match the physical world for a short time span. On a video display, the entire combined view lags behind the physical world, whereas on an optical see-through display, the virtual information objects lag behind the physical world. To cope with this, cognitive effort is necessary that greatly reduces the benefit of augmented reality support [113]. For head mounted displays, it has been estimated that one millisecond delay causes about 0.33mm registration error on average and 1mm in the worst case for objects at arm's length [141]. The resulting offset must be cognitively corrected by a user.

Humans perceive time frames of about 20ms to 30ms as one moment in time during which no order of events is registered [228, 139]. However, this is not the lower threshold below which latency does not need to be optimized. Even with latency as low as 10ms, negative effects on interaction with virtual objects were found [91]. The reason for this are spatial offsets that are created by latency or low frame rates [206]. A user can see that a virtual object follows its intended mounting the physical world with a certain distance based on the speed of movement and the system delay. For latency, this distance is constant while for low frame rates the user can see jumps. As already mentioned for visual integration, the difference of physical and virtual objects of +/-1 mrad (approx. 3.5 arc minutes) should not be exceed [113].

Besides algorithmic measures like predicting movement, measures to cope with this are similar to those needed for tracking errors, as both kind of interference

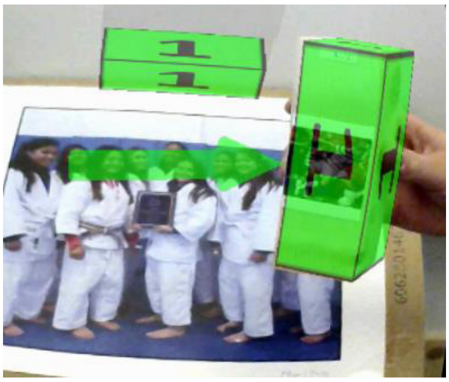


Figure 4.4: The virtual building block makes the physical block nearly invisible and hard to interact with (source: Syberfeldt et al. [279])

manifest in a spatial offset. However, this does not yet answer the question of how latency behaves in comparison to frame rate and how they can be balanced against each other. Ellis et al. [92] created a model to predict how these two factors influence perception by users. They found that these two factors can in fact counterbalance each other, e.g., a low frame rate with low latency has a similar effect as a high frame rate with a high latency. They also determined specific curves for these phenomenon, which can serve as a decision support and can be found in their work [92].

4.3 Visibility Challenge

Information objects convey information for a user that is relevant for completing a work step in a procedural task. However, if they are not visible, this information is not accessible. At least additional effort, e.g., by moving the viewpoint to get a better look, will be necessary, or it may even happen that augmented reality support is rendered useless. If safety relevant parts like sharp edges or moving parts are covered, lacking visibility might also become a safety hazard. It is therefore absolutely necessary that a user can see all information objects. Thus, this challenge is about making fully or partly occluded information objects, which cannot be seen even if a user is looking in their direction, visible. However, only the cases in which information objects overlap each other are relevant. Otherwise a solution is rather trivial, because it is obvious that the information object has to be visible while while the environment can be concealed. Information objects outside the view frustum are not addressed by this challenge because a user would just have to look in the right direction and, instead, needs to be aware which direction this is. Thus, it is created by a different combination of characteristics and results in a different challenge, orientation, which will be discussed later. The visibility challenge should also not be confused with the occlusion problem, which is about how accurate and perspectively correct occlusion of virtual and physical objects can be achieved [308]. This, however, is a technical problem.

In many cases, the visibility challenge can be solved by moving the viewpoint. This is not accepted as a valid solution here, though, as it requires extra effort from a user, which should be avoided if possible. Instead, a presentation should be achieved that only requires minimal actions from a user.

An example for the visibility challenges can be found in figure 4.4. The green virtual building block is a spatial virtual information object that is covering the wooden physical building block, which is an indirect physical information object. While this is a correct occlusion, it makes interaction more difficult, hides any details and diminishes visual feedback.

This challenge is divided into two aspects based on two fundamentally different approaches by which it can be solved. One is resolution by transparency, which means that virtual information objects are made partially transparent such that those in the background can be seen through the ones in the foreground. The other one is resolution by relocation, which means that information objects are relocated from their original mounting position or anchor to another place. A further included section is the description of an approach on how to decide which information object has to be relocated in case this approach to resolution is used.

Derivation. The source of possible unnecessary sensory and cognitive effort addressed with this challenge is caused by a combination of three characteristics. The first characteristic is spatial relationship, which indicates that both virtual and physical information objects from one scene are spatially arranged in a common space. The second characteristic is combination. When seen from a viewpoint in a combined view, information objects may occlude each other due to either perspective or because they take up the same space. However, in both cases, not all relevant information is visible. Having a user to change the viewpoint to create a preferable combined view should also be avoided because it creates effort on the user's side. The third characteristic is manipulability. It implies that solving the visibility challenge is not trivial because simply relocating objects into non-overlapping positions is not always possible and may result in complicating the solution of other representation challenges. As described by this characteristic, changing an object's position or orientation may be very difficult or even impossible and doing so may also disturb its meaningful spatial context or move it away from its anchor.

Previously described by. To the best of the author's knowledge, this challenge has not yet been described to the same extend as it is here, especially not with a focus on supporting procedural tasks with augmented reality. Very close are Bell et al. [36], who describe a view management system for augmented reality that requires that predefined physical objects may not be covered by virtual objects and, instead, must be arranged around them. They also recognize that physical objects elude the control of an augmented reality system and it thus has to react to changes in the physical world. For them, however, the central challenge is not to maintain visibility of information but rather to avoid occlusion. Livingston et al. [181] described the problem that multiple virtual objects can occlude each other and suggested transparency-based measures. However, they characterized it as a problem of a correct representation of depth. King et al. [159] found that for augmented reality geo information systems it can be difficult to find the right transparency of virtual overlays such that both they and the physical environment are sufficiently visible. They considered it to be a problem of immersion but not of missing information. Vincent at al. [293] describe what they call the fat-finger problem, which occurs when a user's finger covers information objects on a screen. This can be considered a special case of the visibility challenge. For labels in augmented reality, the challenge has been described as a placement problem, e.g., by Peterson et al. [220, 221, 222] or by Azuma and Furmanski [21]. The central issue of information being non-visible caused by the use of augmented reality is only marginally considered in all of these descriptions. For virtual environments, however, the problem and possible solutions have already been described [93]. These, however, lack the restrictive element of limited manipulability which, in augmented reality, results from the integration into the physical world.

4.3.1 Resolution by Transparency

Using transparency to resolve multiple overlapping objects has been treated as part of depth perception so far [181, 253, 19]. The focus was not to ensure that all objects were visible to fulfill a need to see them, but, instead, that the depth order of the objects was clear. In one case, transparent overlays were also used to improve immersion [159]. Even if these approaches are not intended for the visibility challenge, they can be used as measures for it.

For virtual environments, transparency has already been applied and tested successfully as a means to show otherwise occluded parts [93]. One important result from this research is that the object in the foreground has to stay sufficiently opaque as to not result in reverse occlusion, i.e., the effect that the object further away seems to cover that in the front. In augmented reality, this form of transparency including the necessary opaqueness has been used for supporting depth perception, but without examining the visibility aspect [245, 181]. The downsides are that it is hard to balance the transparency such that all the layers of objects are simultaneously sufficiently visible [159], and that complex transparency, especially in binocular systems, can cause problems with depth perception [151]. When using optical see-through displays, transparency between virtual and physical objects may come as a side effect of their operating principle [292].

A more complex approach is to selectively remove parts of the foreground object. Here, a feature map with the salient features of the foreground and background objects can be used to decide how the objects are overlaid and combined [253]. Details for this approach were already detailed in the depth perception discussion. One possibility is to make uniform parts transparent, while parts with higher dynamics, e.g., corners and edges, remain visible [19]. Similar to this is the idea to only show parts of objects in the background by including a stylized version of them, e.g., the outlines or stylized wire frames with otherwise transparent surface areas [155]. While they are similar, virtual window techniques [118, 267, 39, 180] cannot be applied in many cases because they remove large parts from the information objects in the foreground which is rarely possible without loss of information. Therefore, they would not solve the visibility challenge.

4.3.2 Resolution by Relocation

At some point, too many overlapping objects are no longer easily recognizable, even with carefully selected transparency and such a resolution is no longer useful [181]. In some cases, this may already occur when two objects overlap. Instead, these can be moved apart such that they do not overlap anymore, and a user can see all of them at the same time without any occlusion. For labels, i.e., spatially referenced virtual information objects, this has already been extensively researched with either the focus of not having the labels overlap each other or with important structures in the environment, i.e., direct or indirect physical information objects [220, 222, 221, 21]. For view management, such goals are usually modeled as visibility constraints [36]. While the techniques were developed specifically for labels, there is no reason not to apply them to other types of information objects.

One approach is to make a planar separation and move virtual information objects in arbitrary directions such that they no longer overlap with each other in a combined view. For this, it is of course necessary that all of them are arranged on one plane, e.g., the display plane. The way this must be performed highly depends on the usage situation, especially on the time the viewpoint is expected to stay at a fixed position and how long the information objects will persist [21]. If it can be expected that a user will only have a brief look, it is reasonable to use a greedy algorithm that quickly reaches a relatively good and stable distribution. Contrarily, a cluster-based algorithm that optimizes the position of the information objects over time and minimizes jumps during movements is ideal for a longer observation. Another approach is to map the distance of information objects' anchors to the viewpoint as height in the visual field [221]. However, this assumes that most anchors are at different distances from the viewpoint.

If a device with a binocular display is used, virtual information objects can also be arranged in depth at a distance of 5-10 arc minutes to enable the user to distinguish them more easily [220, 222]. Compared to the previous approaches, this method has the advantage that information objects can be kept at fixed positions and cognitive effort to recognize them and their anchors is reduced. Strictly speaking, this measure does not solve the visibility problem, but makes it easier for the user to solve it mentally on her or his own. Therefore, it only works for sparse graphical visualizations without too much overlap where the problem is not so much the non-existent visibility, but the differentiation of overlapping objects. When, for example, labels with a billboard style are used, they block out any information object behind them. When the distance between the information objects is increased to over 20 arc minutes, user performance decreases, hinting that such a representation creates additional sensory or cognitive effort. Further, their depth order should correlate with those of their anchors.

If overlapping of information objects in a combined view is too difficult to avoid, spatially referenced virtual information objects should be placed on relatively uniform areas [251, 282, 127]. Those can be assumed to convey relatively little information for users and do not interfere with text or other structures of information objects. Of course, it depends on the specific task and work environment, if this assumption holds true. It might very well be that information objects contain uniform surfaces which must be fully visible. However, placing labels or similar information objects that contain text in front of non-uniform surfaces can interfere with legibility [175] and must be handled by visualization techniques like billboard styles. This is addressed more depth in the clarity challenge. Information objects should also stay as close as possible to their anchor and be positioned in such a way that existing connection lines to their anchors do not cross [127, 200].

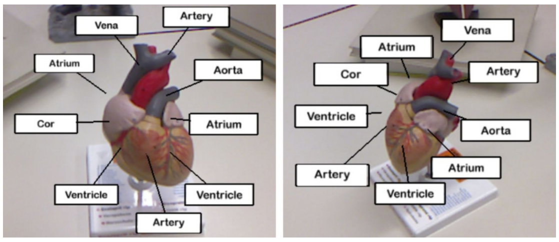
4.3.3 Relocating Information Objects

As discussed in the previous section, mutal occlusion of information objects can be resolved by relocating them. However, a strategy is needed to decide which information object to move and which to leave at their original mounting position. The following section presents an approach to such a strategy in which the basic idea is to use the decreasing spatial relationship and the increasing manipulability over the classes of information objects to decide this.

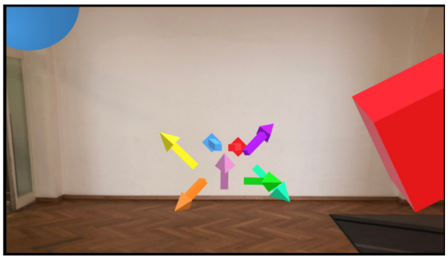
Direct physical information objects cannot be manipulated by an augmented reality system. Therefore, if there is an occlusion of these by any information objects of another class, the resolution is trivial. The other object must be moved to keep it visible. There is also one thing to consider with direct and indirect physical information objects: if an indirect physical information object can cover a direct physical information object without any consequence, the latter one is not an actual information object but is, instead, fully replaced by its reproduction. Therefore, in this case, there is no problem with visibility and no objects have to be moved.

When an indirect physical information object is occluded by a spatial virtual information object or vice versa, the spatial context determines which object must be moved. Depending on which option causes least information to be lost, one or the other must be taken. However, in many cases, this will be the spatial virtual information object when the indirect physical information object is part of its spatial context. Hence, if the indirect physical information object is moved, both of their spatial contexts are lost.

A prerequisite for this consideration is, of course, that the indirect physical information object can be extracted and moved. If this is not the case, the virtual information object has to be moved in both cases. In case an indirect physical information object is occluded by a spatially referenced or detached virtual information object, the solution is again trivial. These two types of information objects



(a) Temporal incoherence forces users to search for object placement over again (Madsen et al. [191])



(b) Arrows point to objects outside the field of view (source: Bork et al. [50])

Figure 4.5: An example of an occurrence of the orientation challenge and an example of a measure to solve the challenge

do not have a meaningful spatial context and must therefore be moved.

When only virtual information objects overlap, spatial virtual information objects must be kept at positions where they have a meaningful spatial context. Otherwise, this would increase a user's cognitive effort and as a result the number of errors made [244, 245]. Spatially referenced virtual information objects do not have meaningful spatial context and can be moved as long as the referenced anchor is easily recognizable [200]. However, a close proximity to their anchors helps to avoid sensory and cognitive effort to understand the connection [200]. With detached virtual information objects, this is not a problem anymore because they simply do not have an anchor.

Lastly, detached virtual information objects must be moved away when occluding any other class of information objects. They can easily be manipulated by an augmented reality system; a meaningful spatial context cannot be disturbed and a proximity to an anchor also does not have to be maintained.

Even with this approach, it remains unclear how a prioritization can be made if objects of the same class occlude each other. Furthermore, it is still to be clarified how important it is to enforce these rules. Other requirements, such as that information objects should show a minimum of movement, may or may not be more import. When, for example, a detached virtual information object, is moved every times it covers any other information object, it may be very hard for a user to follow its position. An empirical validation of this approach is also still missing.

4.4 Orientation Challenge

[2]At any time, users of an augmented reality system must know where they are in the scene and where all the relevant information objects are. On a head-mounted

[2]In a previous publication [199], this challenge was called orientability challenge. However, the term led to misunderstandings, so the more common word orientation is used here.

display, for example, the combined view can contain so many virtual objects that the physical parts of the scene are barely visible, making orientation very difficult. With a handheld device, a similar situation can result in the combined view on the display being difficult to relate to the physical world. Again, orientation becomes difficult, with the additional complication that on the display, the scene is seen from a different angle. If a changed reproduction of the physical world is shown on the display, e.g., false color images, this can become even more difficult. But even if a user knows her or his position and viewing direction in the scene perfectly, this does not mean that the positions of all information objects or even their existence in the scene are known. Some may even be hidden in plain sight, when its not clear that they are information objects, e.g., small direct physical information objects.

The restricted view fustrum in the combined view further complicates this problem, since information objects outside of it are not displayed. If there is no further support available, the user has to search for them and remember their positions. Some information objects may even never gain the user's attention at all and are therefore never seen. When the scene changes, known objects can also move or disappear. Thus, a search becomes necessary, which can last long and, for disappeared objects, ends unsuccessfully. Another aggravating factor is that people tend to develop tunnel vision when using augmented reality [281]. Thus, significant effort may be required for users to obtain and maintain an overview of their own position as well as of the existence and positions of information objects. The challenge is therefore to support users in reducing the needed sensory and cognitive effort to obtain and maintain such an overview.

Examples can be found in figure 4.5. Subfigure 4.5a shows how information objects move over time, making it necessary for a user to find their positions once again. In subfigure 4.5b, three-dimensional arrows are used to point to objects outside the view frustum. This helps users orient themselves more easily in the scene and not have to remember their positions of known objects.

This challenge is further divided into three aspects. These are spatial orientation of a user in the scene, discovery of objects outside the view frustum and keeping track of known objects and their positions. This division is based on a suspected chronological sequence of the orientation of a user. First, an initial orientation in the scene is achieved, followed by the discovery of the individual information objects and finally, the maintenance of the overview obtained in the previous steps.

Derivation. The source of possible unnecessary sensory and cognitive effort addressed with this challenge is caused by a combination of four characteristics. The first characteristic is spatial relationship, which implies that all information objects, whether physical or virtual, are arranged together with the environment in the spatial structure of a scene. A user has to find an orientation in this structure, both in relation to their own position and the positions of the information objects. The second characteristic is combination. It implies that finding an orientation has to be done via a combined view limited to a view frustum. When this is small, it is not always possible to see the overall structure of the scene and all relevant information objects at once, but an increase in size only helps to a certain extent since the human field of vision is also limited. The third characteristic is discrete and continuous change. The change of information objects as described by this characteristic, i.e., the appearance, movement and disappearance of objects, exacerbates this challenge since knowledge already acquired about the structure of a scene can quickly become obsolete. The fourth characteristic is reference systems. In addition to the previously described effects, it implies that information objects can be in different reference systems invisible to the user, which can cause the structure to change in unexpected ways.

Previously described by. To the best of the author's knowledge, this challenge has not been described in its entirety before. However, the relevant aspects have been addressed previously by other authors. Bell et al. [37] recognized the need for a user to be aware of their environment in an augmented reality scene. Their main focus was to help a user discover and understand the physical parts of the world. Discovery of virtual objects was not relevant to them, as in their system every virtual object only described a physical object to which it was attached. Feiner et al. [100] formulated the need to direct a user to specific objects outside the field of view. Later, Biocca et al. [48] showed the importance of pointing to objects within the field of view and Schinke et al. [258] saw the need for referencing multiple objects at the same time. In the context of procedural tasks, the need to indicate places outside the field of vision was presented by Henderson and Feiner [136]. Being able to keep track of information objects has been described as an issue that needs to be solved for augmented reality interfaces by Azuma and Furmanski [21]. They saw it as an algorithmic problem to place labels such that a user can easily find them again. However, to the best of the authors knowledge, objects changes outside the field of view have not been addressed so far.

4.4.1 Spatial Orientation

Spatial orientation is about a user relating their own position respectively the position of what they see on a display or through a combiner to the physical world. For displays outside the line of sight, it becomes difficult when the display gets too cluttered [153]. Because they are not head attached, the angular difference between the perspective of the user and the perspective shown on the display may be very large and contribute to this problem [168]. Generally, it takes a person about one second to mentally rotate an object for 60°[265]. However, this time can be significantly extended in complex situations. With head mounted displays, spatial orientation can become even more problematic because the display or combiner is in the line of sight. Thus, it is important to leave relevant landmarks clearly visible [153]. Therefore, they should be included in the modeling of the scene as information objects. On video see-through displays, these would usually be indirect physical information objects, but also a replacement with spatial virtual information objects is possible. A focus and context visualization can help by displaying such landmarks as a stylized depiction, which allows a smaller graphical footprint compared to a video reproduction [155]. On optical see-through displays, they are naturally direct physical information objects, but it is also possible to use spatial virtual information objects as on video see-through displays. However, the discussed transparency issues of any graphics displayed on the glasses need to be taken into account. A further support for users in orienting themselves in a scene can be the inclusion of a mini map [37]. This gives a user a different perspective in addition to the ego perspective and serves as a further source for orientation.

4.4.2 Object Discovery

Whenever information objects outside the view frustum need to be (re-)discovered, they should be indicated. For this, several pointing techniques have been developed. The probably most simple one is a 2D arrow that points in the general direction, i.e., left or right, of the information object [137]. A visualization comparable to a compass needle which points in the exact direction of the information object relative to the current viewing direction already provides a much more precise direction indication [50, 258, 101]. To convey a distance measure, the length of the needle and or arrow can be used. If an even more precise guidance is necessary, 3D arrows or ray cues can be used. These stretch from close to the viewpoint to the targeted information object and thereby also mark its actual position and guide the user effectively [50, 254, 136]. If it is known with a high degree of certainty that a user is looking at a particular object, the target object can also be indicated by a connecting arrow that goes from the currently viewed object to the target [295]. An attention funnel is a set of concentric circles, rectangles or similar shapes that form a funnel which leads to the target position [48, 47, 263]. One of its openings is fixed at the user's viewpoint and the other at the target. Due to the flexibility with which the individual elements can be positioned, arbitrary paths can be defined that lead around obstacles.

A different approach is not to directly point towards information objects, but to, instead, include a map on which their positions are marked. This can range from a simple mini-map, which only includes their locations shown a 2D pane in relation to user [258] to a much more detailed world in miniature [37, 29]. In a user study, a simple mini-map was outperformed by a compass visualization, but it is not clear whether these results can be generalized [258]. Another possible technique is the so called EyeSee360 [129]. This is a visualization where a user's surroundings are projected onto a flat presentation to a Mollweide projection for maps that is shown in the field of view. Relevant information objects that are not inside the field of view are then marked as points on this projection.

When information objects are inside the view frustum, they still have to be discovered by a user. If they are not very salient, the so called visual search tasks must be supported to minimize the effort which is caused by it [185]. In principle, the same cues that are used for information objects outside the view frustum can be applied. However, they are prone to creating visual clutter while other cues can avoid this problem. One possibility is to superimpose information objects with highly transparent shapes that are still sufficiently colored to enhance the local contrast and direct attention [185, 186]. Similarly, it is also possible to manipulate the information object itself by modulating its features to become more salient, e.g., by increasing the corresponding pixels' brightness [193]. By making the environment around it less salient, the effect can be enhanced. However, this approach is limited to video displays.

4.4.3 Keeping Track

Organizing objects in a spatial structure helps humans to remember them and recall how to find them. Because in augmented reality, virtual objects are placed in the physical world and thus a location is given to them, keeping tracking of the objects is supported in a quasi-natural way [205]. However, information objects and their positions inside the scene may move at unknown times. Reasons are, for example, the start of a new work step, where old information objects disappear and new ones appear, changes to one information object that result in the relocation of others, or a change to the viewpoint that makes it necessary to rearrange the virtual information objects. As a consequence, a user has to react to these changes and find the moved objects again, causing sensory and cognitive effort [21]. Users also tend to use information objects as external memory and simply remember their positions but do not remember the associated information [226]. Thus, losing track of an information object also means losing this part of the working memory until it is found again. Tests have shown that even smaller jumps of text pieces on a CRT monitor decrease readability significantly because their new position must be determined to resume reading, which needs sensory and cognitive effort [164, 126]. In augmented reality, moving objects have even more serious consequences because unlike on a monitor, there is no defined limit as to where an object can move.

Because of the previously mentioned reasons, information objects should be kept at constant positions. A simple solution is to place them in a scene at a fixed position that optimally does not change during the lifetime of the object [127, 191]. Spatially referenced virtual information objects should be close to their anchor [226], and spatial virtual information objects should be placed in their meaningful spatial context. Positioning information objects in the reference frame of the combiner or display can provide fixed positions which are very easy for a user to locate and can be kept comparably stable [226]. A fixed position from the user's perspective can even be more important than a close proximity of information objects to their anchors [226]. The downside of such a solution, however, is that it is relatively inflexible, as it does not allow for major changes in the scene. At startup or after movement of the viewpoint, objects should be repositioned quickly if needed [21]. Increased time needed to find a better positioning only makes sense if later on a longer amount of time is spent during which the objects are in focus and no further movement is expected [21]. If two consecutive work steps share a common information object, the information object should not be moved. If a movement cannot be avoided, the change should

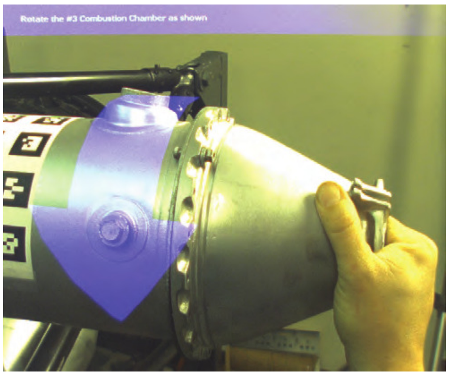


Figure 4.6: The arrow gives a motion cue which indicates how the information object must be moved (source: Henderson and Feiner [137])

be as small as possible and it should be animated such that a user can follow the changes [191]. When an information object is moved while outside the view frustum, the user must be made aware of the new position with the previously discussed measures. Detached virtual information objects should be placed at fixed positions, either relative to the user or the overall scene, where the user can easily find them and no position updates are needed [205].

4.5 Motion Cue Challenge

During the execution of procedural tasks, physical objects in a scene are not static. Instead, the goal of many work steps is to manipulate physical parts, e.g., by installing or removing them, which requires manual actions. Information objects are used in augmented reality support to indicate the needed changes, and describe, with the information they convey, for their physical anchors how and in what form they must be manipulated. Thus, as already described within the characteristic physical change, at least some of the information objects describe manipulations that take place in a three-dimensional space. However, this is not necessarily the case for all of them.

With indirect physical as well as spatial virtual information objects, it is possible to convey this information via the spatial context. However, a single context is not sufficient for this purpose, since it is problematic to represent a progression with only one. Instead, a sequence of them is necessary that corresponds to the spatial and temporal progression of the changes that must be made to the anchors of information objects. These sequences can be displayed in different ways with a greater variety compared to singular contexts. In order to avoid unnecessary cognitive effort, it must be as easy as possible for a user to comprehend the information on a needed change and apply it to physical objects. Therefore, the challenge is to choose the right presentation of these sequences of spatial contexts that to information on manipulations taking place in a three-dimensional space.

With regard to the motion cue challenge, there are several aspects to consider. Due to their limited manipulability, the presentation of such sequences with indirect physical information objects is also limited compared to spatial virtual information objects. Another aspect is that indication of spatial manipulation can also be made indirectly. This is the case when a spatial virtual information object has an indirect physical information object as its anchor, which in turn has a physical object as its anchor. Yet another aspect is that in principle, such a sequence of meaningful spatial contexts can also exist for direct physical objects. However, these cannot be manipulated by an augmented reality system, therefore other possibilities must be found to indicate the manipulation. Thus, one or more information objects have to be defined to convey this information when modeling the scene before the actual representation is considered.

An example is shown in figure 4.6. The arrow indicates a movement to be performed by the user on the combustion chamber. It is a spatial virtual information object that provides a motion cue for a direct physical information object.

This challenge is further divided into two aspects, depending on the kind of measures. One are dynamic cues, which use motion to convey manipulations that take place in a three-dimensional space. They indicate movement by moving information objects by in a scene, i.e., animating them. The others are static cues that translate the dynamics of movements into static representations, with indicators such as arrows or spline curves.

Derivation. The source of possible unnecessary sensory and cognitive effort that is addressed with this challenge is caused by a combination of three characteristics. The first characteristic is physical change, which implies that information objects contain information about required changes of physical objects. Seeing and understanding it should generate as little cognitive and sensory effort as possible for the user, but this can be made difficult by other characteristics. The second characteristic is spatial relationship, which adds that these changes have to be made in a three dimensional world, and every instruction to change an object must be translated into a manipulation that takes place in a three-dimensional space. The third characteristic is combination. It complicates the understanding because the limited field of view may be so small, that not all parts of an indication of a manipulation can be seen simultaneously from a typical viewpoint. This can happen, for example, if objects have to be moved over a distance that no longer fits completely into the field of view at a distance required for manipulation.

Previously described by. To the best of the author's knowledge, this challenge has not yet been described to the extend it is here. However, in contrast to the other challenges, the aspects found for this one were investigated specifically in the context of procedural tasks. Sääski et al. [252], Baldassi et al. [28] and Hou et al. [142, 143] address the user's need to see motion cues for physical objects when working on them. Hou and Wang are specific and describe it as the problem of a user to "acquire the sequent information context" (Hou and Wang [142], 2010, page 6). However, in all previous works known to the author, this challenge is only considered as a challenge of introducing helpful animations of virtual information objects as a means to provide information on movements. Interestingly, there are also examples in the literature where other motion cues have been chosen to represent this information, such as arrows (see, e.g., figure 4.6). To the best of the author's knowledge, these have never been studied in more detail for augmented reality.

4.5.1 Dynamic Motion Cues

Dynamic motion cues include all movements or animations of information objects through which a sequence of meaningful spatial contexts is represented. Typically, a spatial virtual information object follows a path, including the rotations that a direct physical information object must be moved on. An example is a virtual screwdriver that is animated in the scene in such a way that it shows the movement which are necessary with a physical screwdriver. This conveys to a user information on where and how the screwdriver must be hold and turned. Dynamic motion cues cannot be applied to direct physical information objects because these cannot be manipulated by an augmented reality system.

The immediate display of movements with dynamic motion cues supports users in understanding and executing movements [143, 144, 252]. This leads to a reduction of cognitive workload and thus effort [143, 144, 28]. The combination of augmented reality and dynamic motion cues leads to a better execution of movements by the user [28]. Manipulations should be shown from the ego viewpoint because it is easier to transfer them into actions with little cognitive effort [123]. A realistic representation of the movements contributes to a user successfully performing them [140]. In user tests, the use of dynamic motion cues reduced both completion times and error rates compared to paper-based manuals when working on manual procedural tasks [143, 144]. Further, in a study that did not investigate motion cues but augmented reality as a medium to support procedural tasks, promising results were achieved by animating necessary manipulations [136]. Dynamic motion cues can also relieve a person from the burden of mentally performing the necessary rotations of objects. The time required for doing this is linearly dependent on the necessary rotation for both two and three-dimensional objects [67, 265]. For 60 degrees, it is about one second, but larger variations occur depending on the actual shape of an object [265].

4.5.2 Static Motion Cues

As an alternative to dynamic motion cues, it is also possible to indicate motion statically, e.g., by using arrows or splines to show how physical objects shall be moved. Using these static motion cues means to add new virtual information objects to a scene or to modify existing ones. Such cues have the advantage of being temporally invariant, so a user does not have to wait for an animation to end or to remember individual sections of an animation. However, the disadvantage is that the concrete movement is never shown. A combination with dynamic motion cues is of course possible to ameliorate this problem. For example, arrows used to indicate a motion can be moved themselves. Other than dynamic motion cues, static motion cues can also be applied to direct physical information objects by adding a new information object to indicate the motion.

Arrows, e.g. [137, 237], and graphical paths or splines, e.g. [137, 136], have already been used in augmented reality systems to represent movements. While some of the published applications have also been subjected to user tests, e.g. [137, 136], the author is not aware of any systematic investigation of static motion cues for augmented reality support of procedural tasks. However, static motion cues are often used in instructional graphics and some information can be gathered from literature on this topic.

Furthermore, for augmented reality, there is research on displaying navigational instructions in vehicles with head-up displays that indicate in which direction a vehicle shall be steered. Elements such as arrows placed on the road or a virtual guide cable floating above the road were already investigated [224, 296]. It was found that spatially correctly registered cues, i.e., spatial virtual information objects in the language of the IRMAR framework, performed better than unregistered cues [223, 296]. It was also shown that splitting large cues such as arrows into smaller sections whereby the motion information was repeated locally was helpful to understand these cues [223]. One reason for this is the limited field of view of head-up displays, in which in many cases only parts of a larger cue are visible at any time. This leads to a significant loss of information that can be avoided by splitting it into smaller parts and thus locally repeating the information. However, this is a rather different use case and therefore applicability of these results to the support of manual procedural tasks is not ensured.

For static illustrations, various ways have already been developed to depict movements with them [76]. One possible representation of movement is to attach arrows to information objects that indicate the direction of movement [76, 16]. However, they were not developed for use in augmented reality and therefore no empirically validated knowledge is available on them. Even though these have already been used in augmented reality applications, e.g. [237], to the best of the author's knowledge their benefit has never been examined in detail.

Also used in static illustrations is the depiction of one information object in different stages of a movement such that a user sees a sequence of objects describing the path and the poses of the object [76]. When this presentation becomes so dense that the instances touch each other, it is called overlapping multiples [76]. Another form of static motion cues are guide lines, which represent a path an object has to be moved on [76, 16]. The associated information object can be integrated at one point of this line in order to establish a connection between the two. To be able to show a chronological order unambiguously, transitional thickness or gradual density can be used, e.g., by increasing line thickness or color saturation according to the time sequence of the movements [76]. A more detailed overview on motion representation in illustrations can be found in the work by Agrawala et al. [76].

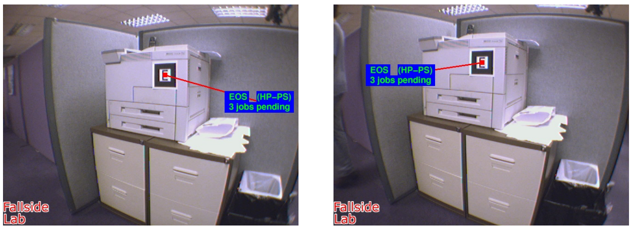


Figure 4.7: Direct lines can show a connection between two information objects independent of their realtive position (source: Rosten et al. [251])

Even though dynamic and some static motion cues were presented here, a more detailed investigation on the use of the different methods for conveying motion in augmented reality is still missing. No comparison of static and dynamic motion cues currently exists to the best of the author's knowledge.

4.6 Information Linking Challenge

As already described by the characteristic connectedness, information objects are semantically connected with each other. This is modeled via anchors which information objects can have. To support a user understand these connections, it is important to visualize them such that they are easily recognizable. A first approach might be to simply position objects close to each that are connected. However, this is not always feasible because of multiple reasons, e.g., because information objects might occlude each other. Thus, different ways of showing connections have to be found to convey this information in scenes where the trivial approach is not possible. For this, it is important that minimal sensory and cognitive effort is caused on the user's side. Effects such as insufficient tracking complicate this challenge because a user needs to compensate for these errors if the presentation of connections is very sensitive to them.

Figure 4.7 shows an example of how the information linking challenge can be solved by using a direct line, which is a type of continuous visual connection. In this form of presentation, the connection is recognizable regardless of the relative spatial positioning of the two information objects. If, for example, a spatial assignment is chosen, this would not be the case.

This challenge is discussed in depth in the following chapter 5 Information Linking. Therefore, only the derivation and previous descriptions of the challenge are given here, while further details are discussed in the following chapter.

Derivation. The source of possible unnecessary sensory and cognitive effort that is addressed with this challenge is caused by a combination of four characteristics. The first characteristic is connectedness. As described by it, information objects are connected with each other and thus it must be easily recognizable which connections there are. This is, however, made difficult by other characteristics. The second characteristic is spatial relationship, which indicates that information objects are located in a spatial structure based on the physical world. This means that they may not always be positioned such that the connections are unambiguously and easily recognizable. The third characteristic is combination which implies that information objects and their connections from a scene can only be seen in a combined view that depends on the viewpoint, and is constrained by a view frustum. Particularly, when the field of view is small or the viewpoint is at an unfavorable position, it is not always possible to clearly see all connections. Moving the viewpoint to obtain a better view of the connections in a scene involves additional effort by the user and is thus not desirable. The fourth characteristic is fluctuation. When the effects described by fluctuation become too strong, it may become difficult to recognize the connections when virtual and physical objects are not correctly aligned.

Previously described by. For augmented reality, this challenge was first described in the author's work [200]. Even though labeling has been intensively researched for augmented reality, e.g., [21, 36, 191, 266], the focus was not on showing connections, but on how best to arrange labels. However, for fully virtual environments without any physical parts, Polys et al. [54, 227] described the challenge that connections between virtual objects must be understood by users and investigated possible visualizations to support this.

4.7 Conflicting Resolution Measures

While the measures described for the six challenges can help designers in creating augmented reality systems supporting procedural tasks, they do have limitations. One of them is that many combinations of them are contradictory and thus, using a measure to solve one challenge can aggravate another challenge. Brennecke and Keil-Slawik [55] call this a design conflict and describe is as a contradiction between different needs respectively requirements concerning the user interface. While for some of these, a general solution can be found, for many, a decision as to when which measure must be applied cannot be made in general, but only depending on the respective usage situation [157]. This implies that except for simple cases, it is hardly possible to solve all challenges in a real world augmented reality system to the same extent. Instead, at least some, if not all, will remain not fully solved. In this case, it is up to the respective system designers to find solutions that are as good as possible for the individual system. In the following, a few examples for design conflicts in the context of IRMAR are given and discussed.

A relatively simple example can be found for the combination of clarity and consistency challenge. If a spatial virtual information object with text is positioned far away from a user, it will be hard to read because it becomes too small. To solve the clarity challenge, a designer might remove it from its spatial context and position it such that its more easily readable. However, its relative size is a depth cue that contributes to solving the consistency challenge. Thus, this solution will remove depth cues and make it harder for a user to understand where in the scene this object is actually positioned. Additionally, the object's meaningful spatial context is lost. For virtual environments, Polys et al. [226] have shown that in such a situation there is no perfect resolution by a one-sided emphasis on solving one challenge. Neither perspective-correct scaling nor ensuring legibility was superior at all times.

Another example is the visibility challenge, for whose full solution no information object may occlude another one. One possible way to achieve this is to make the occluding ones partially transparent. However, this negatively affects any solution of the clarity challenge because for any overlapping parts of information objects the contrast will decrease and colors will blend [181]. As an alternative to partial transparency, the information objects can be relocated so that there is no more overlap [36]. This in turn negatively affects any solution to the information linking challenge. The reason is the increased distance of information objects and their anchors, which makes it more difficult for a user to recognize any connections [200]. As a result, at least one of the three challenges - visibility, clarity or information linking - cannot be completely solved. Instead, for a concrete solution the relevance of the individual challenges in a usage situation has to be weighted against each other and a reasonable compromise has to be found on this basis.

Another, albeit slightly different, example is the video freeze interaction [173] for hand-held devices. When triggered, it freezes the video to a static image, including the virtual objects currently shown on the display. Until the user returns to a live video feed, the same image is displayed continuously. This may be useful, for example, in situations where relevant information objects or connections between them are only visible in the combined view from otherwise unfavorable viewpoints. This involves briefly moving to such a viewpoint, freezing the combined view, and then moving to another location. The challenges - here the visibility and the information linking challenge - are therefore solved through interaction, which causes effort for a user. Further, it can have a negative impact the orientation challenge, because the frozen video is harder to relate to the scene the more the overall scene or the viewpoint change. This effectively leaves it up to the user to resolve the contradiction of which challenge is more important to solve.

As these three examples show, it is impossible to always solve all challenges completely. However, this is not an aspect specific to augmented reality but can be also be found in all types of user interfaces [55]. Instead, a solution has to be identified that is sufficiently adapted to the respective usage situation in order to ensure a good information representation. One approach that is often pursued is the use of patterns which are general, reusable designs, that each represent a solution to a recurring problem [157, 284, 49]. Their distinctive quality is that they combine and balance different requirements, called forces in this context, where there is no perfect solution. Depending on the pattern, the individual forces are given greater or lesser weight in a solution such that different alternative patterns can emerge for similar problems in other usage contexts. A collection of patterns that are linked together is called a pattern language [49]. In the context of these challenges, such a language could be a useful approach to simplify the design of information representation in augmented reality. For this purpose, the challenges would have to be modeled in these patterns as forces for which it is necessary to determine the extent to which they must be solved.

5 Information Linking[1]

In this chapter, the information linking challenge identified in the previous chapter is examined in more detail. For this purpose, a taxonomy is proposed for systematizing different forms of visualizations for information linking. Further, two studies with human subjects are conducted to identify a suitable way of visualizing connections such that sensory and cognitive effort is minimized. This contributes to achieving the goal defined in the motivation to develop measures that help solving the representation challenges.

Of all the identified representation challenges, information linking is the one that appears to be least explored. While there are already relevant research results in the field of purely virtual environments, it has previously only been marginally considered for augmented reality. Therefore, the challenge is examined more closely in this chapter. For this purpose, possible ways in which a connection can be presented to a user are investigated in order to contribute to possible solutions for this challenge. In the following, these ways are referred to as visual information linking or short information linking.

First, a review of information linking in information-rich virtual environments is made, where there are already relevant previous works on the general structure of the subject area and on human factors. These are examined with regard to the aspects that can be applied to the information linking challenge in augmented reality.

Next, a taxonomy for information linking in augmented reality that comprises three dimensions is proposed. This is done to create a structure that summarizes the basic decisions when designing visual information linking. Concerning the IRMAR framework, this creates an opportunity to describe and classify measures for the information linking challenge in a structured way, as well as the possibility to integrate and organize existing works. Only those forms of information linking are considered in the taxonomy where a connection is visually established.

[1] This chapter combines results from two previous publications to which I was the main author respectively co-author. These two publications are presented in this chapter and their results are integrated into the IRMAR framework. The first one is *A Taxonomy for Information Linking in Augmented Reality* [200] and the second one is *An Evaluation of Information Connection in Augmented Reality for 3D Scenes with Occlusion* [73]. Co-author was Ralf Dauenhauer, a master's student whose master's thesis I supervised on the test system used and the execution of the empirical studies to investigate this challenge [72]

When, for example, a connection can only be derived from a textual description or through interaction, it is not included. This would contradict the objective of the IRMAR framework to leave out semantics of information objects as well as interaction aspects. For each of the dimensions, a description, an explanation including examples and an overview of human factors, are given. To show its applicability and to deepen its understanding, several augmented reality systems from scientific publications are classified.

Through the structure of the taxonomy, gaps in the understanding of human factors become evident. In order to at least partly close them, two empirical studies with human subjects were conducted, which are presented here. These were designed to help determine how connections between virtual and physical information objects have to be presented. The first study compares different visualizations for information linking on the three device types handheld devices, video see-through glasses and optical see-through glasses. The second one extends this comparison on a handheld device to a more complex scene with varying distances of information objects and occlusions. Through these two studies, the understanding of how information objects can be displayed as connected with each other is substantially advanced and possible measures for the information linking challenge are developed.

5.1 Information Linking in Information Rich Virtual Environments

In contrast to augmented reality, the topic of information linking has already been investigated in more detail for information rich virtual environments (IRVE). These are purely "virtual environments [that] consist not only of three-dimensional graphics and other spatial data, but also include information of an abstract or symbolic nature that is related to the space." (Bowman et al. [52], 1999, page 1). Relevant in this context is their similarity to augmented reality in terms of their three-dimensionality and the embedding of various forms of information into a three-dimensional space. A scheme to classify the way information is linked in an IRVE was proposed by Bowman et al. [54]. Even though with IRVEs they consider a different application domain, their work pursues a similar goal as is done here with the taxonomy for information linking in augmented reality. Therefore, it is presented in more detail and its applicability to a taxonomy for information linking in augmented reality is discussed. Three dimensions are contained in their classification scheme that represent decisions a designer has to make when deciding how abstract information should be embedded:

- *Display Location*: The place or reference frame where abstract information is located. It can be either world-fixed, display-fixed, object-fixed or user-fixed.
- *Association*: The way abstract information, e.g., labels, is connected to perceptual information, e.g., three-dimensional objects in the virtual environ-

ment. It can be either spatially explicit (association is presented via spatial proximity), visually implicit (association is presented via a joint highlighting during interaction) or visually explicit (association is presented via a connection line or a similar graphical element).

- *Level of aggregation*: How much abstract information is combined into one visualization. It can range from simple pieces of information to complex visualizations with a high information density.

Bowman et al. differ in some points from what is intended to be achieved with this work. First, they are concerned with how abstract information can be integrated into an IRVE and connected with objects located in it, while the focus here is exclusively on how connections between information objects can be conveyed. Therefore, there are also some limitations in the transferability of their results. The dimension association by Bowman et al. is not fine-granular enough to sufficiently classify different visualizations that can be used to establish a connection. This applies in particular for the analysis of human factors, since, for example, no distinction is made which form a visually explicit connection has. Nevertheless, there must be a comparable dimension in the taxonomy developed here. The situation is a bit different with regard to their dimension display position, which offers several categories of locations where an object can be positioned. Since the positioning of objects obviously has a significant influence on how easily a user recognizes connections between them, a comparable dimension must be included in the here created taxonomy. However, other possible classifications have also been proposed [98, 287, 250], so that there are already various possibilities for the concrete categories. One aspects that is completely missing from this dimension, as well as the alternative categorizations, is the orientation of an object. Unlike the last two, the dimension aggregation level cannot be transferred, since the semantic content of information objects is not considered in IRMAR. In conclusion, this classification scheme provides a basis that must be incorporated into the development of a taxonomy for information linking in augmented reality.

5.2 A Taxonomy for Information Linking in Augmented Reality

In the following, a taxonomy is developed and presented which describes the design space for visual information linking in augmented reality. The taxonomy focuses on connections of virtual information objects to their physical anchors, but it can also be applied to connection between other classes of information objects, e.g., connections from spatially referenced to spatial virtual information objects. The classification made with it covers a combination of one information object and one anchor. Thus, if an information object has several anchors, the classification is done for each of the connections individually. For readability reasons, throughout the remainder of this chapter the terms information object or anchor refer to one of the two elements involved in the connection, when no further context is given. As in the other parts of the IRMAR framework, the concrete design of the information objects or their semantics are not considered.

In total, the taxonomy consists of three dimensions. These stand for design decisions, which must be made when displaying information linking, and the individual classes represent basic directions the decision can take. The first dimension is reference systems, which correspond to the decision what the reference for orientation and positioning of an information object should be. It is closely related to the characteristic reference systems, which states that there are various sources for such a reference. The second dimension is visual connection, which corresponds to the decision of how a connection should be graphically presented. Finally, the third dimension is context, which corresponds to the decision if and how a graphical context should be included to support conveying the connection.

One challenge in creating such a taxonomy is to find the balance between a very specific collection of details, which may be relevant only in a few cases, and a comparably general structure, which may miss these details. Here, the approach is to keep the taxonomy rather general, even at the risk that individual details are not included. As an advantage, it is relatively compact and clearly structured. In order to be able to provide the relevant information regarding human factors, their description addresses these details and also describes relevant aspects that may differ within a class.

5.2.1 Dimension: Reference System

The first dimension of this taxonomy is *reference system*. It classifies in which reference systems an information object is positioned and oriented, i.e., the the mounting of information objects. Its relevance for the information linking challenge results especially from the need that a connection between a virtual information object and its anchor must be easily identifiable across the different reference systems, which they may have. Two types of systems are distinguished: the world coordinate systems (WCS) that are bound to the overall scene or an object inside of it, and the spectator coordinate systems (SCS) that are bound to the user's point of view, the user's body or a user-controlled device, e.g., a tablet[2]. These four combinations of position and orientation are possible:

- *WCS Positioned / WCS Oriented*
- *WCS Positioned / SCS Oriented*
- *SCS Positioned / WCS Oriented*

[2]The naming as coordinate system and not as reference system, i.e., world reference system or spectator reference system, has the reason that the technical aspect of object positioning at coordinates was especially considered when creating it. Although this is no longer relevant, the naming has been kept in order to remain consistent with previous publications.

- *SCS Positioned / SCS Oriented*

Some of these combinations of orientation and positioning have established names. A combination of WCS positioned and WCS oriented is called world space, while a combination of SCS positioned and SCS oriented is called screen space [226]. WCS positioned and SCS oriented information objects are called billboards [287].

Similar concepts to classify reference systems have been previously proposed. From the already described classification scheme by Bowman et al. [54], it corresponds to the dimension display location, whose possible classes are world-fixed, display-fixed, object-fixed and user-fixed. Similar classification schemes were proposed by Feiner et al. [98], with the classes surround-fixed, display-fixed and world-fixed, and by Röltgen and Dumitrescu [250], with the classes user, object and environment. Also, Tönnis et al. [287] include a comparable dimension, mounting, as part of their taxonomy for information representation in augmented reality. It describes the reference system to which an object is bound with the possible classes human, environment, world and multiple mountings.

While the distinction into different reference systems for positioning by all of the classes is more fine-grained than the one used here, none of them distinguishes between orientation and position. The reason for the different approach used for this taxonomy is to keep the differentiation in the taxonomy simple and comprehensive. There are various possibilities to define classes of reference systems for information objects, which may also depend on factors such as the use case and the used display type. In order not to become too specific or too wide-ranging, no individual classes for the scene and for objects within the scene are included as origin for a reference system. However, the orientation of an object is a fundamentally different aspect, therefore it was specifically included here, while this is not the case in the other approaches. The possible reference systems classes for orientation are limited to the same two as for positioning to remain concise.

Example. An augmented reality application is running on a handheld device with a video display and shows an information object. There are several ways in which the object can move relative to the device screen. If the device is moved, but the information object remains at a fixed position in the scene, i.e., moves relative to the screen, the object is WCS positioned. If instead the object remains at the same position on the screen and moves relative to the scene, it is SCS positioned. When the information object is WCS positioned and the device is rotated around this position, there are two ways in which the orientation of the object can relate to the screen. If it maintains the same orientation with respect to the scene, i.e., behaves like a physical object placed somewhere in the scene, the object is WCS-oriented. If, however, the information object continuously shows the same side on the screen and thus changes its orientation with respect to the scene, it is SCS-oriented.

Human factors. Especially for IRVEs, studies on positioning of information objects have already been carried out. However, in these, only situations were considered where, in the language of IRMAR, spatially referenced virtual information objects were connected with spatial virtual information objects. Furthermore, it must be noted that IRVEs are not augmented reality and therefore it is unknown whether there are restrictions on the transferability of the results. In augmented reality, this aspect has achieved less attention. To the best of the author's knowledge, only one applicable study was published where reference systems are examined for information linking as part of temporal coherence strategies [191].

Polys et al. [227] found that when using a small field of view for comparison and search tasks, SCS positioning of information objects is preferred in terms of completion time, correctness and user acceptance for IRVEs. This is particularly the case when a wide software field of view is used, such as a scene shown as seen through a wide-angle lens. For an enlarged field-of-view which spanned over multiple monitors, the effect seems to be reversed. While no significant difference was detectable in a study conducted by Polys et al., the tendency was that when objects were positioned in WCS, test persons performed better. This result seems to be because positioning information objects in SCS prevents a guaranteed proximity to their anchor, so a user has to mentally overcome the distance. If the information object and the anchor are far away from each other, this might even lead to an inversion of the law of Gestalt proximity, i.e., that users perceive the objects as not belonging together due to the great distance [227]. However, the conclusion that SCS should be used for small screens and large WCS positioning cannot be clearly drawn, because SCS positioning enables screen management that can support the presentation of information linking by assigning objects fixed positions relative to the display respectively user. This may be even be more important in some cases than the immediate proximity of an information object to its anchor [226]. A similar study to the one by Polys et al. was conducted by Chen et al. [64]. They evaluated two text layout techniques: within-the-World Display (WWD), here classified as WCS positioned / WCS oriented, and Heads-up Display (HUD), here classified as SCS positioned / SCS oriented. In the WWD, condition texts were placed in the virtual environment in close proximity to their anchor with an orientation based on the anchor's orientation. In the HUD condition, the texts were fixed to the user's viewpoint and were connected to their anchors through lines. Their results showed that the HUD condition was generally favorable for search tasks. The authors attribute this to the layout that increases legibility of the texts.

Specifically for augmented reality systems which use tablets as handheld devices, a study conducted by Madsen et al. [191] found that users perform best when labels are part of the scene, i.e., WCS positioned / WCS oriented, and are near their anchors. While this may seem like a difference to the previously discussed studies, it is not. Instead, the good performance can be explained by the stable layout relative to the anchor chosen for this condition. This is also supported by a further result from their study, as they found that when SCS positioned / SCS oriented labels were used, users performed better if their positioning on the screen was algorithmically stabilized such that the relative order of labels was kept stable.

For WCS positioned objects, depth cues have to be given in order to support a user in estimating the correct position and easily identifying the correct anchor [226, 118]. This becomes especially relevant when no explicitly graphical connection is shown, and instead cues like proximity are used to indicate information linking.

5.2.2 Dimension: Visual Connection

The dimension *visual connection* classifies the way in which connections from an information object to one of its anchors are visualized and describes which graphic means are used to make the connection visible. The following classes are distinguished:

- *Spatially assigned*: An unambiguous relative positioning of information object and anchor is used to show the connection
- *Continuously connected*: A graphic connection, e.g., a line or an arrow, that leads from the information object to the anchor. Small interruptions, e.g., like spaces in a dotted line, are included as long as an unambiguous continuation is visible. Further subclasses are distinguished by how they support keeping these graphic connections apart:
  - *None*
  - *Color*
  - *Style*
  - *Both*
- *Symbolically connected*: A connection that is visualized symbolically, e.g., by having each information object and its anchor share the same color. This means that the user must recognize the information linking on the basis of a specific visual coding. Subclasses are based on the type of coding that is used:
  - *Color*
  - *Shape*
  - *Both*

For better readability, the class names are not always referred to directly and instead, substantiations are used. Thus, instead of spatially assigned, the phrase spatial assignment is used; instead of continuously connected, the phrase continuous connection is used; and instead of symbolically connected, the phrase symbolic connection is also used.

The similarity to the Gestalt laws is noticeable for these classes. For example, one can consider spatial assignment as an application of the law of proximity or a continuous connection as an application of the law of connectedness. This is not surprising, since this dimension is intended to classify the graphical means to present connections of information objects to their anchors. Fittingly, part of the Gestalt laws describes shapes that humans recognize as being connected to each other [247]. However, not all Gestalt laws are appropriate, such as the law of good continuation, which is not applicable in this context. The difference is that they describe perceptual phenomena, while with this dimension, classes of visualizations are distinguished from each other that may utilize these phenomena.

For spatially assigned connections, in combination with an SCS positioned information object, it may not be immediately obvious what a presentation can be like. There is the apparent contradiction that a connection is to be represented by spatial proximity while the virtual information object is simultaneously bound to the user. However, an assignment does not have to be created via proximity, it only has to be unambiguous. For example, when on a handheld device, one physical anchor is in the middle of the display and another one is on the right side, the corresponding information objects could be positioned at the bottom of the display, at the middle, and at the right side. The relative positioning guarantees an unambiguous assignment, while anchor and information object are still positioned in two reference systems.

Unlike the other two dimensions, reference system and context, different alternatives can be used simultaneously to show the connection. For example, it is possible to display information objects in the immediate proximity of their anchor and still use a common color coding to emphasize the connection.

Example. A label provides additional information on a direct physical information object as its anchor. When it is WCS positioned and placed closely to its anchor such that an unambiguous assignment is made, the visual connection is classified as spatially assigned. When the label is placed at a distance and a connection line is used to show to which anchor the label belongs to, the visual connection is classified as continuously connected with the subclass none. However, if there are several labels and line colors are used to support a user to distinguish between the individual connecting lines, then the subclass is color. If instead of a connection line, a star is shown next to the anchor and next to the label to mark them as connected, the class is symbolically connected with the subclass shape.

Human factors. To the best of the author's knowledge, no specific publications from other authors concerning human factors of visual connection exist for augmented reality. Therefore, alternative sources have to be consulted.

For view management in augmented reality, it is usually assumed that an information object should be as close as possible to an anchor [36, 191]. For IRVEs with large screens, this is supported by empirical studies [227]. Proximity is also one of the strongest Gestalt laws [247]. With smaller monitors, however, other factors predominated concerning the understanding of connections. A fixed position of an information object over time seems to be more important, at least partly, than proximity [191, 226].

In the continuously connected class, direct connecting lines are often used for visualization, e.g., by Azuma and Furmanski [21, 36], even though other forms are possible as well. The shape even has a direct impact on task performance. For example, Polys et al. [226] found in IRVEs that for search tasks, polygons in the shape of a triangle performed best, while for comparatively more complex comparison tasks simple connecting lines led to best results. Semi-transparent polygons are a compromise that performed relatively well for both task types. Chen et al. [64] also found that in IRVEs connection lines are an appropriate means to show connections of information objects to their anchors. The fact that the human fovea only covers a few degrees of the visual field, and thus eye movements are necessary to follow a line or comparable shape suggest that visualizations in the class continuously connected have disadvantages compared to a direct spatial proximity. However, a direct connection conforms to the law of uniform connectedness, which is comparably powerful [217]. Furthermore, lines can be followed very will with saccadic eye movements as a part of scanpaths and support finding a target [209, 208]. This suggests that continuous connections are also a strong form of connection. When using visualizations from this class, information objects should stay close to their anchor because every kind of visual connections becomes more difficult to follow with greater distance [127, 271].

When a visual connection from the class symbolically connected is used and information object and anchor are further away from each other, a user must perform an activity called a visual search [288]. This means that a target stimulus has to be found among other stimuli, termed distractors. The target stimulus is the coding to be found, e.g., a certain color, and the distractors are other, distracting stimuli which can be any other colors, shapes etc. that are also part of the scene. Such a search becomes less efficient when differences between the target and the distractors become smaller [309]. The number of targets, however, scales well [230]. Thus adding new information objects including codings for their anchors only degrades people's search performance to a limited extent. Overall, search time has sublinear growth with respect to the number of stimuli [230]. While cognitive processes for an initial selection may run in parallel, a second part for a final selection can only be executed sequentially. Therefore, the increase in time required when further complex stimuli are added is still significant [309]. Colors in a visual search task seem to be easier to find than shapes, but this is highly dependent on the difference of the colors, and respectively, shapes for the visual search task [309, 230]. For parts of a visual information linking, that are supposed to be near or at the anchors position, depth cues must be given, especially for non-trivial scenes [226, 118, 86, 168]. Especially when symbolic connections are used, it otherwise becomes difficult for a user to recognize an connection correctly.

5.2.3 Dimension: Context

Context is a stylized representation of the physical world included as part of a computer-generated image [245, 155]. Thus, the *context* dimension is used to classify whether a graphical context is visualized around the anchor, around the information object, or around both. The following classes are distinguished for this dimension:

- *Both without context*: A context is neither shown around the information object nor the anchor
- *Anchor with context*: A context is shown around the anchor but not the information object
- *Information object with context*: A context is shown around the information object but not around the anchor
- *Both with context*: A context is shown around the information object and the anchor

If the information object is WCS positioned and located at the position of its anchor, the context of both elements merges and the assignment of the context can no longer be determined. In this case, only the classes both without or both with context are possible.

Example. A virtual model of a screw is being used to show how its physical counterpart has to be screwed into a drill hole, and it is displayed with some offset next to the hole. If the surroundings of the drill hole are displayed around the model of the screw, e.g., to clarify the exact position of the screw, this presentation is in the class information object with context. If, on the other hand, the context of the drill hole is displayed around the drill hole, e.g., to mitigate the consequences of inaccurate tracking, it is in the class anchor with context.

Human factors. Depending on where the context is used, it has different effects. A context displayed at the anchor's location makes it easier for a user to compensate for tracking errors [245]. By replicating structures of the environment or other information objects, it is easier to see to which extent virtual objects are misplaced relative to the physical world. This can make a difference between helpful and non-helpful augmented reality support [245]. But even if tracking works near perfect and does not create any ambiguities, a user's confidence in the representation is increased through context [245]. Further, a context can also avoid depth ambiguities that make it difficult for a user to determine exactly where the anchor is [29]. This solution, even though it contributes to information linking, can also be seen as a measure for the consistency challenge.

When a context is displayed around the information object, it helps the user to better recognize the link between an information object and its anchor [244]. The graphical context makes it easier for a user to understand which position in a scene it is connected to. This is especially helpful when the information object

Figure 5.1: Information Linking Example 1 (Source: Bell et al. [36])

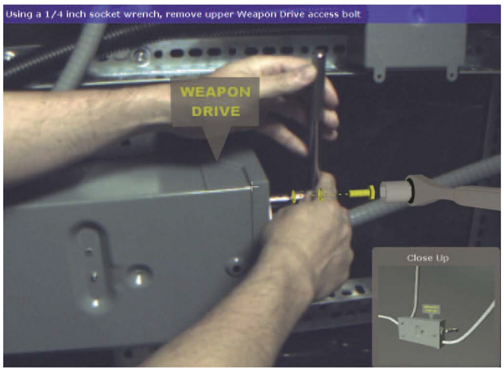


Figure 5.2: Information Linking Example 2 (Source: Henderson and Feiner [136])

is further away from its anchor or, for a spatial virtual information object, is not positioned in a meaningful spatial context. In case an information object is displayed at the position of its anchor and it is no longer possible to distinguish to which part the context is assigned to, a context supports the two aspects addressed simultaneously [244, 245].

5.2.4 Example Classifications

To show the applicability of this taxonomy to real applications, and to give examples of how diverse visual information linking can be designed, different systems are classified in the following. For this purpose, scientific publications were selected that describe systems containing information linking in augmented reality. However, different forms are used even within one system and the description of the user interface is often not available in its entirety. Thus, individual images are taken from these publications, analyzed, and classified depending on the available information. Only parts that are relevant for information linking are described and other aspects, e.g., discussions around which parts are information objects and which are not are left out in order to keep this classification concise.

In the first example (figure 5.1), two different types of visualization are used for information linking. The text labels are all spatially referenced virtual information objects, for which the buildings serve as anchors. For both variants of the linking, the classification for the dimension reference system is WCS positioned / SCS oriented, because all labels are placed somewhere in the scene while they are oriented towards the viewpoint. Concerning the visual connection dimension, the directly superimposed labels are spatially assigned because the connection to the building is conveyed through their position. The labels connected by lines are continuously connected, whereby the subclass is none since the lines do not differ in shape or color. Regarding the dimension context, anchors and information objects do not have any context, so classification is both without context.

The application from the second example (figure 5.2) shows two or three dif-

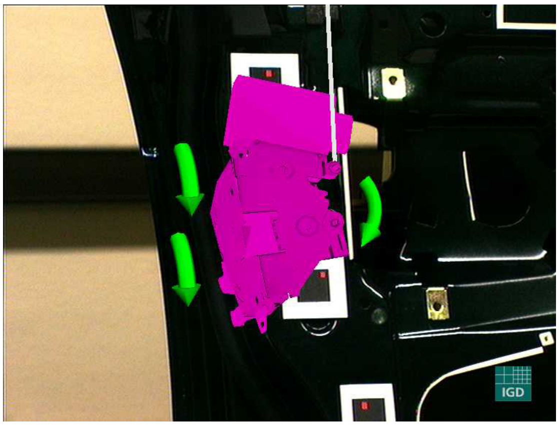


Figure 5.3: Information Linking Example 3 (Source: Reiners et al. [237])

ferent types of visual information linking. The first one is used for the label with the text weapon drive. This is a spatially referenced virtual information object whose anchor is the weapon drive towards which the tip points. The classification in the dimension reference system is WCS positioned / SCS oriented because the text seems to be oriented towards the user's viewpoint. Concerning the dimension visual connection, the classification is continuously connected since the actual label is positioned a little away from the anchor and the connection is visualized with by an arrow. The subclass cannot be identified because only one text label is displayed here. Neither the anchor nor the information object includes a visual context, thus the classification in the dimension context is both without context. The second type is used for the models of the two washers and the bolt. These are all spatial virtual information objects whose anchor is their mounting position at the thread hole. The classification in the dimension reference system is WCS positioned / WCS oriented. Their visual connection is continuously connected via the yellow dotted line, whereby the subclass cannot be defined here either. Again no context is used, so the classification in the dimension context is both without context. The type of information linking for the wrench model is not completely clear. It is clearly WCS positioned / WCS oriented and does not have any context. For the dimension visual connection, however, the classification depends on what one assumes to be the anchor. If the virtual model of the bolt is taken, it is spatial assignment. If the physical wrench is taken instead, there is no visual information linking at all and thus no classification.

Only one type of information linking is used for the application from the third example (figure 5.3). All four virtual objects are spatial virtual information objects. The anchor of the door lock model (purple) is the door frame, while the anchor of the arrows is the model. All virtual information objects are WCS positioned / WCS oriented. The visual connection is created by spatial assignment and a context is not present, therefore they are in the class both without context.

In the upper image of the fourth example (figure 5.4), the model of a sensor and

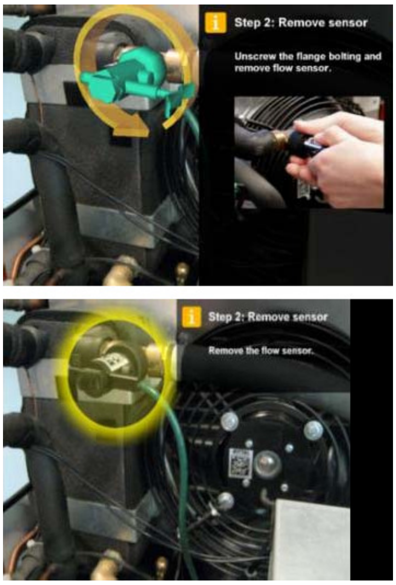


Figure 5.4: Information Linking Example 4 (Source: Webel et al. [301])

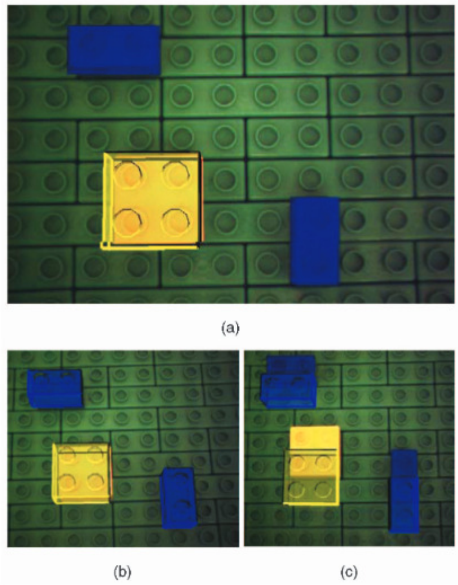


Figure 5.5: Information Linking Example 5 (Source: Robertson et al. [245])

the virtual arrow are both spatial virtual information objects, which are connected to a sensor behind them as their anchor. Both virtual objects are WCS positioned / WCS oriented, matching the position and orientation of the physical sensor. The visual connection is achieved via spatial assignment. A context is not used, thus both are without context. The general information on the current work step displayed at the right side is a detached virtual information object and no connection exists that would have to be taken into account. The lower image in the fourth example (figure 5.4), on the other hand, is not conclusive. Two possible interpretations are whether no visual information linking is used or whether the highlighted sensor is a visual information linking to the description text on the right. In the second case, the description text is a spatially referenced virtual information object whose visual connection to the anchor is symbolic. However, a subclass cannot be determined here, since there is only one connection, so that no differences are visible. No context is used, so the classification is both without context.

The fifth example (figure 5.5) shows three views from an application that supports setting toy blocks via augmented reality. In all views, the information object for which information linking is made is the virtual yellow brick in the middle. Its anchor is always the position where it has to be placed. It is a spatial virtual information object that is WCS positioned / WCS oriented. For the visual connection, spatial assignment is used in all the cases. However, the representations differ concerning the use of context. While the representation in figure 5.5a is in the class without context, figures 5.5b and 5.5c both contain context. Due to the positioning of the information object, it is not possible to distinguish between a context around the anchor and around the information object. Consequently, this results in the class both with context in both cases. Figure 5.5c also clearly shows how the context helps to set the block correctly. Without context, a user would have to assume that the block has to be placed one line further down and thus not recognize the correct connection.

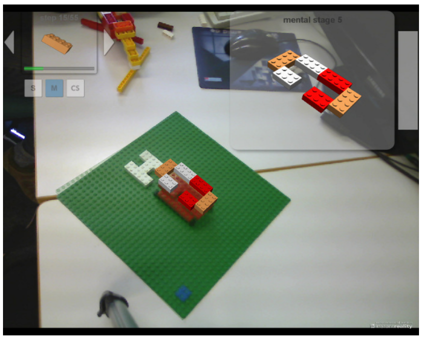


Figure 5.6: Information Linking Example 6 (Source: Engelke et al. [95])

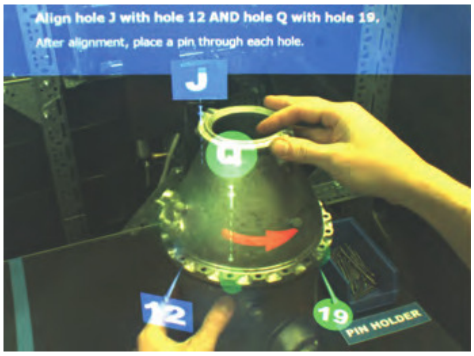


Figure 5.7: Information Linking Example 7 (Source: Henderson and Feiner [137])

The sixth example (figure 5.6) shows a view from another system that also supports setting toy blocks with augmented reality. The virtual brick models displayed above the green base plate are spatial virtual information objects. Their anchors are the positions at which the physical bricks must be positioned. Regarding the dimension reference system, they are all classified as WCS positioned / WCS oriented. The visual connection is done by spatial assignment. To create a context, the base plate is included in the virtual graphics. However, it cannot be distinguished whether the context refers to the anchors or the information objects, therefore it is classified as both with context. The models at the upper left and upper right are detached virtual information objects which do not have anchors, so they do not need to be taken into account here.

In the seventh example (figure 5.7), two different types of visual information linking are used. The labels with the letters and numbers are spatially referenced virtual information objects whose anchors are located at the positions of the, albeit hardly visible, blue or green dots. Since they are positioned in the scene, but seem to be always oriented towards the observer, they are classified as WCS positioned / SCS oriented in the dimension reference system. Regarding the visual connection, they are continuously connected through the dashed lines. Even though colors are used, the subclass of the visual connection is none, since they are not used to distinguish the connections but the classes of information on the labels. Context is not used and thus the classification is both without context. The red arrow is a spatial virtual information object whose anchor is the movable metal component. With respect to the dimensional reference system, it is classified as WCS positioned / WCS oriented since its pose is aligned with its physical anchor. In the dimension visual connection, it is classified as spatially assigned, since its connection to the anchor is conveyed through its position. Again, no context is used and thus he virtual bolt models are spatial virtual information objects, whose anchors are the same points that are the labels' anchors. Their information linking is identical to the arrow in all dimensions, and therefore not repeated.

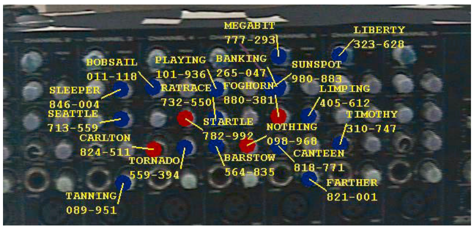


Figure 5.8: Information Linking Example 8 (Source: Azuma and Furmanski [21])

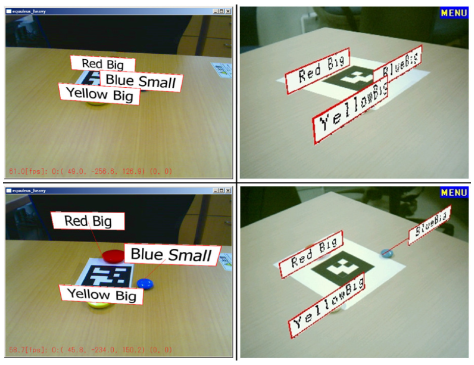


Figure 5.9: Information Linking Example 9 (Source: Shibata et al. [266])

In the eighth example (figure 5.8), text labels, which are spatially referenced information objects, are used to describe their anchors. These are rotary knobs that are highlighted by blue and red dots. Concerning their reference system, the information objects are classified as WCS positioned / SCS oriented because they are positioned relative to their anchors but oriented towards the viewpoint. Their visual connection is a continuous connection with the subclass none, since lines are used that all have the same color and shape and thus offer no further differentiating features. Even though the labels are close to the anchors, it is not possible to establish a connection from this, thus spatial assignment is not used. Context is not included for information objects or anchors such that the class is both without context.

The ninth and final example (figure 5.9) shows several text labels which are spatially referenced information objects. Their anchors are the red, yellow and blue stones. Their reference systems are WCS positioned / WCS oriented. The WCS orientation can be seen clearly in the two right images, where the labels are not oriented towards the viewpoint. The difference between the upper two and lower two images is that in the upper images, the visual connection is established by spatial assignment, while in the lower two images, a continuous connection with the subclass none is used. Context in all four images is neither used for the text labels nor their anchors, and thus the classification is both without context.

5.3 First Study: Visual Connection[3]

From the literature on human factors that was collected for this taxonomy, a wide range of information on how visual information linking must be designed can be derived. This includes, for example, that a graphical context is useful to show a connection from a virtual information object to its anchor. However, the taxonomy also shows that there are still gaps in knowledge, and not all of its dimensions have been fully explored yet. One open research question is how all the possible ways to design a visual connection affect a user's sensory and cognitive effort, and by this, the task performance. Overall, this topic has not received much attention in the scientific literature for augmented reality. In studies for IRVEs, it was found that a close proximity of information object and anchor was most advantageous, while others found that fixed positions and connecting lines might even be better. Concerning different ways to establish a symbolic connection and their influence on user performance in comparison to other options, not much is known. Since the existing results are neither particularly conclusive nor focused on augmented reality, further investigations are necessary for this dimension.

Thus, as an approach to create more insights on possible measures for the information linking challenge, the dimension visual connection is examined more closely in an empirical study with test persons. The reason for selecting this dimension and not one of the other two is that it has received particularly little attention so far. Concerning the context, it is already clear from the available literature that including it has a positive influence on a user's understanding of visual information linking [244, 245]. However, the extent to which it should be used is still an open question including, for example, what structured need to be integrated into it. The reference system in which the virtual objects needs to be positioned seems to be strongly dependent on the available display area as well as on its type [227]. Therefore, there are already some clues as to which positioning should be chosen under which circumstances, even if they were developed for IRVEs. Similar conclusions cannot yet be drawn for the dimension visual connection.

5.3.1 Operationalization and Measurements

Before a test setup for this study can be designed, it needs to be determined what shall be measured. The goal when solving the representation challenges is to reduce sensory and cognitive effort. Therefore the different ways of visual information linking must be evaluated according to how well they contribute to this goal. However, these are not directly measurable values and they do not have a fixed scale such as time or weight. Therefore, an operationalization is necessary. Until now, there has been no consensus on how a measurement and thus

[3] The results of this study were previously published in the publication *A Taxonomy for Information Linking in Augmented Reality* [200]. A more detailed description of the used test system can be found in Ralf Dauenhauer's master thesis [72].

how an operationalization should be made [167]. Nevertheless, there is extensive literature on cognitive effort, also called cognitive load, work load or mental load, and multiple measurement methods have been identified that can be used here. Therefore, even if the most suitable method has not yet been determined, the existing ones can provide good indications of how well different types of visual connections compare to each other.

As one approach, it is possible to measure task performance, which can either be primary or secondary [59, 77]. Possible metrics for this are, for example, time or error rate. Primary task performance is the performance that a test person achieves when performing the actual task of a test. Secondary task performance is the performance that is achieved in a task which a test person has to perform besides the primary task. The difference is that in the first case it is measured how well a task is performed, while in the second case it is measured how much mental capacity is not demanded by the primary task. If it is reasonably possible, the primary task performance should be measured, as it provides more reliable results [59]. Task performance measures are limited in that they only give an indirect measure of the cognitive and sensory effort that a person has to put into completing a task. It is, for example, possible that the same task performance is achieved with varying actual load [214]. However, combining them with other measurement approaches helps to ameliorate this effect [215, 77]. These different measures can then be combined to a common value, which, however, places higher requirements on the operationalization [215].

Another approach is the use of subjective measures [59], where test persons are asked to self-asses on one or more scales. The basic idea behind this kind of measurement is that test persons are able to perceive their cognitive load during the completion of a task through introspection, and can report it correctly. There are various assessment tools for this, such as the NASA Task Load Index (NASA TXL) [134], with the scales mental demand, physical demand, temporal demand, performance, effort and frustration. In this test, the test persons themselves can also determine the weighting of the individual scales. The Subjective Workload Assessment Technique (SWAT) [236] with the scales time load, mental effort load and stress load is simpler compared to the NASA TXL, and also widespread. Again, it includes the possibility for test persons to weight the scales. An easy-to-use assessment tool is the Rating Scale Mental Effort (RSME) [316]. It only has one scale going from 0 to 150 and includes anchor points with textual descriptions of effort which support test persons to use it. These are only a few examples for such rating scales, and many others have been developed so far [59]. However, they are not relevant for the further course of the work.

Other possible measures are physiological measures like heart rate, pupil dilation or eye movements [114, 56]. However, for the here presented experiment, these were not available and thus not further discussed. Eye tracking was first considered to get a better understanding of the eye movements, but then dismissed. The reason is that it is very challenging to achieve correct results when it is used in combination with head-mounted displays [145, 148].

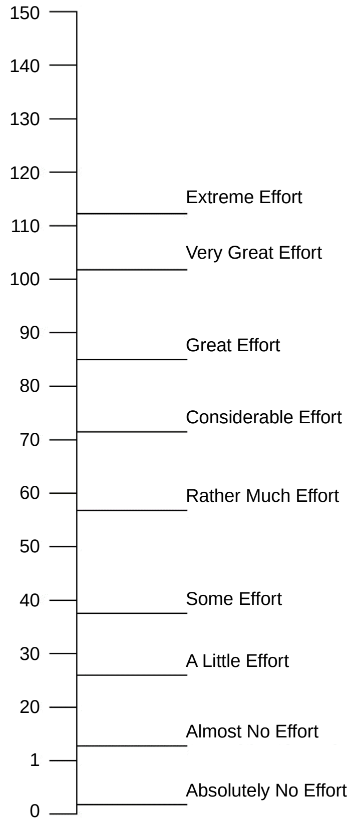


Figure 5.10: The RSME Scale according to Zijlstra [316] (source: own work with translation by de Ward [77])

5.3.2 Experimental Setup

The goal of this experiment was to determine how different kinds of visual connection for information linking perform compared to each other. For this purpose, a simple assignment task was designed. Silhouettes of the five animals, fish, rabbit, bird, crab and pig, were printed out and mounted on a wall. The participants were then placed at a distance of about two meters away from them. For each of these animals, a random number from one to five was shown as information object via augmented reality. Different ways of showing visual connections to the animal silhouettes, e.g., spatial assignment, were then used to display the connections. The participants had to name these pairs (e.g., crab three, bird four, fish one...) as quickly as possible. The animal shapes were chosen because they have no fixed order like letters have. This was done to avoid the so called Stroop effect, which describes mental interference resulting from contradictory information, e.g., when assigning the number 2 to the letter a and the number 1 to the letter b [189]. To not limit the results to only one type of device, the test setup was designed to work with an optical see-through device, a video see-through device and a handheld device with video display.

As already mentioned, the type of visual connection was chosen as independent

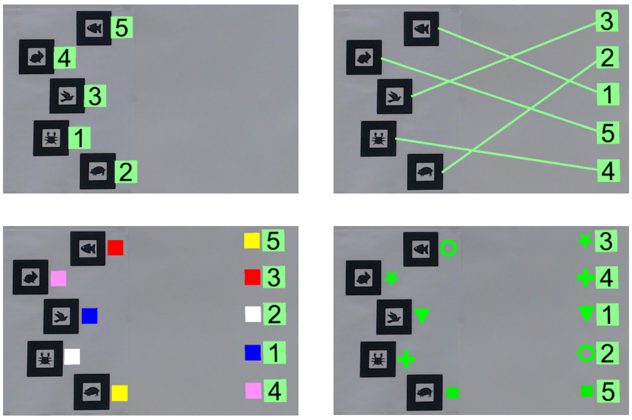


Figure 5.11: The four conditions used for the first study (top left: PROX; top righ: CONT; bottom left: COLO; bottom right: SYMB)

variables. The following four conditions were used for it in the experiment:

- PROX: Spatially assigned (via proximity)
- CONT: Continuously connected (subclass none)
- COLO: Symbolically connected (subclass color)
- SYMB: Symbolically connected (subclass shape)

In every condition, the numbers were included as spatially referenced virtual information objects. The shapes of the animals were used as their anchors. Their reference systems were WCS positioned and WCS oriented. However, because the setup did not allow much movement, the orientation was hardly distinguishable from SCS oriented. A context was not used for either the anchors or for the information objects. A graphical depiction of the four conditions can be found in figure 5.11.

For the proximity condition, which is an instance of the class spatial assignment, the information objects were placed directly next to their anchors. Their positions were carefully controlled so that they did not overlap to ensure that no visibility problems occurred. Another option would have been a direct overlay with semi-transparent information objects. However, this was not done because it might have caused problems with clarity and decreased text legibility. For all other conditions, the information objects were either placed at the edge of the field of view respectively at the edge of the display. However, this was only an approximation, since the WCS positioning of the objects did not allow a direct connection to the view frustum or screen. In this way, it was ensured that the spatial distances for the three conditions CONT, COLO and SYMB were identical. For the CONT variant, continuous, uninterrupted lines in green were used for all connections, i.e., the subclass of visual connection was none. This was

Figure 5.12: The Surface Pro 3 and the modified Lumus DK-32 as used in the test (source: own work)

done to ensure that the test persons could not use any color or drawing style to infer information. In addition, the lines were allowed to cross each other, because otherwise test persons could have inferred the connections based on the order of the information objects. In the symbolic connection variant (SYMB), five different, similarly large, simple symbols (triangle, star, cross, ring and square) were selected, which were all displayed with the same color. Each of the symbols was displayed two times, one time next to the information object, the label with the number, and one time next to the anchor, one of the animal symbols. For the color condition (COLO), colored squares were displayed next to the information objects and the anchors instead of the symbols. They all had the same size as the squares from the symbolic condition. The colors were selected such that they have an appropriate distance in the color space and were therefore easily distinguishable (red, yellow, blue, white and pink).

The device type was not included as an independent variable because the goal of this experiment was not to determine which device performs better, but instead to determine if the effects found were the same for all device types. As a handheld device, a Microsoft Surface Pro 3 was used. It has a resolution of 2160x1440 Pixels on a 12” display [12]. In a pretest, the build in camera of the Surface Pro 3 performed well, so no external camera was used. As an optical see-through head-mounted display, the Lumus DK-32 [8] was used. It has a field of view of about 40°and a resolution of 1280x720 pixel per eye. Because it needs to be attached to an external computing device, the Surface Pro 3 was used for this. It also does not have a built-in camera, therefore a USB 2 uEye XS industrial camera [13] was mounted to its frame. For a head-mounted video see-through setup, the Lumus DK-32 was used as well, but this time the glasses were covered and the video image from the camera including virtual graphics were shown on its displays. Pictures of the devices can be found in figure 5.12.

The software used was programmed specifically for this experiment. For the tracking functionality, ARToolKit [2] was used. It fulfilled the requirements concerning tracking accuracy, was freely available, had good documentation and an active development community.

The experiment was designed as a within-subject experiment where each subject experienced every condition with each of the device types. This was done to avoid an influence on the results caused by differences between individuals in terms of task solving performance. At the beginning of each run, participants were asked to complete a questionnaire asking for demographic information such as age, gender, possible visual impairment and previous experience with augmented or virtual reality systems. After completing the questionnaire, the participants were given an explanation of the test procedure. They were told that it is important to name all animals and their numbers as quickly as possible, but to make sure they are correct. Before the actual tests began, a test run was carried out on each of the used devices. During this run, all four conditions were displayed and it was checked whether the test person could clearly see and understand the information and was able to carry out the tasks.

One assignment task in this experiment consisted of five animals to which five numbers had to be assigned. Each participant had to make 10 random assignments per device type and per condition, here referred to as a block. This means that each individual participant had to complete a total of 120 assignment tasks in 12 blocks. At the end of each block, a participant was asked to estimate the average mental load. The previously described RSME was used for this purpose, as it can be completed very quickly without the need for a longer interruption between two blocks, as would be the case with other evaluation methods, such as the NASA TXL. In order not to generate a systematic error resulting from the order of the conditions or devices, the individual blocks were completed in random order.

As one dependent variable, the time in milliseconds needed to solve an assignment task was selected. Another dependent variable was perceived mental workload (RSME value) for each block. It was known from preliminary experiments that the test persons made mistakes in the assignment, but mostly recognized these themselves and corrected them immediately. It was rather rare, however, that an error was made unnoticed. Instead, each error had an effect on the time needed to complete an assignment. Therefore, the errors made were recorded, but not treated as dependent variables.

To control the system, a wireless presenter was used. An assignment task started when a button was pressed and marked as completed when the same button was pressed again. The time that passed between the first and second pressing of the button was then written to a log file. When the button was then pressed another time, the next assignment task started and the process repeated. Pressing the button was the responsibility of the investigator.

The investigator was also responsible to note errors during each assignment task. Five different types of errors were recorded, one of which led to the exclusion of an assignment task from the evaluation. Details can be found in table 5.1.

5.3.3 Hypotheses

The hypotheses regarding the time to complete the assignment tasks are the following:

Error Code	Error Description	Exclusion
E1	Test person was distracted, problem with the system or error in recording	Yes
E2	Anchors got mixed up	No
E3	Anchors got mixed up, but corrected by test person	No
E4	Information object got mixed up	No
E5	Information object got mixed up, but corrected by test person	No

Table 5.1: Codes for possible errors during an assignment task.

- H1: The use of spatial assignment with proximity (PROX) results in a lower mean time to assign the numbers to animal silhouettes than all other variations across all devices.
- H2: The use of continuously connected visual connections (CONT) results in a lower mean time to assign the numbers to animal silhouettes than the use of symbolically connected (subclasses shape and color) visual connections (COLO, SYMB) across all devices.
- H3: The use of symbolically connected visual connections with subclass color (COLO) results in a lower mean time to assign the numbers to animal silhouettes than symbolically connected visual connections with subclass shape (SYMB) across all devices.

The hypotheses regarding the subjective mental workload are the following:

- H4: The use of spatial assignment with proximity (PROX) results in a lower subjective mental workload than all other variations across all devices.
- H5: The use of continuously connected visual connections (CONT) results in a lower subjective mental workload than symbolically connected (subclasses shape and color) visual connections (COLO, SYMB) across all devices.
- H6: The use of symbolically connected visual connections with subclass color (COLO) results in a lower subjective mental workload than symbolically connected visual connections with subclass shape (SYMB) across all devices.

These hypotheses were formulated on the basis of literature collected for the sections on human factors in the taxonomy. Therefore, only brief justifications for the individual formulations are given here. Hypothesis H1 and H4 were formulated because when anchor and information object are placed directly next to each other, it is expected that no eye movement is necessary to see a connection, keeping completion time and mental workload low. H2 and H5 were formulated because uniform connectedness is a strong Gestalt law and a continuous line can be followed with saccadic eye movements via a scanpath. For symbolic connections, instead a visual search is necessary. Finally, H3 and H6 were formulated because color seems to be a more salient feature compared to symbols in visual search.

The entire family of hypotheses was selected in such a way that if all hypotheses are confirmed, an order of presentation variants with respect to fast recognition and low cognitive effort is the result. The spatial assignment would then be preferable to a continuous connection, which would be preferable to a symbolic connection. Within the symbolic connections, the color coding would be preferable to coding with shapes. By considering two dependent variables as well as three different classes of devices a general validity, although limited to similar scenarios, is also ensured. Thus, this selection of hypotheses results in an evaluation of possible measures to solve the information linking challenge.

5.3.4 Results

Overall, 26 test persons participated in the experiment, 7 of which were female and 19 male. Most of them were students who were working as interns or on their bachelor's or master's thesis at Bosch at the time of the study. However, this was not systematically recorded. The average age of the participants was 26.4 years (SD = 6.8), with the youngest being 19 and the oldest 47 years old. 12 (46%) subjects had a visual impairment corrected by either glasses or contact lenses. One participant stated to be color-blind. 23 (88%) of the subjects reported that they were technology affine. 19 (73%) of the subjects had no or very little experience with augmented reality, and 13 (50%) of the subjects had no or very little experience with virtual reality. 7 (27%) of the subjects played video games regularly.

The interactions between completion time and RSME value were not of interest in this study. Therefore, one-way analysis of variance (ANOVA) with repeated measurements for assignment times and RSME values were performed to test for significant difference in the results. A more detailed comparison of the values was then carried out using appropriate post-hoc tests. The evaluations were carried out separately for each device type (handheld, video see-through, optical see-through).

Errors. A total of 3120 individual measurements were made, with one measurement per completed assignment task. In total, there were 136 errors (E1-E5). Of these, only 22 were due to technical faults or an external distraction (E1). The corresponding measurements were therefore marked as invalid and not taken into account for further evaluations. The main cause of these errors were problems using a head-mounted display. In another 69 cases, the subjects corrected themselves (E3, E5) when they noticed that they had stated an incorrect number or animal. This increased the time need for the assignment task, therefore this kind of error also had an effect on the overall result. Only in 45 cases wrong assignments went unnoticed (E2, E4) which results in a low error rate of 1.45%. Since it was already known from preliminary tests that a low rate was to be expected, the number of errors was not used as a dependent variable. Therefore, they are

	HAR		VST		OST	
	M	SD	M	SD	M	SD
PROX	5269.25	944.20	5550.82	814.65	5522.93	715.12
CONT	6150.19	1090.93	6502.66	1178.92	6851.51	1294.83
COLO	7054.94	1338.17	7387.00	1217.51	8119.65	1295.22
SYMB	8693.66	1507.36	9162.75	1271.35	9273.82	1079.98

Table 5.2: Mean values and standard deviations for the completion times on all four device classes (in milliseconds)

not further considered in this evaluation.

Completion Times. The average time of each block was calculated by summing up the recorded values for the combination of display condition and device for each user and then dividing it by the number of recorded values. Usually these were ten values, but if there were invalid measurements, such as i.e., faults with an E1 code, they were omitted. A one-way ANOVA with repeated measurements showed significant differences for the task completion times between the four variants on all device types. The prerequisites for an evaluation with ANOVA were ensured by a Shapiro-Wilk test including a visual check of the Q-Q diagrams when the test was significant and a Mauchly's sphericity test with subsequent Greenhouse-Geisser correction. The found F- and p-values of the conducted ANOVA tests are the following:

- handheld: $p < 0.001$ (F value including Greenhouse-Geisser correction: $F(1.91, 47.82) = 175.79$)
- video see-through: $p < 0.001$ (F value: $F(3, 75) = 201.15$)
- optical see-through: $p < 0.001$ (F value: $F(3, 75) = 113.75$)

The raw values can be found in table 5.2, the graphic representation can be found in figure 5.13.

In post-hoc tests, the individual results concerning completion times for the conditions were compared using a t-test including a Bonferroni correction with adjusted p-values. The spatial assignment condition (PROX) performed significantly better and resulted in shorter completion times than all other display variants (CONT, COLO, SYMB) on all three device types. The highest p values for each device type were $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} < 0.001$. Therefore hypothesis H1 can be accepted. The continuously connected visualizations (CONT) condition resulted in a significantly lower mean task completion time on all devices compared the symbolical connection with color (COLO) or shape (SYMB). The highest p-values on the different device types were $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} < 0.001$. Thus, hypothesis H2 can be accepted. When comparing the conditions symbolically connected with color (COLO) and shape

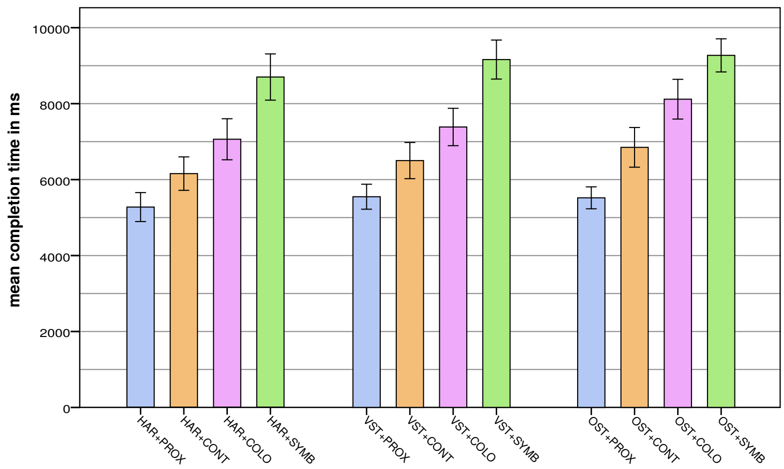


Figure 5.13: Mean completion times needed to assign numbers to animal silhouettes. The bars show the 95% confidence interval.

	HAR		**VST**		**OST**	
	M	SD	M	SD	M	SD
PROX	15.69	8.62	17.92	11.38	20.12	11.17
CONT	31.77	15.54	33.19	18.73	37.12	20.41
COLO	37.38	18.36	40.65	20.74	44.23	19.89
SYMB	58.00	22.34	61.96	26.58	61.73	26.72

Table 5.3: Mean values and standard deviations for the RSME values.

(SYMB), a significantly lower task completion time was found, which was consistent across all tested device types with the hightest p-values being $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} < 0.001$. Thus hypothesis H3 can also be accepted.

Mental Effort. A one-way ANOVA with repeated measurements also showed significant differences for the RSME values recorded for each condition on each device. The prerequisites for an evaluation with ANOVA were confirmed by a Shapiro-Wilk test including a visual check of the Q-Q diagrams when the test was significant. A Mauchly's sphericity test didn't show any violation of sphericity. The found p-values and F-values are the following:

- handheld: $p < 0.001$ (F-value: $F(3, 75) = 62.26$)
- video see-through: $p < 0.001$ (F-value: $F(3, 75) = 46.88$)
- optical see-through: $p < 0.001$ (F-value: $F(3, 75) = 30.47$)

The raw values can be found in table 5.3, and the graphic representation can be found in figure 5.14.

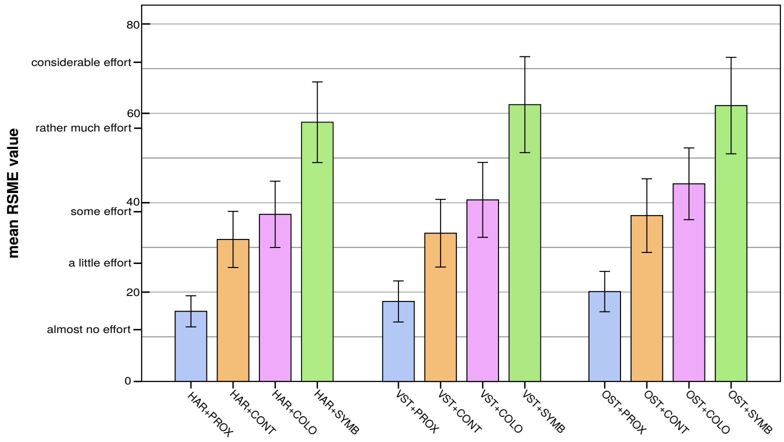


Figure 5.14: Mean RSME values when assigning numbers to animal silhouettes. The bars show the 95% confidence interval.

In post-hoc tests, the individual results concerning RSME values for the conditions were compared using a t-test including a Bonferroni correction with adjusted p-values. The spatial assignment condition (PROX) performed significantly better, i.e., resulting in significantly lower RSME values for all conditions on all devices. The highest p values for each device type were $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} < 0.001$. Therefore, hypothesis H4 can be accepted. Compared to the symbolically connected with shapes condition (SYMB) the continuously connected condition (CONT) showed significantly lower RSME values on all devices with the p-values being $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} < 0.001$. However, compared to the symbolically connected with color condition (COLO), the differences were not significant with the p-values being $p_{har} = 0.374$, $p_{vst} = 0.112$ and $p_{ost} = 0.637$. Thus, hypothesis H5 cannot be accepted and must instead be rejected. The comparison of the symbolically connected with color (COLO) condition compared to the symbolically connected with shapes (SYMB) condition resulted in a significantly lower mean RSME values. These were consistent over all devices with the p-values being $p_{har} < 0.001$, $p_{vst} < 0.001$ and $p_{ost} = 0.002$. Thus, the hypothesis H6 can be accepted.

5.3.5 Discussion

The achieved results show a clear order of the four examined classes in their influence human performance. Since H1 is supported, it follows that the time required to recognize connections is lowest for spatial assignment with direct spatial proximity. This is followed by a continuous visual connection, which results from H2 being supported. In combination, the support of H1 and H2 implicates that symbolic connections are the class with the slowest completion times. However, there is also an order between the subclasses. Since H3 is supported, it follows that color coding leads to faster completion times, while the symbolic connection via shapes leads to slower and the slowest overall completion times. Concerning cognitive effort, the order is the same, but with the restriction that no difference could be found between continuous connections and symbolic connections via color. This is because H4 and H6 were supported, but no significant differences were found for H5.

Perceptual psychology research gives indications that the results obtained will still be valid for a changed number of anchors and information objects and are therefore applicable to similar scenes [230, 288]. However, there are several limitations to this study. One is that no interactions with the other two dimensions reference system and context were examined. Indications that this may influence the results, can be found in literature on search tasks in IRVEs [227, 226]. In addition, the scene used was relatively simple structured and not fully representative for real life scenes. It was missing varying distances between the information objects and the viewer, and no occlusion was included. A further limitation here is that connections from physical to virtual information objects were examined. When connections in the opposite direction have to be recognized, the results may be different. Especially the flexibility in terms of layout, which virtual information objects allow, may be an important factor [227, 226]. Also, with respect to the test subjects, another limitation is that the subjects did not move the view point, which is presumably not the case in a real-world application.

From this study, measures for solving the information linking challenge can be derived, based on the time differences for completing the assignment tasks as well as differences of RSME values reported by the test persons. Since a spatial assignment via proximity caused the least effort, both in terms of completion time and subjective mental effort, it can be considered the preferred measure and should be used where possible. The second in line of preference is a continuous connection. Even though the needed cognitive effort does not significantly differ from a symbolic connection via color, the time required for the assignment provides a good indication that a continuous connection is the preferred from these two. Finally, when a symbolic connection is used in a solution for the information linking challenge, it should be created by using colors. A symbolic connection via shapes is not recommended since the reported cognitive effort was significantly higher and the time required significantly longer than in all other conditions. These measures are nevertheless limited by the above-mentioned restrictions on their general applicability.

Of course, this order of preference only applies if there is freedom of choice. The other challenges such as orientation and visibility must be solved as well. As already mentioned, situations can arise in which solutions must be balanced against each other. A better visibility, for example, may have to be weighed against a spatial assignment of information objects in terms of the lowest sensory and cognitive effort. For IRVEs, the orientation challenge has proven to be a case where solutions have to be counterbalanced against solutions to the information linking challenge [226, 21].

In summary, these results provide an indication of how a connection between information objects and their anchors can be represented in a way that minimizes sensory and cognitive effort. However, as already mentioned, the results are limited by the use of a comparatively simple scene, so that this aspect was investigated in more detail in a follow up study.

5.4 Second Study: Visual Connection with Occlusion[4]

As already mentioned in the discussion for the first study, the used scene was relatively simple and did not include objects at different distances from the viewpoint or any occlusions. These traits, however, are a regular part of real-life scenes. Particularly when using connection lines, i.e., instances of the class continuously connected from the dimension visual connection, occlusion becomes relevant. Either the lines are fully shown and the depth cue occlusion is disregarded, or lines are interrupted at sections where they are located behind other objects. Both variants have the potential to increase sensory and cognitive effort. While incorrect occlusion in the scene can lead to conflicts in depth perception, missing segments of connection lines have to be mentally compensate by a user. In the literature, only limited information exists how this affects the perception of information linking. Therefore, it is particularly relevant to compare how well humans can recognize connections between information objects when continuous connections are correctly occluded compared to a not occluded, but perspectively incorrect, presentation. Further research is needed on how to handle occlusion in visual information linking and what type of visual connection is preferable in scenes with occlusion. Thus, a second study with the goal to examine the influences of different visualizations from the dimension visual connection with and without correct occlusion in a complex scene is presented in this section.

With regard to the information linking challenge, the goal of this study is to refine the measures available to solve it. Many real scenes in the context of procedural tasks contain objects at different distances from the user which can lead to occlusion. In a vehicle, for example, the components are often installed in a complex spatial arrangement that makes it impossible to see all parts at the same distance and without mutual occlusion. This must be taken into account when solving the challenge and measures must be selected accordingly. This study is intended to create a basis for designing a suitable visual information linking in such scenes.

Generally, missing or incorrectly used depth cues make it difficult for a user to interpret an augmented reality scene [168], as was already described with the

[4]The results of this study were previously published in the publication *An Evaluation of Information Connection in Augmented Reality for 3D Scenes with Occlusion* [73]. A more detailed description of the used test system can be found in Ralf Dauenhauer's master thesis [72].

consistency challenge. To compensate for this, additional sensory and cognitive effort is required. However, when occlusion is used as depth cue, virtual information objects may be partially or not at all visible. Concerning information linking, this is particularly relevant for connection lines and similar visualizations that are instances of the class continuously connected. For these, it may happen that some sections of them are not visible. This does not mean that the connection is no longer recognizable or only recognizable with great effort. Humans are able to mentally reconstruct these parts when a sufficiently large proportion is visible [247]. Especially simple shapes can be interpolated based on the visible parts which is also called the Gestalt cue of good continuation. However, additional mental and sensory effort may be needed to compensate for missing parts. Thus, the focus of this study is, if the connections must be completely visualized or if correct depth cues are more important. This includes whether interruptions are acceptable and if other types of visual connection are more beneficial for complex scenes.

For purely virtual environments, Maass et al. [187] conducted a study on how inconsistent occlusion influences the perception of label assignments to their anchors. The subjects were shown static images depicting potentially contradictory depth cues for connection lines and recorded which of these images the subjects felt were appropriate. They found a negative influence of conflicting depth information on the assesment of the presentations by the test persons. However, the significance of the study is limited by the fact that no concrete impact on factors such as task performance was recorded. Furthermore, the effect found was rather small.

5.4.1 Experimental Setup

A special focus for this study was on using a scene with objects at different distances from a user and obstacles that cover parts of connection lines, i.e., instances of the class continuous connection. For this purpose, a setup was developed in which several smaller objects such as a rubber duck or a key were distributed on a table at different distances from the viewer. In an augmented reality view, these were then randomly connected with virtual information objects showing the letters A, B, C, D and E. Several vertical blocks of cardboard were also placed on the desk. One served as the location for the letters, while the others served as occluders, which when using connecting lines, either covered them or lead to invalid depth cues. Overall, seven objects were placed on the table, but for each test run only five randomly selected were used. This was done to avoid any memorization effects by the test persons and to require them to actually look at which letter was connected with which object. For a picture from an ongoing experiment, see figure 5.15.

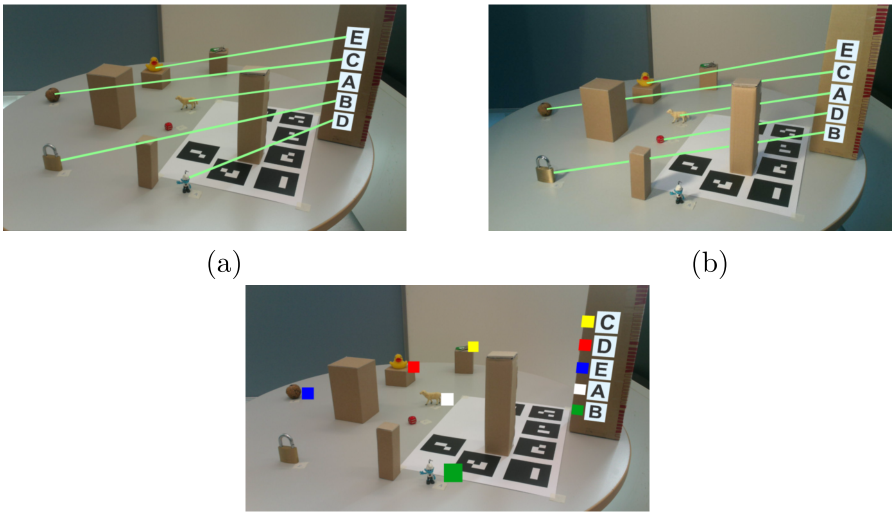
As an independent variable, the type of visual connection was chosen. The following three conditions for it were used in the experiment:

Figure 5.15: A photo from a running experiment, where the condition CONT+DEPTH (connection lines with perspectively correct overlap) is shown (source: photo taken by Ralf Dauenhauer)

- CONT+SIMPLE: direct lines (class: continuously connected) that simply overlay the 3D structures without considering the correct depth information (see figure 5.16 a).
- CONT+DEPTH: direct lines (class: continuously connected) that are spatially correctly integrated into the scene (see figure 5.16 b).
- SYMB+COLOR: symbolic connections (class: symbolically connected with subclass color) via colored squares (see figure 5.16 c).

These three conditions were selected for several reasons. The selection of CONT+SIMPLE and CONT+DEPTH is obvious as they are needed to investigate the effects of occluding parts of continuous connections in complex scenes. SYMB+COLOR was also added because it was the type of symbolic connection that performed best in the previous experiment. It can therefore provide an indication of how symbolic connections perform in comparison to continuously connected ones in complex scenes. A spatial assignment via direct proximity was not included, since this had already performed best in the previous experiment and there was no reason to assume that this would change in complex scenes with a greater distance between anchor and information object. A symbolic connection via shapes was also not included, since it performed significantly worse than symbolic connections via color in the previous experiment.

From the previous experiment, it was also known that all three devices showed the same order of visual connection concerning mean task completion time and mean cognitive effort. Thus, to make this experiment less time-consuming, only one type of device was used. Because occlusion is a monocular depth cue, no stereo vision is required, and since the Lumus DK-32 glasses showed problems concerning



(a) (b) (c)

Figure 5.16: (a) Condition with incorrect occlusion (CONT+SIMPLE), (b) condition with correct occlusion (CONT+DEPTH) and (c) condition with connections via colored squares (SYMB+COLOR) (source: photos taken by Ralf Dauenhauer)

precise calibration and color fidelity, the Surface Pro 3 tablet was chosen as the test device.

At the beginning of each run, participants were asked to complete a questionnaire asking for demographic information such as age, gender, possible visual impairment and previous experience with augmented or virtual reality systems. After completing the questionnaire, the participants were given an explanation of the test procedure.

Similar to the previous experiment, this one was also designed as a within-subject experiment. Each of the test persons experienced all three conditions in a random order. Ten different combinations of objects and characters were displayed for each of the three conditions. For these, the test persons had to recognize the assignments from the objects to the characters (e.g. lamb A, dice B ...) as fast as possible and speak them out loud. After that, the experimenter switched to the next assignment task and the time between start and completion was automatically recorded. When the ten combinations for each condition were completed, each test person was asked for the RSME value. The RSME scale was again used here due to the good experiences in the first experiment. Then the next condition was started. Overall, each test person had to recognize 30 assignments.

Again, the completion time, measured in milliseconds, and the mental effort, measured by the RSME test, were selected as dependent variables. Errors were recorded, but neither treated as dependent variables nor evaluated, since it was clear from a pre-test that they rarely went unnoticed, similar to the previous study. For error recording, the same codes were used as in the last experiment.

5.4.2 Hypotheses

The following four hypotheses were formulated for the experimental evaluation:

- H1: The condition with the direct lines and the incorrectly used depth cues, i.e., the lines pass in front of the occluders, (CONT+SIMPLE) will lead to a lower mental effort compared to the condition where the lines are integrated with a spatially correct occlusion (CONT+DEPTH).
- H2: The condition with the direct lines and the incorrectly used depth cues (CONT+SIMPLE) will lead to a lower mean completion time compared to the condition where the lines are integrated with a spatially correct occlusion (CONT+DEPTH).
- H3: Continuous connections are robust in case of false depth cues (CONT+SIMPLE) or occluded sections (CONT+DEPTH) and still result in a lower mental effort than the connections via color coding (SYMB+COLOR).
- H4: Continuous connections are robust in case of false depth cues (CONT+SIMPLE) or occluded sections (CONT+DEPTH) and still result in a lower average task completion time than the connections via color coding (SYMB+COLOR).

Since Maass et al. [187] found only small negative impacts regarding incorrect depth cues, it can also be assumed that no particularly strong negative effect will occur in the context of information linking in augmented reality. Therefore, it seems reasonable to assume that larger gaps in the connection lines due to occlusion cause more mental effort and longer completion times than inconsistent depth cues do. This expectation is reflected by H1 for mental effort and by H2 for completion time.

The first experiment showed that there was a significant advantage of continuous compared to symbolic visual connections via colors in terms of time required to recognize the assignments. With an increased complexity of the scene used and thus a more complex visual search, it can be assumed that this also holds true for more complex scenes. In addition, people are able to easily reconstruct gaps in simple shapes like connection lines as described by the Gestalt law of good continuation [247]. This is reflected by the hypothesis H3 for mental effort and H4 for the time needed.

Regarding the information linking challenge, these four hypotheses are structured in such a way that, if confirmed, they provide a further refinement of the measures to be taken for solving the challenge. With H1 and H2, it is determined whether it is important to display connection lines without interruptions, even if this means giving inconsistent depth cues. With H3 and H4, it becomes clear whether direct connections are still the superior measure in complex scenes compared to symbolic connections.

5.4.3 Results

A total of 24 volunteers took part in the experiment as test persons. Most of them were students who were working as interns or on their bachelor's or master's thesis at Bosch at the time of the study. Their average age was 27.42 years (SD = 8.46). The youngest participant was 21 while the oldest was 55 years old. 10 (42%) of the participants were female, the rest were male. All participants reported having only limited previous experience with augmented reality or virtual reality but were generally technology affine.

Errors. Overall, 720 measurements were made. From these, 7 were marked with an E1 code and thus invalid. In 19 cases, subjects made a wrong assignment, which was then self-corrected (E3, E5). As in the previous experiment, the assignment time is prolonged by this so that these cases are included in the evaluation due to their higher time requirement. Finally, 13 cases remained unnoticed (E2, E4). However, these are treated as inadvertent mistakes and are not excluded from the evaluation like E1 errors. In particular, the low error rate of 1.82% resulting from E2 and E4 errors confirms the decision not to use the errors as a dependent variable.

Completion Times. For evaluation of the experiment, the mean time required to solve the individual assignments was determined per test person and condition. Invalid measurements, i.e., faults with an E1 code, were omitted. With a Shapiro-Wilk test, it was checked whether the results for completion times were normal distributed. This was not the case for the condition with a continuous connection without correct depth cues (CONT+SIMPLE: $p = 0.038$) and the condition with a symbolic connection via colors (SYMB+COLOR: $p = 0.015$). This might be some artifact from the way the experiment is designed, for example, e.g., that certain combinations of objects and characters are more difficult to recognize or take more time to speak. However, the cause was never determined.

Due to a lack of normal distribution, the completion times were evaluated with a non-parametric Friedman test. It shows significant ($p < 0.001$) differences between at least two of the conditions, which can be found in figure 5.17 for the recorded time.

In order to decide which of the hypotheses can be accepted and which must be rejected, post-hoc tests were performed. However, since several tests were performed on the same data, the alpha level was adjusted. To evaluate H2 and H4, three comparisons on the completion times were needed. Thus the alpha level was adjusted via a Bonferroni correction to $\alpha' = \frac{0.05}{3} \approx 0.016$[5].

To evaluate H2, a Wilcoxon-signed-rank test was used to compare the average time needed for the conditions with correct (CONT+DEPTH: $Mdn = 6138.95$) and without correct depth cues (CONT+SIMPLE: $Mdn = 6363.95$). It was used because CONT+SIMPLE is not normally distributed and thus a nonparamet-

[5]This is different from the first study where the individual p-values were adjusted. Both variants correspond to the way it was done in each of the original publications and were not changed for this thesis to stay consistent with them.

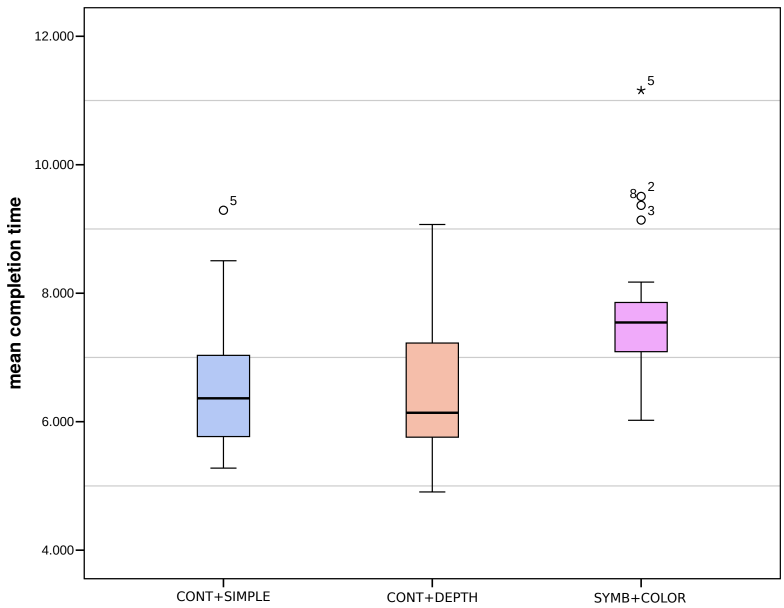


Figure 5.17: Mean completion times needed to assign objects to characters. The boxes denote the two quartiles between 25%-75% of the results, the whiskers denote up to 1.5 IQR.

ric test had to be used. However, no significant difference (CONT+SIMPLE / CONT+DEPTH: $Z = -0.257$, $p = 0.797$) could be found. Thus, H2 must be rejected.

For the evaluation of H4, a Wilcoxon-signed-rank test was used to compare the average time needed for the conditions with continuous connections with correct (CONT+DEPTH: $Mdn = 6138.95$) and without correct depth cues (CONT+SIMPLE: $Mdn = 6363.95$), to the condition with symbolic connections via colors (SYMB+COLOR: $Mdn = 7544.60$). Both comparisons showed significantly lower completion times for the conditions with connection lines (CONT+SIMPLE / SYMB+COLOR: $Z = -4.257$, $p < 0.001$ and CONT+DEPTH / SYMB+COLOR: $Z = -4.200$, $p < 0.001$). Thus H4 can be accepted.

Mental Effort. To confirm that the results for mental effort were normal distributed, a Shapiro-Wilk test was used. Then, the recorded RSME values were evaluated with an ANOVA that showed significant differences ($p < 0.001$) between at least two of the conditions. The results can be found in figure in figure 5.18 for the recorded RSME values

To evaluate H1 and H3, three comparisons on the RSME values were needed. Thus, the same Bonferroni correction as for the completion times was used. The alpha level was again adjusted to $\alpha' = \frac{0.05}{3} \approx 0.016$.

To evaluate H1, a pairwise t-test was used to compare the subject's mental effort in the two conditions with correct (CONT+DEPTH: $M = 29.88$, $SD = 14.41$) and without correct depth cues (CONT+SIMPLE: $M = 31.88$, $SD = 13.42$). No

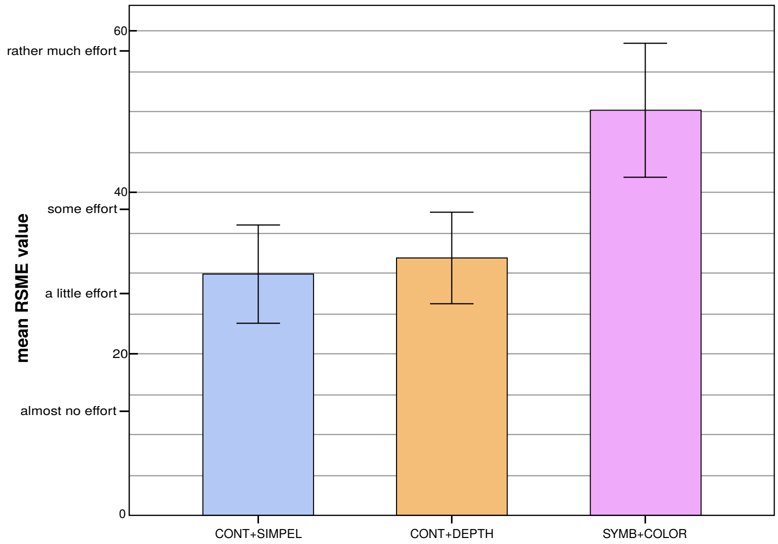


Figure 5.18: Mean mental effort to assign objects to characters. The bars show the 95% confidence interval.

significant difference between the two conditions could be found ($p = 0.294$) and thus the hypothesis H1 must be rejected.

For the evaluation of H3, a pairwise t-test was used to compare the RSME values from the conditions with a continuous connection with correct (CONT+DEPTH) and without correct depth cues (CONT+SIMPLE), to the condition with symbolic connections via colors (CONT+COLOR: $M = 50.17$, $SD = 19.65$). Both comparisons showed significantly lower RSME values for the conditions with connection lines (both $p < 0.001$). Thus. H3 can be accepted.

5.4.4 Discussion

The results from this second study provide further insights on how the information linking challenge can be solved. First, they are an additional indication that connection lines are a suitable way to present information objects as connected when direct spatial assignment is not used. Further, they also give information that could not already be derived from the previous study. This is especially true for complex and more realistic scenes which include objects in different distances from the viewpoint and occluded connection lines.

As in the first study, there are some limitations to consider when interpreting the results. Again, the task was to first recognize a physical object and then identify the assigned a character. A transferability of the results to an assignment in the other direction, however, is not certain. Furthermore, the other two dimensions reference rystem and context were not varied this time either, which further limits the generalizability of the results, especially since in IRVEs relevant interactions were observed [227, 226]. Another limitation is that in this study the results were obtained with only one type of device, namely a handheld device. However, the results from the first study suggest that the type does not play a role concerning the preference of the type of visual information linking.

The rejection of H1 and H2 can be interpreted in at least two ways. One possibility is that the negative effects due to occluded sections and inconsistent depth cues are very small and hardly noticeable. Another possibility is that both effects are significant but similar in size and thus no difference is found. However, the fact that H3 and H4 could be confirmed hints at the first possibility being true because of the clear superiority to symbolic connections that was found. For a final clarification, however, a further study would be necessary in which the presence or absence of occluding elements is used as a further condition. Regarding the information linking challenge, it can be concluded that both variants of continuous connections can be used as measures, and that neither has particular (dis-)advantages. Thus, connection lines can be interrupted to achieve correct embedding into a scene. However, in this case it is important that the geometry is sufficiently simple and enough sections of the line can still be seen so that the Gestalt cue good continuation can take effect [247]. If no sufficient model of the scene is available, simple continuous lines can also be used without negative effects.

The confirmation of the hypotheses H3 and H4 again shows that continuous connections are preferable to symbolic connections for the given combination of task and scene. With this study, it becomes clear that it also stays true in complex scenes, where either connection lines or similar visualizations have to be interrupted to achieve correct occlusion or have to be displayed incorrectly in front of other objects. Reasons for this result may be that no colors need to be memorized and no visual search is necessary. Based on the feature-integration theory of attention by Treisman and Gelade [288], it can be assumed that this result will also holds true for a varying number of anchors and information objects. Regarding the information linking challenge, this means that when either a continuously connected or a symbolically connected visual connection can be used, a continuous connection is preferable, regardless of whether a correct occlusion can be achieved or not.

6 Using IRMAR in Human-Centred Design

In this chapter, it is demonstrated how the the IRMAR framework can be applied as part of human-centred design as defined in ISO 9241-210. This is the first of two such applications shown in this thesis. It contributes to achieving the goal defined in the motivation to demonstrate the application of the IRMAR framework including the representation challenges.

The IRMAR framework is intended to facilitate creating augmented reality systems that support manual procedural tasks by supporting the design of information representation. However, it has not yet been outlined how it can be used for this purpose in practice. To this end, it is not necessarily helpful to only show how the challenges can be solved for a single system. Instead, it is useful to demonstrate an approach for its application in a more general way. To achieve this, a methodical integration into an existing design framework is made in the following.

One often-followed method in designing interactive systems based on users' needs is human-centred design, which is defined in the ISO norm 9241-210 [80]. It describes an iterative approach during which a system is repeatedly evaluated from a user's perspective, optimally with actual users, and improved based on the gathered information. This framework is very adaptable to different contexts, and can also be used to develop the information representation for an augmented reality system. Therefore, an integration of IRMAR into this norm is proposed. In addition, practical examples are given to show how it can be used as part of human-centred design activities.

ISO 9241-210 is chosen because of several reasons. First, it is one generally-accepted standard for designing interactive systems that has a relevant distribution. It can therefore be assumed that providing an integration into it will remove barriers and support the use of IRMAR in an engineering context. Furthermore, it is very flexible in the kind of systems designed with it. Many other existing standards or design guidelines are oriented towards established forms of user interfaces, which are based on the features of PCs, laptops or mobile devices. However, as already argued in the fundamentals chapter, augmented reality does not fit into any established form of user interface. Additionally, ISO 9241-210 is a constructive standard that is focused on creating new systems rather than analyzing or evaluating existing ones. Since the norm deals with the activities necessary for the creation process, choosing it supports specifying activities associated with the use of IRMAR.

However, there are alternatives to using this norm as a target for integration. For example, there are two other norms from the ISO 9241 standard family alone, namely ISO 9241-125 Guidance on visual presentation of information [81] and ISO 9241-110 Interaction principles [82], which might seem suitable for integration. Both focus on aspects of how user interfaces must be designed and are both open for extension. In the following, these are taken as examples to explain why they and similar approaches are not used for integration.

ISO 9241-125 focuses, as the name Guidance on visual presentation of information suggests, on the presentation of information in visual computer interfaces controlled by software. It is structured in the form of a catalog in which a variety of guidelines is given on how to realize individual aspects of information presentation. This includes, for example, the structure of icons or how characters can be used for visual encoding. To add IRMAR as an extension to the norm, the measures for the identified representation challenges could be added to the catalog as additional parts. However, they would not quite fit in, as they are on a different conceptual level. The aspect of information presentation in IRMAR is intentionally not considered and instead the focus is on the embedding of information objects into a scene. Furthermore, ISO 9241-125 is designed for traditional forms of user interfaces. This becomes apparent, for example, with the description of layouts for windows or the design of cursors. Most of the recommendations from ISO 9241-125 are not or only marginally relevant for the design of augmented reality systems supporting procedural tasks. Nevertheless, there is some limited overlap with the measures for representation challenges, e.g., concerning the arrangement of elements or the use of colors. Therefore, an integration of IRMAR would make only this small, shared part accessible for augmented reality systems. The large parts of guidelines available in the standard cannot be transferred and would still be, from the perspective of augmented reality system design, unusable. Moreover, this standard does not describe activities associated with the design of information presentation. Therefore, an integration does not create additional benefit in this respect either. In sum, it is questionable whether an integration creates sufficient additional value since there are simply too many differences. Only little of the standard can be applied to designing an augmented reality system to support procedural tasks. Further, embedding IRMAR in this standard does not provide any guidance for solving the representation challenges beyond the already described measures. For comparable catalogs with guidelines on how user interfaces or aspects of them should be designed, such as those given by Smith et al. [269] that exist for PCs, laptops or mobile platforms, the line of argument applies in a similar way.

In ISO 9241-110, interaction principles are established which are relatively general. An example of such a principle is that user interfaces must be self-descriptive. The representation challenges from IRMAR, however, do not match with these principles because they relate to a variety of aspects which a user interface must provide. Presentation or representation of information such that sensory or cognitive effort is minimized is not one of them. Instead, an appropriate representation of information can support the implementation of each of the principles. However, an integration of IRMAR with this focus would be very extensive or unspecific because it very much depends on the realization of the principles. A self-descriptive user interface could be implemented in many different ways, for example, using tooltips activated by interaction. Accordingly, IRMAR would have to be applied to a wide range of these possibilities. Therefore, if IRMAR were integrated to support realizing these principles, the integration would have to either be very general or be applied to a large number of different ways, resulting in a corresponding complexity. Another way to integrate IRMAR would be to add minimizing sensory and cognitive effort through an adequate information representation as a further principle. However, this would offer little added value compared to how the representation challenges are already described. As in ISO 9142-125, no activities are described in this standard, therefore, an integration does not create additional benefit in this respect either. For guidelines or heuristics that apply a similar structure of principles, such as those provided by Nielsen [207], the line of argument applies in a similar way.

The focus here is on how information objects can be identified and defined and how an appropriate information representation can be designed as part of the activities specified in ISO 9241-210. Therefore, neither a concrete development process nor the general use of the norm for augmented reality systems is described. Further, no change of ISO 9241-210 is proposed. Instead the norm is general enough to allow an inclusion of the activities required to use the IRMAR framework.

The integration of the IRMAR framework and its application is demonstrated using an exemplary repair of a vehicle in which an airbag must be removed for later replacement. Multiple fictional examples throughout the chapter show the first iteration of a human-centred design process for a fictional augmented reality system that supports this repair task. Its goal is to make it easier for technicians to carry out this task when they are not sufficiently familiar with the vehicle. In this way, practical examples are given how the IRMAR framework can be used and how it helps to develop an adequate information presentation. To remain concise, only those parts are included in the examples that have a direct relation to the described way of including IRMAR into ISO 9241-210. Other aspects of human centred-design are left out. The examples are simplified in that they do not attempt to define all information objects and constraints for solving the representation challenges in each activity, but rather a number sufficient to explain the presented ideas. The examples are written from the perspective of a functional development team with the goal to develop this system with human-centered design. The descriptions are based on the author's experience during his work at Bosch, but heavily condensed for the use as an example.

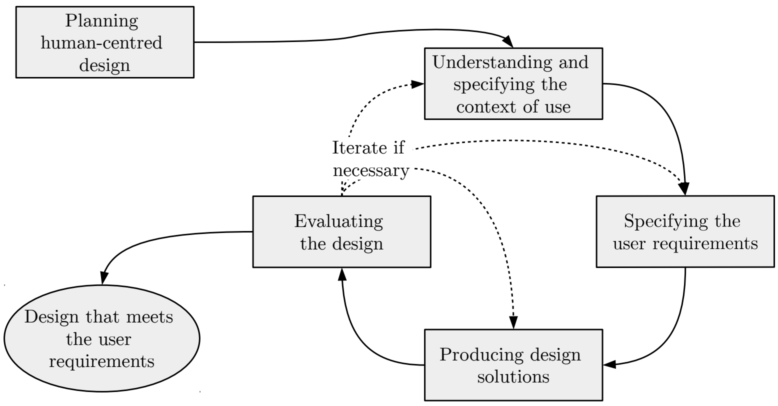


Figure 6.1: Human-centred design activities and and their temporal sequence (source: own work based on ISO 9241-210 [80])

6.1 ISO 9241-210

ISO 9241-210 is a norm that defines human-centred design as well as the activities associated with it. In its introduction, its authors note that using it “enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance” (ISO 9241-210 [80], page vi). However, it does not describe an actual development process but instead defines which activities in which order must be part of a product life cycle to support human-centred design. In the following a brief overview of the norm is provided.

A total of five distinct activities are defined, which have a fixed order as depicted in figure 6.1. The sequence of activities is iterative in order to produce increasingly better results based on the information gained from practical evaluations. These are either conducted with users or at least consider a user’s perspective. The iterative approach is needed because in many cases, it is not yet clear at the beginning of a development what the exact requirements are. Instead, these only arise through interaction of users with prototypical or real systems. Therefore, it is necessary to evaluate intermediate results together with users or at least from a user’s perspective and to plan further steps on this basis.

The following activities are described in the norm:

- “Planning human-centred design” (ISO 9241-210 [80], section 5): In this activity, it shall be planned how human centred-design can be integrated into a product life cycle. This includes estimating the importance of human factors for the product, planning how the activities for human-centred design can be successfully integrated in an organization and assigning time and resources to performing these activities.

- “Understanding and specifying the context of use” (ISO 9241-210 [80], section 6.2): In this activity, a description of the context in which the system will be used shall be created. The context consists of a description who users and stakeholders are, what their specific characteristics, goals and tasks are and what the environment for the system is. The goal is to later be able to use this description to derive requirements and an information design.
- “Specifying the user requirements” (ISO 9241-210 [80], section 6.3): In this activity, it shall be collected what the requirements from the users are. These may be acquired directly from users but also other sources like organizational requirements are possible. Because requirements are in many cases not consistent, conflicting requirements must be balanced.
- “Producing design solutions” (ISO 9241-210 [80], section 6.4): Based on the context of use and the user requirements, possible solutions for the system design shall be produced in this activity, including designing the interaction of a user with the system and the user interface. In early stages of development, these do not need to be actual implementations but can as well be non-functional prototypes. This activity is not limited to creating one solution per iteration. Instead, multiple solutions can be created that are compared in the following activity to determine the most adequate design.
- “Evaluating the design” (ISO 9241-210 [80], section 6.5): In this step, an evaluation of the previously-created design solutions shall be made. The goal is to obtain feedback from a user’s perspective and decide on the next activity based on this feedback. However, because obtaining feedback from actual users, e.g., with empirical studies, can be very expensive, other approaches such as expert interviews are also possible. Based on the results, either a new iteration can be started or alternatively the development can be completed.

6.2 Example Use Case

The use case used to demonstrate the inclusion of IRMAR in ISO 9241-210 is briefly presented in the following. It is a procedural task to remove an airbag installed in the driver’s door of a vehicle. The steps shown here are part of a larger repair process, which also includes the installation of a new airbag. However, in order not to make the use case too complex, only this limited segment is used. Figures 6.2 - 6.17 show its individual work steps.

Figure 6.2: Step 1: the cover behind the lock lever must be removed with a screw driver. It is important for the screw driver to be used in the shown position. (photography by Clemens Günther)

Figure 6.3: Step 2: the SRS clip must be removed with a screw driver. (photography by Clemens Günther)

Figure 6.4: Step 3: the cover clip at the bottom side of the arm rest must be removed with a screw driver. (photography by Clemens Günther)

Figure 6.5: Step 4: the screw holding the lock cover must be unscrewed and then the lock cover must be removed. (photography by Clemens Günther)

Figure 6.6: Step 5: the screw behind the lock lever must be unscrewed. (photography by Clemens Günther)

Figure 6.7: Step 6: the screw below the arm rest must be unscrewed and removed. (photography by Clemens Günther)

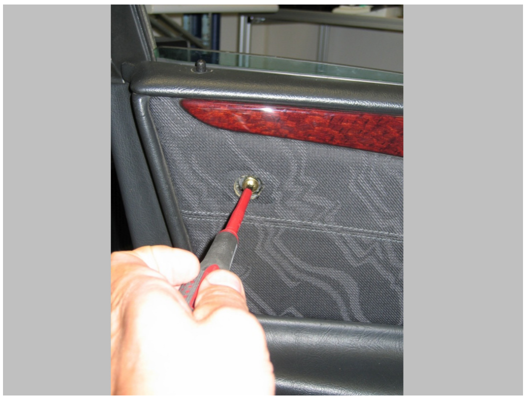


Figure 6.8: Step 7: the screw below the SRS clip must be unscrewed and removed. (photography by Clemens Günther)

Figure 6.9: Step 8: the door cover must be pulled on the right and left bottom. Both hands must be used and it may not be pulled more than 5cm to not damage the upper side of the cover. (photography by Clemens Günther)

Figure 6.10: Step 9: the door cover must be fixed at the bottom side such that it is positioned with a little distance to the door body. Then it must be pulled upwards to slide out of the the fasteners. It is important to be careful to not damage the upper part of the cover. (photography by Clemens Günther)

Figure 6.11: Step 10: before the door cover can be fully removed the bowden cable must be detached from the lock lever. (photography by Clemens Günther)

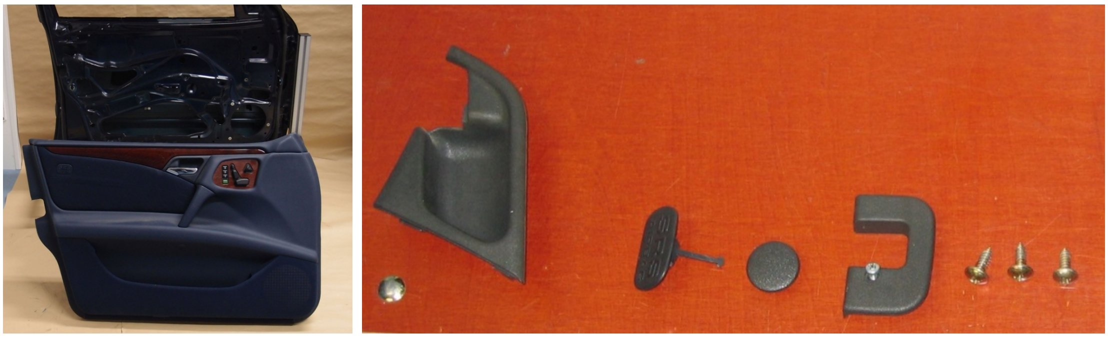


Figure 6.12: Step 11: the door cover can be removed. The parts must be stored for reassembly later on. (photography by Clemens Günther)

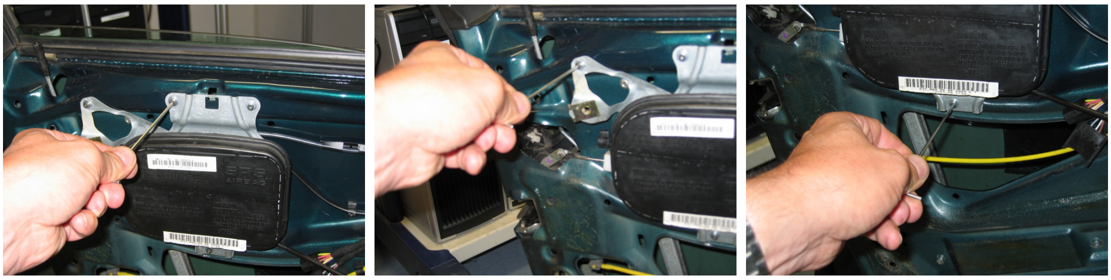


Figure 6.13: Step 12: the bolts from the airbag must be removed. (photography by Clemens Günther)

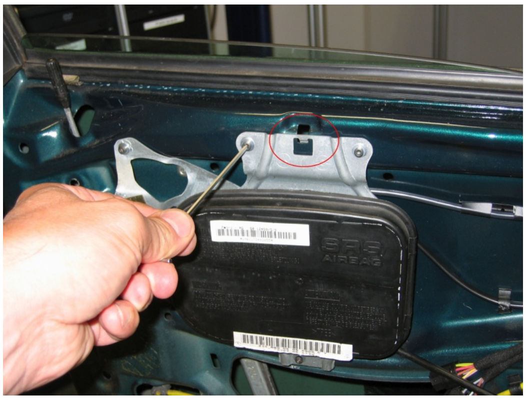


Figure 6.14: Step 13: the airbag must be pulled upwards to be unhooked from the door body (see red circle). (photography by Clemens Günther)

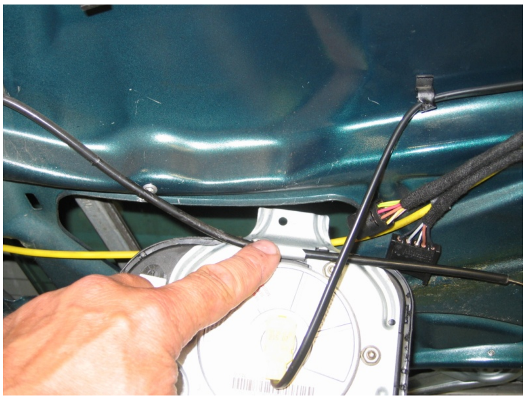


Figure 6.15: Step 14: the bowden cable must be removed from the airbag body. (photography by Clemens Günther)

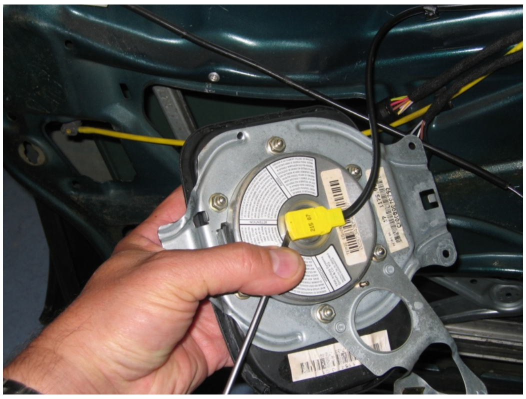


Figure 6.16: Step 15: the connector at the back of the airbag cable must be removed. (photography by Clemens Günther)

Figure 6.17: Step 16: the airbag must be placed with the soft side up at a safe place for later reassembly. (photography by Clemens Günther)

6.3 Integration into ISO 9241-210

For this approach to facilitate using the IRMAR framework as part of ISO 9241-210, it must be integrated into four of the five activities defined by the norm. Only planning human-centred design is left out because it is mostly concerned with organizational aspects that are not relevant for identifying information objects, nor do they provide information on how the challenges can be solved. In the following, for each of the remaining activities it is explained what is necessary to integrate IRMAR into them and what the outcome of each activity is with respect to IRMAR. To further demonstrate the integration, examples are given based on the previously described use case.

The general approach is to derive two types of information from the findings collected in each of the three activities Understanding and specifying the context of use, Specifying the user requirements and Producing design solutions. The first type is (prototypical) information objects. When it is determined that a certain piece of information must be displayed to a user, it is recorded as an information object. However, they do not need to be fully described when they are first discovered. Instead they can be refined at a later stage which may also include splitting them up or merging multiple into one. At the latest in the activity Producing design solutions when designing concrete solutions, they must be worked out in sufficient detail. An example of this would be the requirement of a user to obtain a description of a physical object. When it is first discovered, it is collected as an information object, then in a later activity refined as a text and finally for the design solution made to a label. The other type are constraints on how the representation challenges can be solved. These can be found wherever there are restrictions on how information objects can be integrated or represented in a scene. An example for such a constraint could be that due to an organizational regulation, the color red may only be used for warnings, but not for other purposes.

As part of the activity Producing design solutions information objects must not only be collected but also be embedded in a scene to create a design solution. Thus, they must be first visualized, which is not the focus of IRMAR, and then integrated in a scene while addressing the six representation challenges. The measures presented for the challenges can help with this. However, the constraints limit how these measures can be applied. When, for example, red was the best suited color for a text, e.g., due to a good contrast to the environment, the constraint from the previous example would not allow using this color. Finally, in the activity Evaluating the design, it is evaluated if the correct information objects were identified and the challenges were solved sufficiently. This can be done by using methods such as user tests to compare different information representations.

Not part of this integration of IRMAR is an approach of how the requirements that lead to the identification of information objects or constraints are collected. This is part of requirements engineering, which, for example, has been covered by Kotonya and Sommerville [166], Sommerville and Sawyer [270] or Chung et al. [65]. Instead, it is assumed that it is possible to collect these requirements.

This approach to integrate the IRMAR framework into ISO 9241-210 is described in detail in the following sections. In accompanying examples, an augmented reality system to support the previously presented use case is developed. As part of these, the necessary information objects and constraints are identified, and the information representation is designed within the human-centred design activities. As already mentioned, the examples are given from the perspective of a team that develops this system.

6.3.1 Integration into “Understanding and Specifying the Context of Use”

The context of use is defined in ISO 9241-210 as “the characteristics of the users, tasks and organizational, technical and physical environment” (ISO 9241-210 [80]) that are relevant to the application of the system to be developed. The knowledge contained in its description is also a source of information objects and constraints. Therefore, information on them can be gathered as part of creating the context description. In total, the norm defines four different aspects that are part of a context of use. With regard to information objects and constraints, these aspects contribute in different ways. The following list therefore describes what the individual aspects need to be examined for. However, this should not be considered exhaustive. Constraints or information objects may also emerge from aspects which are not explicitly listed as source. It is therefore possible that an information object results from the analysis of the system environment, whereas this is not explicitly specified in the list. Instead, it should to be understood as an indication, on which sources of information objects and constraints one should especially concentrate.

- As part of “the characteristics of the users or groups of users” (ISO 9241-210 [80], section 6.2.2 b), it can be determined if solutions to the challenges are subject to special constraints by the user group. An example are requirements on the used colors and needed contrasts as part of the clarity challenge. This point results from the goal to represent information objects in such a way that they are adapted as well as possible to users. Therefore, limitations resulting from the user group must be accounted for.
- As part of the description of “the goals and tasks of the users” (ISO 9241-210 [80], section 6.2.2 c), information objects can be collected because each procedural task also implies some sort of interaction with objects that must be manipulated, and these objects can be identified at this stage. Additional information objects can be identified that are necessary because they provide relevant information on the needed manipulation. However, the task and the goals can also result in new constraints for solving the challenges. The norm lists here, for example, dangers to health that must be considered.

- As part of "the environment(s) of the system" (ISO 9241-210 [80], section 6.2.2 d), constraints can be identified that arise from the environment[1]. While it does not change the actual task and thus is not a likely source of information objects, it limits how the challenges can be solved. Relevant factors are for example like lighting, work practices or safety requirements.

While information objects are collected in this activity, they do not have to be fully defined yet. For example, their class does not have to be already assigned and no visualization needs to exist. During following activities, it may also happen that information objects are split or merged.

The first aspect of the context of use, "the users and other stakeholder groups" (ISO 9241-210 [80], section 6.2.2 a), is not included in this list, because this part of the context does not have a direct influence on how information must be represented. Instead the actual users are addressed by the aspect "the characteristics of the users or groups of users" (ISO 9241-210 [80], section 6.2.2 b).

In summary, there are two results of this activity concerning the application of IRMAR. The first is a collection of prototypical information objects that may need to be refined and modified, and the second is a collection of constraints that must be considered when solving the challenges.

Exemplary Execution

An augmented reality system shall be created by a development team to support the described example use case. As a starting point, a survey with workshop employees is carried out with the goal of learning more about "the characteristics of the users or groups of users" (ISO 9241-210 [80], section 6.2.2 b). One of the results obtained is that new colleagues often have many questions about airbag removals, concerning aspects like positions of parts, short explanations of work steps, etc. Thus, for these colleagues, such information must be shown to answer these questions in situ. From this result, one constraint can be derived, als shown in table 6.1.

Work Step	Constraints
All	Solutions to the challenges must be robust with regard to the simultaneous display of multiple information objects.

Table 6.1: Collection of constraints after the first one has been identified during the activity Understanding and Specifying the Context of Use.

With regard to the aspect "the goals and tasks of the users" (ISO 9241-210 [80], section 6.2.2 c), successful completion of the described airbag removal is a goal and the way to achieve this is a task. First, information objects can already

[1]While it shares the the same name, the environment from IRMAR is not meant here.

be identified from this task, namely those that are independent of the support provided by augmented reality. This includes all physical objects that must be manipulated during the removal task and must be seen by the user. For safety reasons, it is also important that the hands must be occluded as little as possible to prevent injuries when working on the vehicle door. This means that hands must be treated as information objects as well. Other information objects like instructions, which stem from the support through augmented reality, only become relevant with the following activity. Overall, the information objects shown in table 6.2 can be identified based on the task in this activity.

Work Step	**Information Objects**
Step 1 (figure 6.2)	cover behind lock lever, screw driver, possibly lock lever (when removing the cover, it may collide with the lock lever and thus a user must see it), hands
Step 2 (figure 6.3)	SRS clip, screw driver, hands
Step 3 (figure 6.4)	cover clip, screw driver, hands
Step 4 (figure 6.5)	lock cover, fastening screw, screw driver , hands
Step 5 (figure 6.6)	fastening screw, screw driver, lock lever (must be taken into account to not scratch it with the screw driver), hands
Step 6 (figure 6.7)	fastening screw, screw driver, hands
Step 7 (figure 6.8)	fastening screw, screw driver, hands
Step 8 (figure 6.9)	door cover, lower part of door frame, hands
Step 9 (figure 6.10)	door cover, center part of door frame, hands
...	...

Table 6.2: Collection of information objects identified during the activity Understanding and Specifying the Context of Use. For the sake of clarity, it is cut off after the first nine work steps.

Further constraints come from the “the environment(s) of the system” (ISO 9241-210 [80], section 6.2.2 d). The environment is here a car repair shop. These often have all sorts of things standing around like vehicles, stacks of tires, fully cleared shelves, and so on. In many cases, there is also some sort constant movement caused by people, tools or vehicles. This implies that only an insufficient prediction about the environment, as in the sense of the IRMAR framework, is possible. Further, artificial light is generally used in a workshop. This leads to two new constraints as shown in table 6.3.

Work Step	Constraints
All	Solutions to the challenges must be robust with regard to the simultaneous display of multiple information objects.
All	The environment can hardly be predicted and thus solutions for the challenges must be robust concerning this uncertainty.
All	The augmented reality system will used be under artificial light and use of colors must be adapted adequately.

Table 6.3: Collection of constraints including those identified during the activity Understanding and Specifying the Context of Use. Entries in black indicate constraints that were newly added or modified, while entries in gray indicate those that were carried over from a previous table.

6.3.2 Integration into “Specifying the user requirements”

Acquiring and specifying relevant user requirements is a main part of human-centred design conducted as part of the activity “Specifying the user requirements” (ISO 9241-210 [80], section 6.3). To achieve this, a number of sources such as the context of use, user needs or guidelines must be analyzed. A total of five sources are particularly pointed out by the ISO norm as mandatory to consider (ISO 9241-210 [80], section 6.3.3). However, not only can general requirements be derived from them, but they can also be origins of information objects and constraints on how the challenges can be solved:

- As a first source, the “intended context of use” (ISO 9241-210 [80], section 6.3.3 a) is listed. Because information objects or constraints from it are already derived in the previous activity, this source is not relevant here. Information objects and constraints that result from an interaction of context and user needs are addressed with the next point.
- Part of the “user needs and the context of use” (ISO 9241-210 [80], section 6.3.3 b) is a user’s need for information, i.e., what information needs to be given via augmented reality about the individual work steps. These requirements for information must be collected in the form of prototypical information objects, which, as in the previous activity, do not necessarily have to be final but can be refined or concretized in a later activity. In the same way, constraints can be derived for the solution of the representation challenges.
- Requirements coming “from relevant ergonomics and user interface knowledge, standards and guidelines” (ISO 9241-210 [80], section 6.3.3 c) may also lead to information objects. However, in many cases they imply constraints

on how the challenges may be solved, and can often lead to specific demands for representation, such as font sizes or distinguishability of colors.

- Both "usability requirements and objectives" (ISO 9241-210 [80], section 6.3.3 d) and "requirements derived from organizational requirements" (ISO 9241-210 [80], section 6.3.3 e) are comparably unspecific concerning what kind of requirements they contain in terms of information objects and constraints. Thus, these sources should be analyzed for both.

During specification of the user requirements, the ISO norm demands that conflicts between contradicting requirements shall be identified and resolved if necessary via trade-offs (ISO 9241-210 [80] section 6.3.4). However, this should not be done with the identified constraints. In this activity, it cannot yet be known how they affect the scene or a combined view on it. In this respect, they differ fundamentally from the functional requirements, which are addressed at this part of the norm. Instead they have to be kept for the next activity and then be resolved for the actual design as discussed in this thesis' section 4.7 Conflicting Resolution Measures.

Results of this activity are the two updated collections of information objects and constraints. At this point they do not need to be completed yet and instead, a refinement takes place during the following activity.

Exemplary Execution

Within this activity, the development team is able to find several user requirements, which lead to either constraints or information objects. One finding from analyzing the user's needs that leads to constraints is that it is desirable for novice mechanics to be supported at every step. A further finding is that mechanics are used to the textual descriptions from the manuals and want to have access to them. Further, most mechanics want to know what step they are in and how many steps there are in total. It is also important for them to be shown the location of the next work step so they do not have to search for it. The updated list of constraints including the constraints that can be derived from these findings, can be found in table 6.4.

Further, two new information objects can be identified that are present in each work step. The first one is a progress indicator that shows the number of the current step as well as the overall number of steps. The second one is a textual descriptions of each work step. For the individual steps, many more information objects can be identified. As the users want to be supported in each work step, these are based on what information users want to have on each individual work step. The updated list can be found in table 6.5.

It is noticeable that there are some information objects present in all work steps, such as the hands or the progress bar. Nevertheless, the development team lists these for each individual step to have a complete overview for each work step when the design is created during the next activity.

Work Step	Constraints
All	Solutions to the challenges must be robust with regard to the simultaneous display of many information objects. (This constrain was already a result of the last activity and now found again from a different source.)
All	The environment can hardly be predicted and thus solutions for the challenges must be robust concerning this uncertainty.
All	The augmented reality system will used be under artificial light and use of colors must be adapted adequately.
All	It is important to help users to reorient after a work step ended and find the relevant place respectively information objects for the next step.

Table 6.4: Collection of constraints including those identified during the activity Specifying the user requirements. Entries in black indicate constraints that were newly added or modified, while entries in gray indicate those that were carried over from a previous table.

6.3.3 Integration into "Producing design solutions"

The integration of the IRMAR framework into the activity "Producing design solutions" (ISO 9241-210 [80], section 6.4) can be divided into two steps, which are both part of "Designing user tasks, user–system interaction and user interface to meet user requirements, taking into consideration the whole user experience" (ISO 9241-210 [80], section 6.4.2).

The first step is to identify new information objects and refine previously identified ones. This can be done as part of the activities described as "Designing the tasks and interaction between user and system" (ISO 9241-210 [80], section 6.4.2.2). Source for new information objects is the interaction between user and system as designed in this activity, whereby especially two parts are relevant. The first part is the information architecture of the system, i.e., the way information and access to information is structured in an interactive system [286]. It may introduce new information objects whose purpose it is to refer to or to include other information objects. For example, an information architecture could require that a specific set of information objects is accessible in each work step. An interaction with a proxy object could then be used to access them, which would make this object an information object as well. In addition, further information objects can be assigned to a work step that originate from other work steps. This is the case when, for example, a preview of the next work steps is given. The second part are the interaction objects, i.e., objects that enable or facilitate user interaction with the augmented reality system. Even if these do not contribute any information

Work Step	**Information Objects**
Step 1 (Figure 6.2)	cover behind lock lever, screw driver, lock lever, hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the cover, textual description of the work step, progress indicator
Step 2 (Figure 6.3)	SRS clip, screw driver, hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the SRS clip, textual description of the work step, progress indicator
Step 3 (Figure 6.4)	cover clip, screw driver, hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the cover clip, textual description of the work step, progress indicator
Step 4 (Figure 6.5)	lock cover, fastening screw, screw driver , hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the screw and cover, textual description of the work step, progress indicator
Step 5 (Figure 6.6)	fastening screw, screw driver, lock lever (must be taken into account to not scratch it with the screw driver), hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the screw, textual description of the work step, progress indicator
Step 6 (Figure 6.7)	fastening screw, screw driver, hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the screw, textual description of the work step, progress indicator
Step 7 (Figure 6.8)	fastening screw, screw driver, hands, virtual model showing required movement of the screw driver, virtual model showing required movement of the screw, textual description of the work step, progress indicator
Step 8 (Figure 6.9)	door cover, lower part of door frame, hands, virtual model showing how to grip and pull the cover, virtual model showing required movement of the door cover, textual description of the work step
Step 9 (Figure 6.10)	door cover, center part of door frame, hands, virtual model showing how to fix the cover position, virtual model showing where to grip and pull the door cover, textual description of the work step, progress indicator
...	...

Table 6.5: Updated collection of information objects, including those identified during the activity Specifying the user requirements. For the sake of clarity, it is cut off after the first nine work steps.

on a work step or a procedural task, they provide information on possible interactions with the system. An example for this would be a break or cancel button. Of course, such objects also need to be seen by a user with as little sensory and cognitive effort as possible.

Furthermore, the information objects found in this or the previous activities must be refined. Whereas they were only collected up to this point, in this activity they must be transformed into a suitable granularity by merging or splitting them. This can be supported through the structure of the previously-created information architecture, depending how detailed it is and its granularity. On this basis, it can also be determined which spatial information an information object carries, i.e., into which of the five classes it must be classified. However, this does not yet imply a specific design, e.g., that a virtual spatial information object must be designed as a 3D object.

After bringing them to the right granularity, the identified information objects must be integrated into scenes where each work step is one scene. For this, it is necessary to find concrete visualizations for them. While this is not covered by this work, there are extensive studies available in the area of information visualization, over which Ware [300], Card [61] or Spence [272] give an overview. Nevertheless, there is one important aspect of information objects that needs to be considered for their visualization: the relevant spatial information must be included. This includes, for example, visualizing a virtual spatial information object in such a way that it can be integrated into the scene in a way that creates a meaningful spatial context. In addition, there may be cases in which a concrete visualization only permits an inadequate solution to the representation challenges. Therefore, it may be necessary to work iteratively during this activity as well to achieve a suitable combination of visualization and representation.

When integrating information objects in a scene, the six representation challenges must be solved as well as possible. The constraints limit how this can be done, since they should not be violated. It may happen, however, that this is not entirely possible if there are contradictions between the constraints. In this case, it must be considered to what extent which of these may be violated. As with design conflicts, no conclusive general answer can be given, and a solution can only be found for a specific context.

Measures to solve representation challenges can be found in the discussion on the challenges in chapter 4 Representation Challenges. It should be noted that the collection of measures is not complete and cannot be completed. Instead, it is foreseeable that further research will lead to new solutions to the challenges. For concrete application scenarios, there may also be specific measures that cannot be found in the literature. Therefore, the sources of unnecessary sensory and cognitive effort described by the challenges must be the focus, and it must be questioned whether a concrete design solution actually minimizes the effort caused by it. As already mentioned in the discussion on design conflicts, it may happen that different challenges cannot be solved to the same extent, and instead it is necessary to find compromises.

Found solutions cannot be generally valid but are always adapted to a context of use. Therefore it is necessary to evaluate the found solutions for practical suitability in the next activity “Evaluating the design” (ISO 9241-210 [80], section 6.5) of a human-centred design process. It may also be necessary to design several solutions in order to compare them with each other.

In summary, the results of this activity are twofold. The first one is a collection of the identified constraints and information objects including their classification in the different classes. Second, an integration of the information objects into each scene is defined. The integration consists of one or more visualizations for the information objects, except direct and possibly indirect physical information objects, as well as one or more designs solving the representation challenges while taking the determined constraints into account.

Exemplary Execution

Creating concrete designs for information objects for all the steps described would be very time-consuming. Especially for the first iteration when no evaluation has taken place so far, the risk is very high that such work would be useless because the design would need to be changed significantly in a later iteration. Thus, the development team decides to design information visualization and representation for only one selected example work step. The first step of the airbag removal task is used for this.

Before designing the information representation can be started, the target hardware must be defined as part of the “producing design solutions” activity (ISO 9241-210 [80], section 6.4). While this is not part of the integration of IRMAR into human-centred design, it is relevant for the further procedure. Two factors are particularly relevant for the selection. First, it is important in a workshop to have one’s hands free when working. Secondly, a direct view of the workspace must be maintained, which may not be distorted or influenced by recording and subsequent playback on a video screen. Therefore, an optical see-through, head-mounted display is chosen.

One additional information object is added to this work step, which is the next step interaction. It is an object that a user must look at while making a tap gesture to indicate the current work step as completed[2]. Together with the information objects identified in the previous activities, this results in the collection of information objects for work step 1 shown in table 6.6.

[2]This is a standard gesture for the Microsoft HoloLens used for the test system at Bosch and thus chosen here

Object	Class
hands	direct physical information objects
cover behind lock lever	direct physical information object
screw driver	direct physical information object
lock lever	direct physical information object
virtual model showing movement of the screw driver	spatial virtual information object
virtual model showing movement of the cover	spatial virtual information object
work step description	spatially referenced virtual information object
progress indicator	detached virtual information object
next step interaction	detached virtual information object

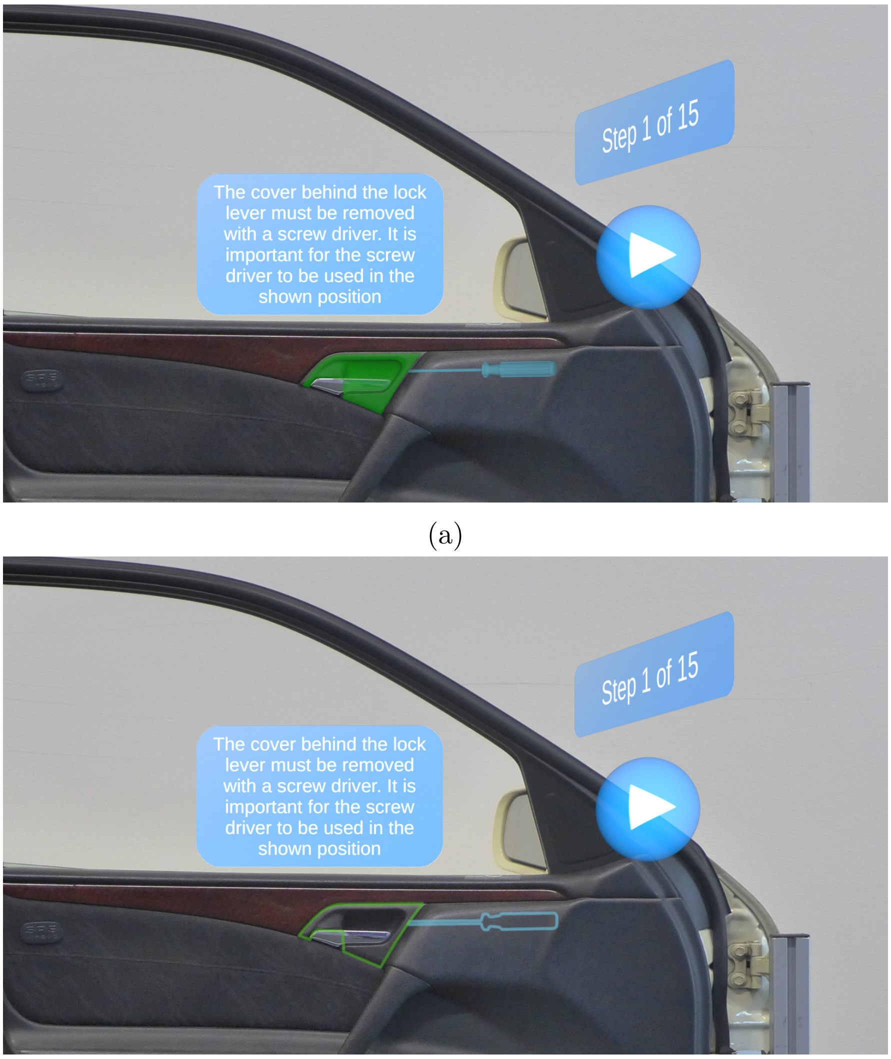
Table 6.6: Updated collection of information objects for work step 1, including those identified during the activity Producing design solutions.

As part of this activity, the information objects are finally assigned to the individual classes. As required by IRMAR, the connection to the physical world is chosen as the criterion which object belongs in which class. The progress indicator, for example, has no spatial relationship to a specific location or another information object and is therefore classified as a detached virtual information object. However, this assignment does not imply a certain visualization, e.g., using a 3D model of the physical screwdriver for the virtual screwdriver.

Several of the information objects have anchors. For the virtual model of the cover, this anchor is the physical cover that must be removed. Similarly, the screwdriver model has the physical screwdriver as anchor. For the description of the work step, the area where the work is done is the anchor. Even if this definition of an anchor is somewhat fuzzy, it is still valid since it is not required that a precise spatial delimitation is given. The other information objects do not have an anchor because they are either direct physical or detached virtual information objects.

The two spatial virtual information objects also have meaningful spatial contexts. These are the positions that their physical counterparts, the screwdriver and the lock cover, must be moved along during the work step.

After their definition is completed, these objects are brought together in a scene to form a concrete representation. The first information objects to be placed are the models of the cover behind lock lever and of the screwdriver. These are both spatial virtual information objects, i.e., their concrete positioning and orientation conveys information to the viewer via a spatial context. Since there is no certainty about a suitable visualization yet, two different visualizations are created that can later be compared with each other when evaluating the design during the following activity "Evaluating the design" (ISO 9241-210 [80], section 6.5). For the first visualization, both objects are designed as 3D models whose shape is realistically based on their physical counterparts and which have a solid appearance. In order to create a contrast to the interior lining of the door, a green color is used for the



(a)

(b)

Figure 6.18: Example representation for the first work step with a) complete model and b) outline visualization. In these images the physical information objects hands and screw driver are not included. (The images were post-processed in a 3D graphics program, as photographs of the optical see-through setup had a very low visual fidelity.)

virtual cover instead of dark gray, which is very poorly displayed on an optical see-through head-mounted display (see 4.1.2 Clarity Challenge: Color Perception). The screwdriver is displayed in a different shade of green for the same reason. For the second visualization, only the schematic outline of the objects is displayed while the colors used are the same. This is done to keep as much as possible visible from behind the objects (see 4.3 Visibility Challenge). In both cases, the information objects are placed in their meaningful spatial context, which also solves the information linking challenge for the cover plate (see 4.6 Information

Linking Challenge). The IRMAR framework does not give any indication about the suitability of these visualizations. Instead, it helps to define to how they are embedded into the scene.

In both visualization variants, the necessary movements of the screwdriver and the cover plate are represented by animations of their two virtual counterparts (see 4.5.1 Motion Cue Challenge: Dynamic Motion Cues). In the first visualization, the virtual model of the cover occludes the physical cover and partly the lock lever, such that the visibility challenge is not solved (see 4.3 Visibility Challenge). The movements during the animation slightly ameliorates this, since it allows a partly unobstructed view during some parts of the animation, but it does not completely solve the challenge. The situation is also similar with the model of the screwdriver, which may cover the physical screwdriver when completing the work step. Therefore, both virtual objects made partially transparent in order to let the underlying objects shine through (see 4.3.1 Visibility Challenge: Resolution by Transparency). Shifting them to a nearby location is not done because this would destroy their meaningful spatial context (see 4.3.2 Visibility Challenge: Resolution by Moving). Aside from that, it is known from the collected constraints that a larger amount of information objects can be expected. By shifting, the information linking challenge (see 4.6 Information Linking Challenge) as well as the orientation challenge (see 4.4 orientation Challenge) are aggravated, especially with many present information objects. However, this is not a problem for the currently designed step, but this is to achieve a certain consistency throughout the entire task. For the second visualization, occlusion of information objects is not a problem because only an outline of the virtual cover and the screwdriver are shown, which do not occlude objects behind them. Thus, the visibility challenge does not need to be explicitly handled in this case.

The next information object to be placed is the description of the current work step. A fixed position slightly above the work area is selected as its position. Because it is a spatially-referenced information object, it can be moved without a loss of spatial context and instead can be placed at a position that is sure to not occlude any other information objects (see 4.3.3 Visibility Challenge: Moving Information Objects) and is close to its anchor (see 5.2.2 Information Linking Challenge: Visual Connection). A salient type of visual connection, e.g., with a direct visual connection, was not used, since proximity is an adequate way of assignment and there is no possibility of confusion (see 4.6 Information Linking Challenge). From the constraints, it is known that the background is difficult to control and can contain anything that is found in a car repair shop. Since the text is located at the height of the window and thus in front of such an uncontrollable background, it must be embedded accordingly to ensure good legibility throughout (see 4.1.3 Clarity Challenge: Text Legibility). As a solution, it is placed on a blue billboard and the font color is set to white to create a good contrast (see 4.1.3 Clarity Challenge: Text Legibility). When using the system under artificial light as defined by the constraints, blue appears as a suitable color. This is not necessarily the case under natural light due to its comparably poor perceptibility (see 4.1.2 Clarity Challenge: Color Perception). Finally, the scaling is chosen so that the text has the height of about 1°per line from a typical viewing position. This guarantees an adequate size for good legibility (see 4.1.3 Clarity Challenge: Text Legibility).

As last information objects the progress indicator and the next step interaction area are placed in the scene. However, as detached virtual information objects, they do not have any connection to other information objects, a point or an area in the scene. Thus, they can be placed anywhere without loss of spatial context. To support visibility, they should not be placed at a position where they are likely to occlude other information objects (see 4.3 Visibility Challenge). Instead, to support orientation and keep track of information objects, they are bound to the user's body and are both positioned slightly to the front and to the side of the right shoulder (see 4.4.3 Orientation Challenge: Keeping Track). They are also slightly rotated to directly face the head when it is turned. This placement is not limited to the first work step, but instead valid for all. This way a user can remember the position once and always find them at the same relative position (see 4.4.3 Orientation Challenge: Keeping Track). For color and text size the same design is chosen as for the work step description (see 4.1.3 Clarity Challenge: Text Legibility). The two designs can be found in figure 6.18.

One of the constraints that has not been addressed yet is that users want to be guided to the current work area. This constraint addresses an aspect of the orientation challenge. It is solved by including an attention funnel in the middle of the field of view that guides towards it (see 4.4.1 Orientation Challenge: Spatial Orientation). However, it is only shown whenever a user does not look at the work area. This way it is ensured that it does not occlude other information objects (see 4.3 Visibility Challenge). A depiction can be found in figure 6.19.

6.3.4 Integration into "Evaluating the design"

As part of this activity, it is evaluated in how far the identified information objects and the created representation of the information objects meets the needs of the users. This includes whether the representation challenges have been solved to a satisfactory degree. If several designs were created with different solutions to the representation challenges in the previous activity, they are compared with each other to identify their strengths and weaknesses. Recommendations as to how this investigation can be carried out are outside the scope of work and also dependent on the concrete situation. Therefore, only the results that need to be achieved will be discussed here. In contrast to the other activities, concrete references to its aspects as described in ISO 9241-210 are not possible. The reason is that these describe different test procedures, which are, in principle, all suitable for obtaining necessary results. Therefore, in the following, only little will be explained on how to achieve the results in this step, but much more on which results have to be achieved. Details on possible evaluation methods can be found in ISO 9241-210 itself (ISO 9241-210 [80], section 6.5), but also in the general literature like, e.g.,

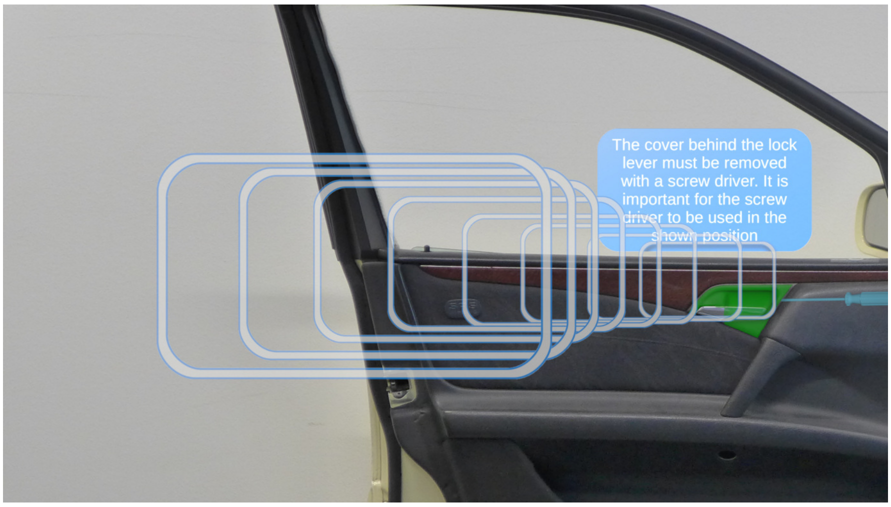


Figure 6.19: An example of the included attention funnel. Normally, it would not be displayed with such a small angular distance to the work area but is done here for the sake of a compact illustration. (the image was post-processed in a 3D graphics program, as photographs of the optical see-through setup had a very low visual fidelity)

in Goodman et al. [125] or in Sauro and Lewis [256].

One aspect of the results is whether the appropriate information objects have been collected. It might be that some were not identified, or that others were included but are not relevant. Also, an inappropriate granularity may have been chosen, i.e., multiple objects were combined into one or one was split into multiple ones, without this being useful. Thus, one result of this activity needs to be information on where information objects are needed but were not identified, are included but are not relevant, and where the granularity is incorrectly selected.

Another aspect are the representation challenges and how well they are solved. It needs to be determined in which scenes which of the challenges have not yet been adequately addressed. Part of this is whether the correct constraints have been identified and if the identified constraints are actually required. An insufficiently solved challenge is an indication that the constraints that lead to the chosen solution for the challenge needs to be further examined. It is possible that these have not yet been sufficiently understood or that missing constraints must yet be determined. Where a design conflict was resolved, i.e., where measures to solve challenges contradicted each other, it is also relevant whether the right balance was found. If this is not the case, it hast to be determined which challenges need to be addressed more thoroughly and which challenges can be addressed less thoroughly. If several designs with different solutions for the challenges were created in the previous activity, it has to be clarified which better suits the users' needs. If possible, strength and weaknesses have to be identified.

In summary, the results of this activity are insights in two aspects. The first one is whether the right information objects were identified in the appropriate granularity during the previous three activities. The second aspect is whether the challenges have been adequately addressed and, if not, where improvement is needed. Constraints, which contributed to the selection of the solutions, have to be specifically considered in the next iteration. These outcomes can then be used in an iterative human-centred design to refine the results of the other activities.

Exemplary Execution

After the information representation has been designed during the previous activities, it is evaluated by the development team. They select to employ user-based testing. Thus, technicians from a workshop are asked to perform the first step of the airbag removal using the designed augmented reality system. Both variants, the solid and the outline models, are compared during testing. The technicians are observed how they carry out the repair task, and afterwards, they are also asked about their subjective impressions of the system. The collected results are then analyzed in terms of findings on information objects and on how well the representation challenges are solved.

The result from the comparison of the two variants is that on average, technicians tend to prefer the solid model over the outline. However, this does not provide any indication on how well the challenges are solved. Instead, it implies a preference for a specific information visualization. However, since the solutions to the challenges may depend on the visualization, the result is relevant for information representation. Indications of how well the challenges were solved are instead based on observations made by the development team during the test and the individual judgments by the technicians. This way the following five findings are made concerning information objects and solutions of the challenges:

- The lock lever needs to be lifted during the removal of the cover. This is not shown so far and caused difficulties when working on the work step as noticed through the observation. This is an indication that further information objects are needed.
- Showing the animation on how to remove the cover behind the lock lever is perceived as not adequate to indicate how the parts need to be moved. Instead, the technicians expressed the wish to have a static display of how to move it that is continuously present. This is an indication that the motion cue challenge must be solved differently via a static motion cue, which also includes adding a further information object.
- Even though the technicians preferred the variant with the solid virtual model of the cover, it seems to be too opaque. During the observation, the technicians showed signs of having problems seeing the physical cover underneath it. This was confirmed by some of them expressing the wish for it to be a little bit more transparent. The implication is that the visibility challenge must be better solved.

- The shade of the background color blue for the description text and the progress indicator was described as too bright by the technicians. This is an indication that to better solve the clarity challenge, the color must be dimmed a little bit.
- The attention funnel was perceived as too invasive and the technicians wanted a more unobtrusive guidance to the work area. This is an indication that a different measure for the orientation challenge must be used.

These insights can be used by the development team in further iterations of a human-centered design process to improve information representation. Now that a visualization variant has been selected for the cover, it is possible to concentrate on designing information representation for this variant. For this purpose, the obtained results help for various activities. While, for example, the missing indication of the lock lever movement is relevant in the activity “Specifying the user requirements” (ISO 9241-210 [80], section 6.3), the shade of blue used as background color for the texts only becomes relevant in the activity “Producing design solutions” (ISO 9241-210 [80], section 6.4).

7 The Augmented Reality Procedural Task Modeling Language[1]

> In this chapter a second application of the IRMAR framework is shown. This time the focus is the creation of a language for authoring augmented reality support for procedural tasks in the automotive domain. It contributes to achieving the goal defined in the motivation to demonstrate the application of the IRMAR framework including the representation challenges.

In the previous chapter, the potential of the IRMAR Framework to support user interface design was shown. However, there are also other application areas that are less obvious. One is authoring of augmented reality documentation for procedural repair and maintenance tasks in the automotive area. For this purpose, the definition of information objects including the different classes of spatial connection as well as the description of the representation challenges in IRMAR are a good foundation.

The form of technical documentation used in the automotive context often has to take many different variants of one product into account which differ in details [274]. Similarly, many products share large parts of their documentation. When using augmented reality as a documentation medium, the combination of physical environment and virtual data makes any inconsistency immediately visible. It is therefore difficult to describe a process once and then transfer it to similar models or products, which, for example, differ in details such as the position of screw holes. At the same time, depending on the context of use, different views of the same description are needed. A certain piece of information, for example, may be needed by a novice but is only distracting for advanced technicians. To the best of the author's knowledge, no appropriate approach exists to create such documentations with a focus on supporting appropriate information representation. Therefore, in the following the Augmented Reality Procedural Task Modeling Language (ARPML)[2] is presented, which uses the IRMAR Framework as a basis for

[1] This chapter is based on a previously published paper *ARPML: The Augmented Reality Process Modeling Language* [201]. My contribution was the development and definition of ARPML.

[2] In previous publications it was called Augmented Reality Process Modeling Language. However, it was renamed in order to make the purpose clearer.

authoring augmented reality documentation in the automotive domain.

The structure of the chapter is as follows. First, existing concepts to model tasks and processes in augmented reality that influenced ARPML are briefly discussed. Then, the requirements for the modeling language are derived and motivated. Finally, the elements of the language are explained, with a special focus on describing the influence and the relevance of the IRMAR framework. The presentation here is focused especially on the overall structure of ARPML, while technical aspects are omitted for length reasons. In particular, the description of the XML implementation, details on the runtime behavior, handling of inputs from users or sensors and technical challenges like tracking and object detection are not discussed.

Throughout the presentation of ARPML, examples are used to better explain its individual elements. These are again based on the airbag removal task, which has already been used to illustrate the integration of IRMAR into ISO 9241-210. Again, not everything is explained in detail, but instead, selected aspects are used to show how ARPML can be applied to author augmented reality documentation for procedural task support. For this example, a second variant of door cover is added to the airbag removal task, which differs in some relevant details to the one from the example use case. The discrepancies between them, however, are not large enough to justify the creation of two independent task descriptions. The reason for such differences are often different equipment variants or changes due to a model updates. The explanations are done using a graphical editor for ARPML that supports the authoring process[3].

7.1 Existing Modeling Approaches

At first, some fundamentals have to be discussed for the development of ARPML, which are not included in chapter 2 Fundamentals of Augmented Reality for Procedural Task Support. The reason for listing them only here is that they are not relevant to the development of the IRMAR framework. Existing approaches to model processes and tasks for augmented reality range from relatively simple, linear descriptions with work steps succeeding each other [95], to complex models, which are e.g., based on activity diagrams [32] or state diagrams [172]. Three of them are particularly influential for ARPML.

The first is the generic document model by Köpfle et al. [274] that was developed for augmented reality manuals in the automotive field. The central elements of this approach are work steps, which contain specific work instructions, each of which can reference other work steps that need to be completed before it can be started. To be able to reuse the same manual for as many products as possible, each work step is only valid for a range of products or variants of a product. When executing a task for an actual product, e.g., a car with a specific equipment package, the last work step of the intended task is selected. The needed work

[3]This editor was created as part of a Bachelor's thesis supervised by the author [243]

steps are then recursively resolved and brought into a linear order while those not valid for the product or variant are omitted.

For the information shown as part of the work steps, instantiatable templates are used [162]. These are graphic elements such as CAD models of tools or components that are defined once but can be used in arbitrarily many work steps. Adaptations, such as changing positions, are possible via parameters. Further, multiple instances can be used per work step. Templates can have two animation directions: one is to show the actual operation while the other is to show the reverse operation. This is necessary because automotive repair involves the reassembly of many parts, for example, reattaching a cover that has been previously removed. Simply playing a reversed animation would not be sufficient because operations are usually not done with an exact reverse movement.

The second relevant previous approach is the use of task models to describe augmented reality support for work processes. Even though their name sounds similar to procedural task, they describe something different, namely an approach to modeling tasks in a tree structure. The overall task or goal is at the root, which is recursively refined into subtasks or subgoals until they are sufficiently fine-grained and not broken down any further. The leaves of this tree therefore represent the single work steps, while the inner nodes are subtasks completed through the execution of their child tasks. Task models have been used to describe augmented reality support for aircraft maintenance by De Crescenzio et al. [75] and Wang et al. [298]. In general, task models have a wide area of applications and approaches like CTTE [219] or VTMB [41] have previously shown the successful application for modeling complex tasks.

The third particularly influential work is the Augmented Presentation and Interaction Language (APRIL) framework which is an approach to authoring augmented reality information presentation [172]. Its terminology is based on theatre plays. Hence, a presentation created with APRIL is referred to as a story. This in turn is divided into scenes in which actors appear. The modeling is done via UML State Charts, where each state is a scene and the whole diagram is the story. Actors are individual virtual objects that exhibit certain behaviors, e.g., showing a specific animation. They are placed on timelines, of which each scene has three. The enter timeline is played once when a new scene is entered. Afterwards, the do timeline is played and is repeated until the scene shall be left, at which point the exit timeline is played once. The places where the story takes place are called stages, which are spatially-registered reference systems.

7.2 Requirements for ARPML

Before the modeling approach is presented, relevant requirements that were taken into consideration during the creation are introduced. However, only a selected subset is included here, since many are not related to this work, such as what types of measuring instruments need to be considered. Instead, the focus here is on those that are relevant for the use of the IRMAR framework or those that have a very large impact on the overall structure of ARPML. Sources for the requirements are, where indicated, scientific literature and otherwise operational needs at the Robert Bosch GmbH, where this work was carried out.

1. It must be possible to model procedural task consisting of work steps including transitions and dependencies between them [274]. This is a basic prerequisite for describing procedural tasks at all.

2. It must be possible to show different instructions based on user interactions, diagnosis results, and similar inputs because automobile repairs are often not linear but depend on information gained during the repair, e.g., diagnostic results.

3. It must be possible to include various types of information objects such as texts or 3D models in one work step because technical authors use a variety of media types to describe repair and maintenance instructions [274].

4. It must be possible to have multiple views, i.e., selections of included work steps and shown information, on one procedural task depending on the product, variant of the product or the user's need for information [274]. Each view must be adapted to the usage situation as well as possible to not have inconsistencies between the physical world and virtual information for it. Besides obvious cases where incorrect work instructions are shown, this also concerns inconsistencies such as incorrectly-positioned virtual objects or unnecessarily shown information. These can cause sensory and cognitive effort [86, 245, 221].

5. It must be possible to show augmented reality support for a procedural task described with ARPML using multiple device types. These may vary concerning display type and size.

6. Visualizations for shown information and work steps must be made reusable for multiple products or variants of a product [162]. This is needed to keep effort for authoring augmented reality manuals low because costs for content creation remains a large impediment for the practical application of augmented reality [204].

7. Solving the representation challenges for the views on procedural tasks must be facilitated. Otherwise it would be likely that information is represented in a way which causes unnecessary sensory and cognitive effort.

7.3 Variant Handling in ARPML

As described in the requirements, the information presented via augmented reality must be very accurate concerning its fit to the physical world. For this purpose,

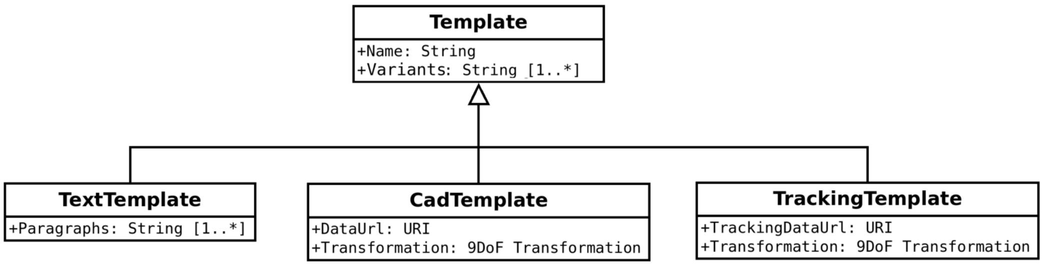


Figure 7.1: UML definition of the ARPML templates

variants are included as a simple mechanism to mark the scope of ARPML elements. These are short strings of characters, e.g., *equipment_variant_a* that can arbitrarily be assigned to each ARPML element that supports variants. The mechanism is thus similar to tagging. Which elements support the variants mechanism is explained in the following sections.

When a concrete instance of a procedural task description is created, the elements are included based on their assigned variants with the use of logical expressions. For example, an equipment variant can be selected by an OR link of two variant tags, like *equipment_variant_a OR equipment_variant_b*, including all elements that are assigned with one of the two variants *equipment_variant_a* or *equipment_variant_b*. Whether additional variants have been assigned to them is irrelevant. The term variant was chosen to express that an element is in the scope of a particular variant of a procedural task description.

7.4 Templates in ARPML

Templates are information-carrying units like 3D models, texts or images that can be instantiated and shown as part of augmented reality support for procedural tasks. They are not limited to specific types but instead an implementation of ARPML can provide the needed ones. The reason for allowing this flexibility is that tooling and development pipelines often already exist and need to be integrated. In the ARPML implementation that was created as part of this work, three types of templates are included: text, 3D and tracking templates. When creating a template, the parameters determined by the type must be provided. Each template type requires at least a name and possible variants. Additionally, the text templates have the individual paragraphs of the texts, while 3D templates also have a reference to the associated 3D model and a spatial transformation for it.

The tracking templates are somewhat different from the other templates. They provide information on how to localize a position, e.g., that of a marker, in the workspace. Besides name and variants, they also include a reference to information for the tracking system and a spatial transformation that translates from the tracking system's reference system to one defined by the author. Their use is explained in more detail in the later sections.

The here described templates are similar to those defined by Knöpfle et al. [162]. However, there are some differences in regards to the scope of a template. In ARPML, general types of templates are defined that must be parameterized, e.g., with 3D models or texts, to create a template. The parameters that can be set for a template instance are independent of the type. Knöpfle et al. do not have such types. Instead each template is independent from the others. For this purpose, each has its own individual data, behavior and parameters associated with it. Thus, template types and the way templates are created do not have an equivalent in Knöpfle et al.'s work, while ARPML templates are more comparable with their templates but with less possible parameterization.

A depiction of the template types defined for the ARPML implementation created for this work can be found as an UML class diagram in figure 7.1.

7.4.1 Application Example for Templates

In this example, individual templates are created for the airbag removal task, more precisely for the first work step (see figure 6.2). In order to create the mapping between the virtual and physical world required for augmented reality, a tracking template is created. This does not define the properties of a tracking system, but rather integrates a configuration into ARPML and makes it available for later use. Specifically, a name (*Door Tracking*), a reference to the tracking configuration (*file:////./trackingdata.dat*) and the variants (*door_cover_a* and *door_cover_b*) are defined. In order to compensate for specifics of the tracking system's reference system, a translation (tilted 90°around the Y-axis) is set. The corresponding entry is shown in figure 7.2.

Additionally, templates for the cover plate and the screwdriver are created. These are both three dimensional objects and thus, two 3D templates are used. For each of them a name (*Cover* resp. *Screwdriver*), the valid variants (*door_cover_a* and *door_cover_b*) and the data URL (*file:////./griffschale.obj* resp. *file:////./schraubenzieherschlitz2.obj*) are entered. A transformation from their local coordinate system is not needed. For the corresponding entry in the ARPML editor, see figure 7.3.

One template is also included to to give information on the progress, i.e., how many steps have been finished out of how many. Since the implementation of ARPML used here does not have a specific template type for this functionality, it must be solved via a text template. It has a name (*Progress Step 1*), the valid variants (*door_cover_a* and *door_cover_b*) and a single paragraph (*Step 1/17*). Similarly, a description of the current work step is created with a name (*Description Door Cover A Step 1*), the valid variants (*door_cover_a*) and a single paragraph (*Door Cover A - Step 1: The cover ...*). Due to the different door covers, this description of the work step is only valid for the variant with door cover A. Thus, a further text template with the correct description for door

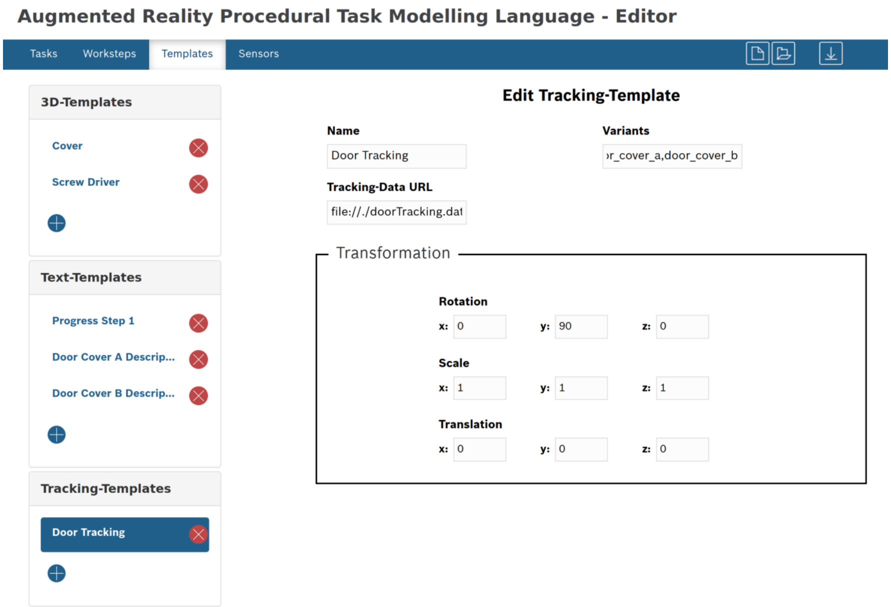


Figure 7.2: Defining a tracking template in the ARPML editor

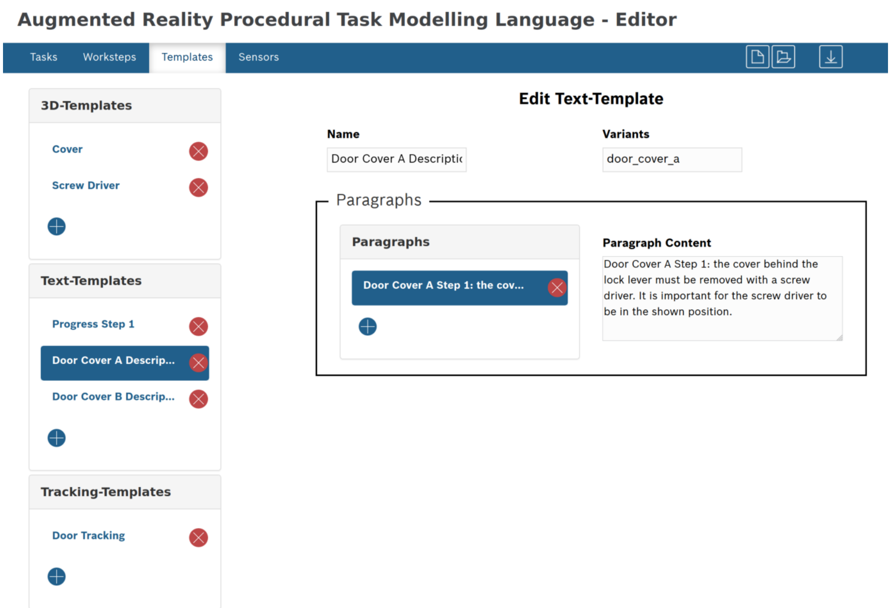


Figure 7.3: Defining a text template in the ARPML editor

cover B is also added. For the corresponding entry in the ARPML editor, see figure 7.4.

While it would be beneficial, the implementation of ARPML created as part of this work, does not have a template type that allows to represent a user's hands and therefore cannot be included here.

7.5 Information Objects in ARPML

While the templates are heavily based on the work of Knöpfle et al. [162] and do not offer much novelty, there is a big difference when it comes to their instances. In ARPML these are *information objects* in the sense of the IRMAR framework including their described properties, characteristics and the related representation challenges. This has a significant impact on the design of the modeling language discussed in the following.

To be able to use information objects in ARPML, it must be possible to fully describe their properties. One important aspect of this is the ability to assign information objects to one of the five classes of spatial connection, which thus become part of ARPML. This also creates the opportunity to model physical objects in IRMAR. Further, information objects in ARPML can also have anchors as in the IRMAR framework. However, for practical reasons, only other information objects but no positions or areas can be set as anchors. Therefore, if necessary, such referenced areas must be modeled as information objects. Each information object can also be assigned a pose and a scale in a reference system as their mounting. This pose of information objects can change over time as described by the characteristic physical change, and may be necessary, for example, to show that a certain motion is required to install a component. To integrate this aspect, key positions with a pose and a scale can be defined for arbitrary times. Between them, the values are interpolated at runtime. For spatial virtual information objects, a meaningful spatial context can also be defined through the pose mechanism. A definition of information objects as an UML class diagram can be found in figure 7.5.

For ARPML, information objects are extended with variants. This way when creating an instance of an ARPML task description, only the relevant ones are included while those irrelevant are left out. With this mechanism it possible to omit objects that are not information objects in the context of a particular instance. This can be useful, for example, if two equipment variants of a product are identical except for a few parts. In this case, only the differing parts must be defined twice while for all other information objects, no redefinition is needed.

The way a view management system chooses the concrete representation of information objects is not part of ARPML. The simplest approach, which does not try to solve the challenges, is to leave all objects at their specified mountings and use only one fixed visualization for motion cues and for information linking. However, a more sophisticated approach could be to optimize representation to-

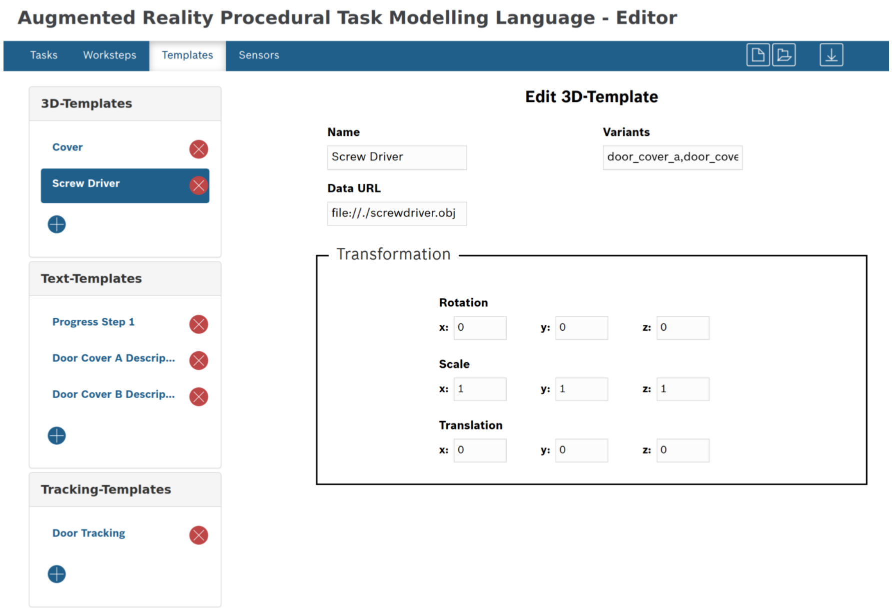


Figure 7.4: Defining a 3D template in the ARPML editor

wards solving the six challenges based on the defined information objects, their classes and relationships. Possible examples of this are orienting texts towards the viewpoint to increase text legibility or arranging information objects in order to ensure visibility. To enable such functionalities, it is particularly relevant that during authoring of a procedural task in ARPML, the necessary information can be specified, which is achieved by including all properties of information objects. Finding a good representation that solves the challenges as well as possible can then be formulated as an optimization problem, like shown for other user interfaces by Oulasvirta et al. [212, 213]. For augmented reality, some layout management systems have already been proposed that automatically optimize individual aspects of the user interface, such as the arrangement of labels [36, 21, 127].

For solving the clarity challenge, it is especially relevant to include the classes of spatial connection for the information objects. This way, a view management system has a basis to decide which objects can be moved or scaled to what extent without losing spatial context in order to ensure a higher clarity. Examples for possible measures are orienting objects towards the user, enlarging them or moving them in front of better-suited backgrounds. Further, the use of different template types also supports creating an adapted presentation of the information objects. For example, the graphical appearance of text templates can be modified based on the situation at hand by an augmented reality system. However, this is less an aspect of representation than of visualization.

Concerning the visibility challenge, it is not sufficient to simply place virtual objects at fixed positions in a user's view and create a scene with a static layout where every object is visible. One reason is that many variants of one procedural task may exist, each with a partially different set of information objects, which may even be used on different device types. Another reason is the viewpoint, which is controlled by a user whose movements are hard to predict. Instead, the view management system needs to be informed about the existing information objects, including the physical ones, and their positions. By doing so, it can then ensure a solution to the challenge. The class of spatial connection for information objects informs a view management system on what spatial context is lost when manipulating, e.g., relocating, resizing or rotating, information objects. This can help deciding which objects can be moved to ensure visibility.

Regarding the orientation challenge, it is important that all information objects are included. This makes it possible to automatically give hints on their position, e.g., when they are outside the field of view. The inclusion of physical information objects makes them available to such a mechanism as well. By using information objects and not, for example, 3D models, whose structure is solely technical in nature and, possibly, can also combine many information objects into one, the necessary granularity is given to reference them individually.

Solving the motion cue challenge is supported by the capability to specify the pose and scale of information objects over time. This way, the relevant information about movements that objects must perform is provided to a view management system. At the same time, the freedom remains to display the motion cues as required in a particular situation. In contrast, animations embedded in models do not distinguish between object movement and visualization and can hardly be translated into other motion cues.

Having anchors supports a view management system to handle the information linking challenges by one of the presented measures, such as showing a visual connection or spatially assigning information objects to their anchors. Again, this approach is more flexible than including such connections as graphical elements when authoring the augmented reality support.

7.5.1 Application Example for Information Objects

With this application example for the user of information objects in ARPML, the example from the templates is continued. In the first step of the airbag removal task, a total of seven information objects are included. Two of them are the physical cover plate and the screwdriver, which are direct physical information objects. They are created as instances of the templates *Cover* and *Screwdriver* with their class of spatial connection set to *Direct Physical Information Object*. Including them as information objects does not imply that these objects will later be graphically displayed to a user but instead, they will be used as placeholders at runtime. The 3D models referenced in the templates provide a view management system with the size and shape of the physical objects. Additionally, names (*Physical Cover* resp. *Physical Screwdriver*) and variants (*door_cover_a*, *door_cover_b*) are also specified. The anchors are not defined since they cannot have one.

Further, virtual information objects that show how the cover plate has to be

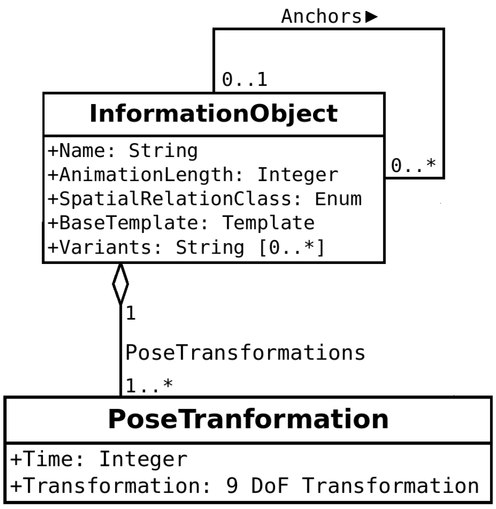


Figure 7.5: UML class diagram of information objects in ARPML

removed with a screwdriver are added. Again, the same templates are used for the virtual cover plate and screwdriver. Name (*Virtual Cover* resp. *Virtual Screwdriver*), class of spatial connection (*Spatial Virtual Information Object*) and variants (*door_cover_a*, *door_cover_b*) are set. The previously-defined direct physical information objects are set as their anchors. However, there is a big difference in the position. Since the cover plate shall be removed with the screwdriver, motions are defined to indicate this by defining key poses to move both virtual objects similar to how the user has to move the physical ones. Two key poses are added to each object: one at the built-in position and the other at a position where the cover is detached. The poses between are generated by an ARPML implementation. The resulting definition of the virtual cover plate is shown in figure 7.6.

Finally, three more information objects are added based on the three text templates. Names (*Progress Step 1*, *Door cover A Description Step 1* and *Door cover B Description Step 1*) and variants (*door_cover_a* for the progress indicator and the description for door cover A, and *door_cover_b* for the progress indicator and the description for door cover B) are set for all three. While the descriptions of the work step are spatially referenced virtual information objects, the progress indicator is a detached virtual information object. Therefore, an anchor must also be set for the work step descriptions. In both cases, the physical cover plate is selected for this since it is to be removed in this step. Unlike the virtual models of the grip plate and the screwdriver, the information about the pose respectively mounting has no special meaning beyond a suitable positioning. Since they do not indicate any motion, one key pose is sufficient.

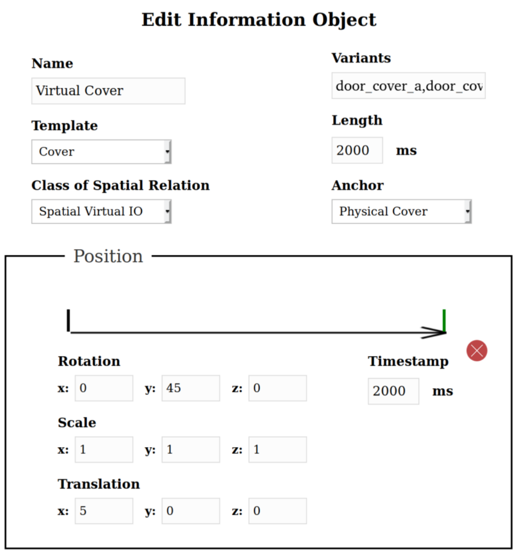


Figure 7.6: Defining the virtual cover as information object in ARPML

7.6 Work Steps in ARPML

A *work step* element in APRML describes the individual steps of a procedural task. Each of them has two directions in which it can be carried out: the actual operation, e.g., removing a part, and the reverse operation, e.g., reattaching the previously removed part. As already described by Knöpfle et al.[162], these two directions are necessary because automobile repairs often require parts to be reassembled that were previously removed during the repair. However, it is not possible to simply perform each step exactly backwards again, because there are differences in how operations must be performed, e.g., different sequences for loosening and subsequent tightening bolts, that make two independent directions necessary. Some operations may not even have a reverse step, meaning that the reverse direction is empty. Each of these two directions is equivalent to an IR-MAR scene. They combine the information objects needed to perform one work step respectively reverse it. Thus, the representation challenges must be solved for each of these two directions.

A direction is further divided into stages which create frames of reference for the placement of information objects. This is similar to an approach previously used by Ledermann and Schmalstieg [172] for the APRIL framework. Each of them is assigned a tracking template which provides the information on how the stage can be registered in the physical world. This is the reason for the special role of the tracking templates since they cannot be instantiated to information objects, but instead are referenced by stages. By tracking physical information objects, e.g., with a marker placed on them, the stage mechanism can also be used to include their virtual representations in a work step. In the ARPML implementation created as part of this work, virtual models, i.e., information objects created from 3D templates, must be added to such a stage to represent the physical objects. However, because the template mechanism is open to new types of templates, it would be possible to include a template type which automatically transfers the shapes of physical information objects without the need for a virtual model.

Besides the aspect of stages as a connector between virtual objects and the physical world, they can be also used to support solving the orientation challenge. They allow for positioning information objects together at a constant relative position while moving independently from those in other stages. One stage, for example, could be positioned relative to the user's body and contain virtual detached information objects that are present for each work step and thus helping a user to easily keep track of them. A view management system can also address a hole stage and point it out to the user to support finding the information objects positioned on it.

Each stage has three timelines. These were previously presented in a similar way as part of the APRIL framework [172]. One is the enter timeline which is played once when the work step that the stage belongs to, is entered. Then, the do timeline is repeated until the step is left, which causes the exit timeline to be played once. Information objects can be placed on these timelines and be shown, hidden and transformed over time. This is achieved with the key poses already explained for the information objects. As already explained, this mechanism is not intended to replace animations and, instead, define information on motion as a basis for motion cues. For direct or indirect physical information objects, the timelines are mostly irrelevant because they are either completely or, for indirect physical information objects, largely not manipulatable through an augmented reality system.

While some of the work steps depend on further work steps being completed beforehand, others can be started without such dependencies. The example task to remove the airbag from the door shows this very well. While removing the covers and screws from the inside of the door (steps 1 to 7) could be done in an mostly arbitrary order with only some dependencies between the work steps, all of them need to be completed before the door cover can be removed (step 8). To include this into ARPML, each work step has a list of other work steps that must be completed before it can be started. This list is empty for those that have no such dependencies. There may never be a full dependency cycle because this would result in a deadlock and no work step in that cycle could ever be executed.

Similar to templates and information objects, variants can be assigned to work steps. In order to allow a finer selection that is still more general than the one on the level of individual information objects, stages can have variants assigned. When an instance of a procedural task description is created, references to work steps that are not part of the instance are removed. When, for example, work step A depends on work step B, and B depends on C, then A also depends on C to be completed. However, if B is not part of an instance, the reference to B is

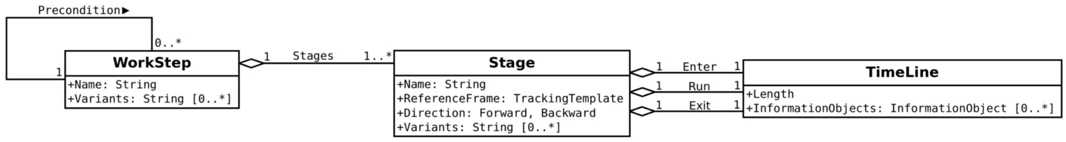


Figure 7.7: UML definition of work steps in ARPML (details on data input are omitted)

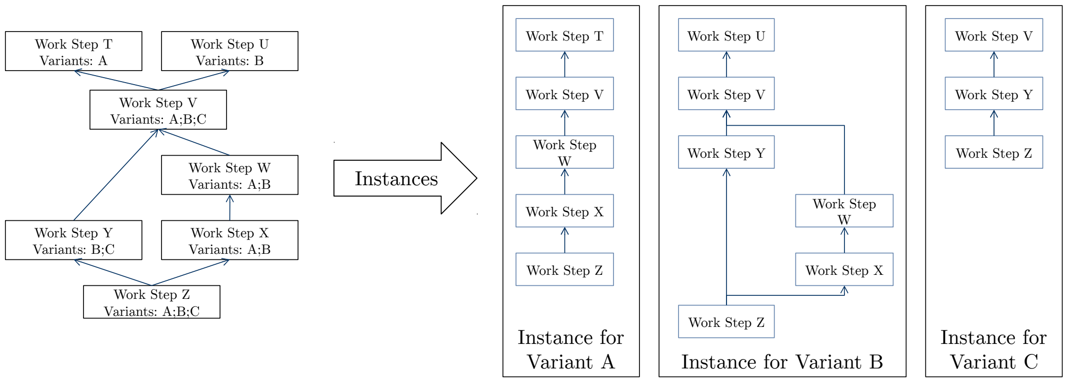


Figure 7.8: References between work steps in ARPML. The example shows how references for work step Z depend on the variant chosen for the instance.

removed from A, and thus there is also no longer a dependency on C (see figure 7.8). A similar approach was previously presented by Stock and Weber [274].

Work steps also have various methods for data input, for example, to enable work step completion and case discrimination. Since this mechanism is complex in itself and has no relation to the use of IRMAR, it is not included in this description.

A definition of work steps described in this section as UML class diagrams can be found in figure 7.7.

7.6.1 Application Example for Work Steps

Again, the previously begun example is continued. Here, the individual information objects are combined into one work step with the name *Remove Cover Plate*. First, the work step is named and the valid variants (*door_cover_a* and *door_cover_b*) are set. Stages are then created for the forward and backward directions of the work step based on the door tracking template (*Forward Stage Step 1* and *Return Stage Step 1*). Creating more than one stage per direction would be possible, even with only one available tracking template. However, this would have no practical effect, since all stages would have the identical reference system. Two timelines are used in this example, the enter and the run timeline. The run timeline has a length of 2000 ms, where the pose changes of its information objects are repeated every two seconds. The corresponding entry in the ARPML editor

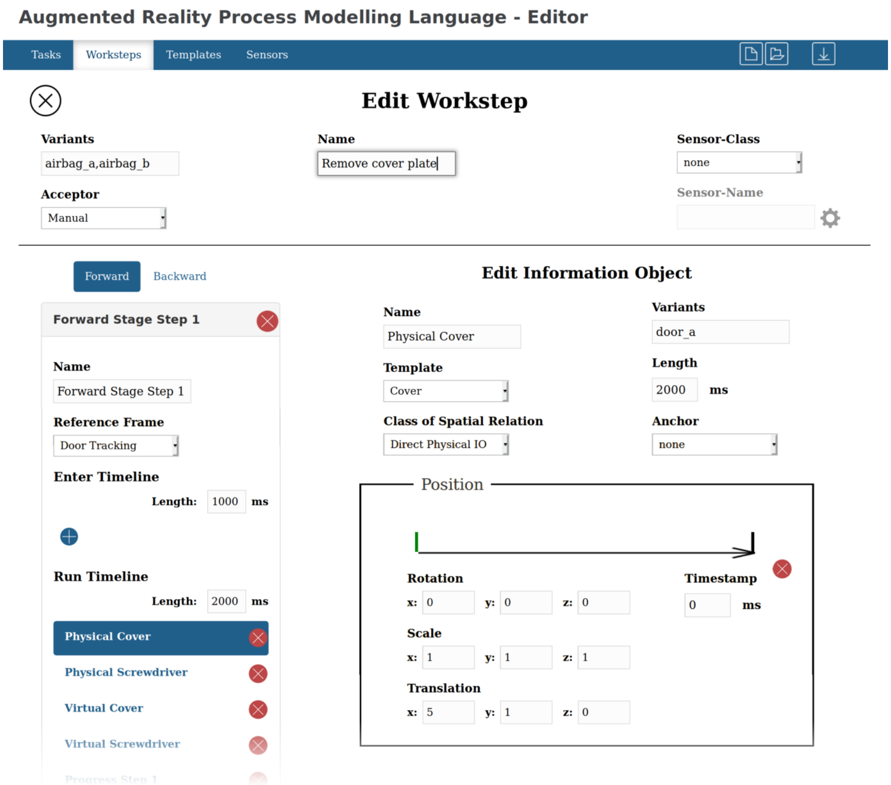


Figure 7.9: Defining the work step to remove the cover plate in the ARPML editor

can be found in figure 7.9. Options concerning data input could also be set but are not included here for the aforementioned reason that this aspect of ARPML is not relevant for information representation.

Aside from defining the single work steps, their dependencies must also be modeled. While some have a fixed order, others do not influence each other. It is obvious, for example, that the cover clips must be removed before the screws underneath can be unscrewed but it does not matter whether one clip or the other is removed first. These sequences are modeled by the dependency list of each work step. For all work steps from the example task, the dependencies are shown in figure 7.10.

7.7 Task Models in ARPML

While the individual work steps and their dependencies are already covered by the already introduced components of ARPML, there are still some aspects missing to author complete repair and maintenance procedures. For example, work steps only model technical dependencies while a further structuring of the task does not take place. In addition, conditional selection of subtasks or work steps may be necessary. This is the case, for example, when a fault memory is read out and,

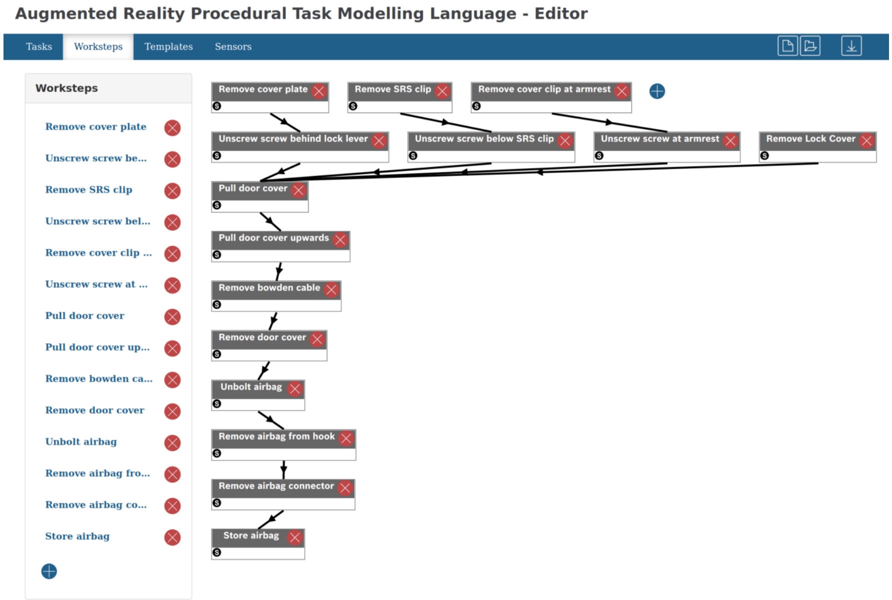


Figure 7.10: All dependencies between the work steps from the airbag removal task as shown in the the ARPML editor. The arrows point from the work step, which must be completed first, to the work step, which can only be started after completing all steps pointing at it.

depending on the result, one of several possible repairs must be carried out, each of which requires different work steps. *Task models* are a means to cover exactly these aspects. Previously, they were already used to model augmented reality support for procedural tasks by De Crescenzio et al. [75] and Wang et al. [298].

Through task models, the overall task is divided into smaller and smaller subtasks. The root as well as the inner nodes represent tasks that are further divided into subtasks while leaf nodes are fine-grained enough to refer to work steps. However, it is not necessary to include a leaf for each individual work step but instead only for selected ones. Through their dependency lists, all required work steps can be automatically resolved, allowing a procedural task to be structured by an author as far as desired without defining a complete sequence of all steps. Task models can also be tagged with variants. However, it is not possible assign variants to single subtasks, as this can quickly lead to a confusing complexity.

Nodes, inner nodes as well as leaf nodes, with the same parent node can be connected by two types of relations. One type is a sequence between two nodes. This means that all work steps referenced by the first node or any of its subnodes need to be completed before the first work step referenced by the second node or any of its subnodes can be started. This relation makes it possible to structure procedural tasks in such a way that a sequence of the work steps does not result solely from the technical dependencies but can be defined by an author. This allows an author to guide through a repair, for example, in a way that makes it less error-prone or easier to perform. The other relation is a case distinction that can be defined between a source node and any number of target nodes. At runtime, after the source node is completed, one of the target nodes is selected, while the other target nodes and their subnodes are discarded. This makes it possible to select the appropriate next steps on the basis of findings during the execution of a task, e.g., based on a diagnosis result. The concrete selection mechanism is, as in the description of the work steps, omitted here because it is rather technical and has no relation to the use of IRMAR.

Compared to the templates, information objects and work steps this part of ARPML are considerably less related to the IRMAR framework. Instead, it is a structure that is needed to make the other parts usable by modeling the overall task. It would also be possible to use other approaches instead, like e.g., state charts as in APRIL [172] or activity diagrams as in DWARF [32]. However, task models were chosen for several reasons. First, they are a possible way to break down tasks to a work step level that supports the authoring process through the given hierarchy. They have the benefit that "they are easy to understand and use" and "are able to structure large sized specifications" (Paterno et al. [219], 1997, page 1). By including sequences and conditions, they are well-suited to model procedures of typical repair and maintenance tasks. At the same time, however, they do not come with the high complexity of modeling approaches that do not define the structure of a procedural task but instead its dynamic flow. Furthermore, the tree structure of task models can help to quickly select the current work step if augmented reality support is only started during an ongoing repair.

A UML definition of task models as used in ARPML can be found in figure 7.11. However, a limitation is the simplified integration of the case discrimination mechanism, as it is not relevant here.

7.7.1 Example Application of Task Models

The task model is created as the final part of the example airbag removal task. Its root is the overall task, which is named accordingly (*Airbag Removal*). This is divided into two subtasks, the removal of the door cover (*Remove Door Cover*) and the actual removal of the airbag (*Remove Airbag*). Between these two subtasks, a sequence is defined which specifies that the inner cover must be removed before the airbag is removed. The sequence of steps for the airbag removal after the cover has been taken off, is clearly defined by the dependencies of the last step *Store Airbag* (see figure 7.10). Thus, it is sufficient to only reference this one in the task model. However, there are several possible sequences for removing the door cover. Thus, a further breakdown into four subtasks is defined for *Remove Door Cover* of which each refers to one work step. Further, sequences are defined by which the individual subtasks have to be carried out. Interestingly, none of the four subtasks and none of the sequences are technically necessary. It would be possible to generate a valid order of work steps based on their dependencies. Instead, the task model reflects the opinion of the technical author on the best

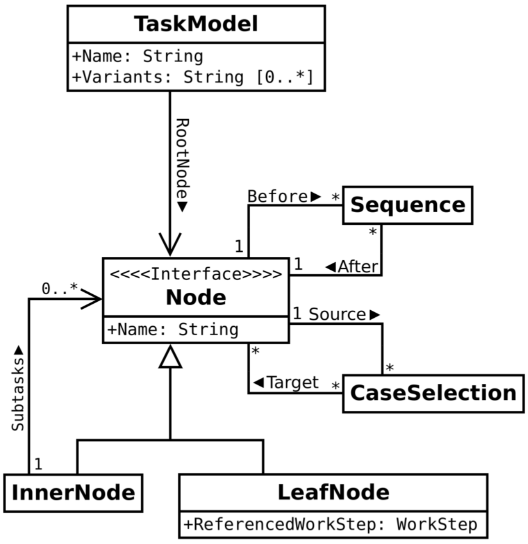


Figure 7.11: UML definition of task models in ARPML. The case discrimination has been simplified, since this mechanism is not relevant here.

sequence of the individual work steps for completing the task. In this case, a working direction from the left towards the right side of the door was chosen. A graphical representation of the task model from the ARPML editor can be found in figure 7.12.

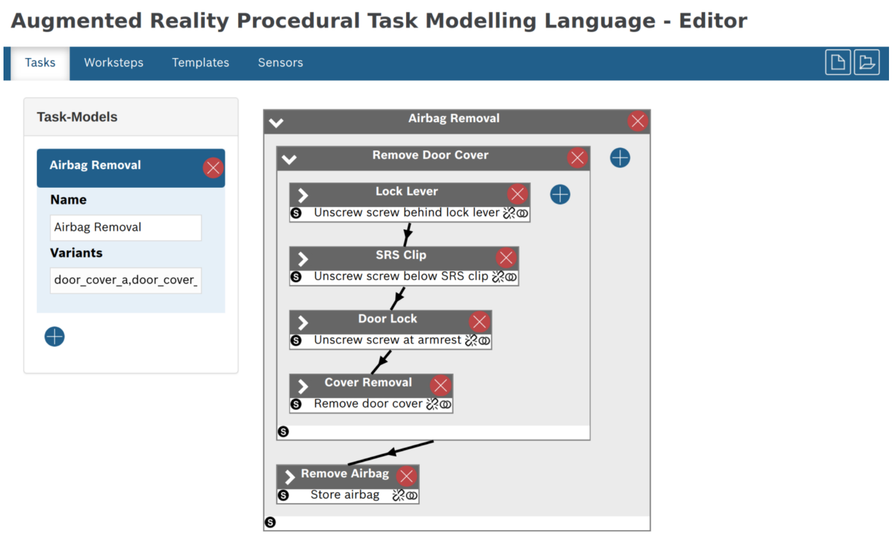


Figure 7.12: Task model for the airbag removal task in the ARPML editor. Subtasks are shown as nested inside their parent tasks. The arrows show the order of execution, with the arrow going from the preceding to the following subtask.

8 Conclusion and Outlook

At the beginning of this thesis, three research goals were formulated, which are pursued with this work. The primary goal is to identify sources of unnecessary sensory and cognitive effort for users of augmented reality systems that represent information to support procedural tasks. Part of this goal is also to formulate these sources as challenges that need to be solved and to find and develop measures that can be used in solutions for them. The two other goals are derived from the primary one. The first is to find a suitable approach for modeling information representation with augmented reality for procedural task support that facilitates the identification and description of the challenges. The other one is to show the application of the identified representation challenges and the measures available to solve them for the design of augmented reality systems. In the following, it is assessed how well these goals have been achieved with this work and which aspects have not yet been fully investigated. Furthermore, it is also described which scientific contributions have been achieved by working on these goals, as well as which further steps can be taken based on the accomplished results.

8.1 Achievement of Goals

The goal pursued first was to create an approach for modeling information represented with augmented reality. For this purpose, the IRMAR framework was developed (**chapter 3: Information Representation Modeling in Augmented Reality**). Its core concept are the information objects, which are individual objects, either physical or virtual ones, that provide relevant information for users of an augmented reality system that supports executing a procedural task. Particularly relevant is that not only virtual objects are included in this modeling framework but also physical ones. This takes into account the nature of augmented reality that it combines physical and virtual objects into a combined three-dimensional space. For the same reason, special attention is paid to the relationships of objects to the physical world and to each other. Because of this, information objects are distinguished by five classes of relation to the physical world, ranging from physically integrated to completely detached. In order to be able to model their spatial arrangement, the position and orientation is defined by their mounting. Their connections with each other that go beyond their spatial positioning are included through the concept of anchors. However, their informational content as well as their visualization are left out intendentionaly. Through this description of information objects, IRMAR particularly contributes to the first goal. By including the relations of information objects with the physical world and each other, it is possible to model a wide variety of relations specific to augmented reality. By leaving informational content and visualization out, a generalizability beyond individual applications is ensured. Through this focus on relationships and a sufficient generality, it facilitates identification of characteristics from which sources of unnecessary sensory and cognitive effort can be derived. This concept differs from previous approaches in three ways relevant for achieving the first goal. The first is the combination of physical and virtual objects in one modeling approach so that they can be examined together in a unified way. This is necessary for the examination of information representation with augmented reality, since an observer perceives both kinds of objects together. The second is to classify the objects according to their connection to the physical world. This is necessary because it is a basis for relevant characteristics, which are later used to derive the challenges. Finally, the third is the determination of characteristics that are generally valid for information represented with augmented reality to support procedural tasks. These are the foundation for the following derivation of the challenges. With these results, the first goal can be considered as achieved. Nevertheless, there are some aspects that have not been fully investigated yet. One important point is that so far, it is not clear which characteristics are relevant for the derivation of representation challenges and which are not. An assessment is only possible in retrospect if a characteristic has contributed to a challenge. At the same time, it can be assumed that there are further characteristics which are relevant for further representation challenges. What these are and how they can be found is still open.

The goal pursued second, which is also the overall goal, was to identify sources of sensory and cognitive effort, formulate them as challenges and identify measures to solve them. It was approached by combining the previously identified characteristics of information objects and their combination, whereby sources of unnecessary sensory and cognitive effort could be derived from some of these combinations (**chapter 4: Representation Challenges**). Through this approach, it was shown that the challenges are an inherent part of the use of augmented reality and are not simply caused by an insufficient user interface design. Overall, six possible sources were identified which were each formulated as a representation challenge that needs to be solved. The identified challenges are the clarity challenge, the consistency challenge, the visibility challenge, the orientation challenge, the motion cue challenge and the information linking challenge. In order to have a comparison with the state of the art and to show which challenges were novel, it is also included where these challenges were previously described. For each of the challenges, measures are collected from the literature that can serve as a (partial) solution in an augmented reality system. Here, the definition of information objects, including their properties and characteristics helped to formulate and organize them. For the information linking challenge, measures were not only taken from literature, but new ones were identified by creating a taxonomy of the different ways to present a connection between information objects and two subsequent empirical studies (**chapter 5: Information Linking**). Specifically, different forms of visual connection between information objects and their anchors were compared in experiments with human subjects regarding the sensory and cognitive effort caused by them. It was also addressed that there are cases where challenges cannot be fully resolved at the same time because the measures required to solve the challenges may be mutually contradictory. Instead, it was argued that a trade-off must be found regarding the degree with which each challenge is solved. The chosen approach is extensible and allows the addition of both new challenges and further solutions to the challenges, which was important since the collection of challenges is probably not complete. Based on these results, the second goal can also be regarded as achieved. However, as with the first goal, there are some aspects that have not been fully investigated yet. For example, it is not yet clear how characteristics have to be combined to create a challenge. Instead, many possible combinations have not been included in this work since they do not seem to form a source of sensory and cognitive effort. Therefore, it is not evident in which direction new characteristics must be investigated to determine further possible sources. The lists with measures also contain large gaps in some parts. This can be seen, for example, in the motion cue challenge. There are no indications under which conditions dynamic or static cues are better suited or how they perform compared to each other.

The goal pursued third was to show the practical use of the identified challenges including the developed modeling approach for developing augmented reality systems. This was pursued in two ways. First, the IRMAR framework was integrated into the ISO norm 9241-210 Human Centred Design (**chapter 6: Using IRMAR in Human-Centred Design**). By thus, a method of using the IRMAR framework to design the representation of information in augmented reality in the context of an established standard was demonstrated. Integration was achieved by creating two lists, one containing potential information objects and one containing constraints, which are filled in the individual activities that are defined in the standard. The use of this approach was explained in detail through an exemplary application. Secondly, the augmented reality procedural task modeling language (ARPML) was developed for authoring augmented reality support for procedural tasks (**chapter 7: The Augmented Reality Procedural Task Modeling Language**). It was created based on the IRMAR framework using information objects as a way of describing the information displayed with augmented reality. This demonstrated another way of using IRMAR besides the maybe more obvious design of user interfaces and is an indication of the framework's versatility. With these two applications, the third goal was achieved as well. However, as with the two previous goals, there are aspects not yet covered. In particular, empirical studies or expert interviews are missing to better prove the usefulness of the IRMAR framework. While the basic applicability has been shown with this work, its benefits would probably become more evident with a more extensive evaluation.

8.2 Scientific Contributions

By pursuing the previously discussed goals, this work contributes to research in the field of augmented reality user interfaces. Three contributions are considered by the author to be the most relevant ones. Their relevance stems from the fact that these results significantly deepen the understanding of augmented reality and its use for supporting procedural tasks. These are the three main contributions:

- The concept of information objects for modeling information represented in augmented reality. Of particular relevance is the combination of physical and virtual objects in one approach that also includes the spatial connection of these objects with the physical world. With this, a comprehensive model has been defined that makes information from physical and virtual sources in augmented reality available to a unified examination. Further, the investigation of their characteristics as part of this modeling approach provides relevant insights into the structures of augmented reality user interfaces for procedural task support.
- The approach to derive representation challenges from the identified characteristics. As a result, these challenges are clearly defined and it is established why they are inherent to the use of augmented reality for supporting procedural tasks. By creating a collection of measures for each challenge, a foundation for solving them is provided. As part of the approach, it is also explained why in many situations perfect solutions are not possible due to different challenges requiring contradictory measures.
- The taxonomy created for information linking in augmented reality and the empirical research based on it. It allows describing the different possibilities for displaying information objects as connected to each other and thus defines the corresponding design space. The empirical studies provide insight into which type of visual connection from this design space is preferable to present connections.

Even though a relatively large part of this work is given to the applications of IRMAR, they are not among of the most relevant contributions. The reason is that in the end they show the use of the achieved contributions, but do not provide independent insights on their own.

8.3 Future Work

Based on this work, different directions can be taken for future research. One obvious option is to extend and refine the measures for the challenges and make it easier for application designers to find adequate solutions for information representation. As already described, there are still gaps. For example, the visibility challenge lacks a technique to move different spatial virtual information objects that overlap in such a way that they are fully visible while adequately preserving their spatial contexts. Another example is the visual integration aspect of the consistency challenge. Different factors are known, but how they interact with each other or to what extent integration is really necessary is unknown. A third example are the already mentioned static and dynamic motion cues from the motion cue challenge. It is not known which of these should be used under what circumstances and whether there is a general preference for one of them. Further, the concrete solutions to the challenges must be found individually for each application without having transferable templates that can be applied in many situations. Design patterns offer such solutions, because they provide transferable templates, which indicate their area of application and also define which influences, called forces, have led to it. Regarding the IRMAR framework, the challenges can be used as forces. Finding such patterns would greatly simplify designing suitable information representation, especially in cases when measures to solve challenges contradict each other.

Another direction is the identification of further challenges. So far, six have been identified, but not every combination of characteristics of information objects and their combination have been considered yet, therefore, there will probably be additional challenges. It is also likely that not all relevant characteristics have been identified yet, which means that it is realistic to be able to define further challenges. Further, the question of what exactly makes a combination of characteristics a challenge remains unanswered. A deeper understanding of this could provide clues as to where and how further challenges can be found.

Until now, the IRMAR framework has only been applied to supporting procedural tasks. However, there are also other application areas, e.g., assisting learning processes with augmented reality or giving drivers information through augmented reality head-up displays. Maybe information objects can also be defined for these areas, but if it is possible, their characteristics will certainly be different. This will lead to at least partly different challenges, which will also require other measures. Whether minimizing sensory and cognitive effort remains a goal or if other goals are needed is also open.

Bibliography

[1] Arkit | apple developer documentation. https://developer.apple.com/documentation/arkit. Accessed: 2022-01-22.

[2] Artoolkit github repository. https://github.com/artoolkit. Accessed: 2022-01-22.

[3] Artoolkit home page. http://www.hitl.washington.edu/artoolkit/. Accessed: 2022-01-22.

[4] Bosch augmented reality applications now also work with the new microsoft hololens 2. https://www.bosch-presse.de/pressportal/de/en/bosch-augmented-reality-applications-now-also-work-with-the-new-microsoft-hololens-2-183680.html. Accessed: 2022-01-22.

[5] Dk-52 - lumus optical. https://lumusvision.com/wp-content/uploads/2018/01/DK-52-Flyer.pdf. Accessed: 2022-01-22.

[6] Extend3d products: Werklicht pro and werklicht video mevis. https://www.extend3d.de/en/produkte.php. Accessed: 2022-01-22.

[7] Google ar & vr | arcore. https://arvr.google.com/arcore/. Accessed: 2022-01-22.

[8] Lumus dk-32 brochure. http://wiki.lumus-optical.com/lib/exe/fetch.php?media=user_manual:brochure_a4_lumusdk-32_oe-32_09.14.pdf. Accessed: 2019-02-24.

[9] Microsoft hololens | mixed reality technology for business. https://www.microsoft.com/en-us/hololens. Accessed: 2022-01-22.

[10] Moverio bt-350 - epson. https://download.epson-europe.com/pub/download/6342/epson634266eu.pdf. Accessed: 2022-01-22.

[11] Solutions for the automotive industry. https://assets.bosch.com/media/global/products_and_solutions/market_specific_solutions/solutions_for_the_automotive_industry/brochure.pdf. Accessed: 2022-01-22.

[12] Surface pro 3 specifications. https://support.microsoft.com/de-de/help/4037669/surface-surface-pro-3-specifications. Accessed: 2022-01-22.

[13] Usb 2 ueye xs industrial camera. `https://en.ids-imaging.com/store/xs.html`. Accessed: 2022-01-22.

[14] Visionlib. `https://visionlib.com/`. Accessed: 2022-01-22.

[15] What is the mixed reality toolkit | mixed reality toolkit documentation. `https://microsoft.github.io/MixedRealityToolkit-Unity/README.html`. Accessed: 2022-01-22.

[16] Maneesh Agrawala, Doantam Phan, Julie Heiser, John Haymaker, Jeff Klingner, Pat Hanrahan, and Barbara Tversky. Designing effective step-by-step assembly instructions. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 828–837. ACM, 2003.

[17] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517. Ieee, 2012.

[18] Takahisa Ando, Koji Yamasaki, Masaaki Okamoto, and Eiji Shimizu. Head-mounted display using a holographic optical element. In *Practical Holography XII*, volume 3293, pages 183–190. International Society for Optics and Photonics, 1998.

[19] Benjamin Avery, Christian Sandor, and Bruce H Thomas. Improving spatial perception for augmented reality x-ray vision. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 79–82. IEEE, 2009.

[20] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, Nov./Dec. 2001.

[21] Ronald Azuma and Chris Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of the 2nd IEEE/ACM international Symposium on Mixed and Augmented Reality*, page 66. IEEE Computer Society, 2003.

[22] Ronald Azuma, Howard Neely, Mike Daily, and Jon Leonard. Performance analysis of an outdoor augmented reality tracking system that relies upon a few mobile beacons. In *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on*, pages 101–104. IEEE, 2006.

[23] Ronald T Azuma. A survey of augmented reality. *Presence: Teleoperators and virtual environments*, 6(4):355–385, 1997.

[24] Ronald T Azuma. The challenge of making augmented reality work outdoors. *Mixed reality: Merging real and virtual worlds*, pages 379–390, 1999.

[25] Ronald T Azuma, Bruce R Hoff, Howard E Neely III, Ronald Sarfaty, Michael J Daily, Gary Bishop, Vern Chi, Greg Welch, Ulrich Neumann, Suya You, et al. Making augmented reality work outdoors requires hybrid tracking. In *Proceedings of the First International Workshop on Augmented Reality*, volume 1, 1998.

[26] Cédric Bach and Dominique Scapin. Ergonomic criteria adapted to human virtual environment interaction. In *Proceedings of the 15th Conference on l'Interaction Homme-Machine*, pages 24–31. ACM, 2003.

[27] Kevin M Baird and Woodrow Barfield. Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality*, 4(4):250–259, 1999.

[28] Stefano Baldassi, Grace T Cheng, Jonathan Chan, Moqian Tian, Tim Christie, and Matthew T Short. Exploring immersive ar instructions for procedural tasks: The role of depth, motion, and volumetric representations. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*, pages 300–305. IEEE, 2016.

[29] Ryan Bane and Tobias Höllerer. Interactive tools for virtual x-ray vision in mobile augmented reality. In *Proceedings of the 3rd IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 231–239. IEEE, 2004.

[30] Thomas Bär, Ulrich Bockholt, Hilko Hoffmann, Eduard Jundt, Matthias Roth, Werner Schreiber, Ingo Staack, Peter Zimmermann, and Konrad Zürl. Das verbundprojekt arvida. In *Web-basierte Anwendungen Virtueller Techniken*, pages 1–16. Springer, 2017.

[31] Olivier Bau and Ivan Poupyrev. Revel: tactile feedback technology for augmented reality. *ACM Transactions on Graphics (TOG)*, 31(4):89, 2012.

[32] Martin Bauer, Bernd Bruegge, Gudrun Klinker, Asa MacWilliams, Thomas Reicher, Stefan Riss, Christian Sandor, and Martin Wagner. Design of a component-based augmented reality framework. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 45–54. IEEE, 2001.

[33] James Baumeister, Seung Youb Ssin, Neven AM ElSayed, Jillian Dorrian, David P Webb, James A Walsh, Timothy M Simon, Andrew Irlitti, Ross T Smith, Mark Kohler, et al. Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics*, 23(11):2378–2388, 2017.

[34] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[35] Benjamin B Bederson. Audio augmented reality: a prototype automated tour guide. In *Conference companion on Human factors in computing systems*, pages 210–211. ACM, 1995.

[36] Blaine Bell, Steven Feiner, and Tobias Höllerer. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 101–110. ACM, 2001.

[37] Blaine Bell, Tobias Höllerer, and Steven Feiner. An annotated situation-awareness aid for augmented reality. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 213–216. ACM, 2002.

[38] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)*, 5(3):185–223, 1998.

[39] Christoph Bichlmeier and Nassir Navab. Virtual window for improved depth perception in medical ar. In *International Workshop on Augmented Reality environments for Medical Imaging and Computer-aided Surgery (AMI-ARCS)*, 2006.

[40] Irving Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

[41] M. Biere, Birgit Bomsdorf, and Gerd Szwillus. The visual task model builder. In Jean Vanderdonckt and Angel Puerta, editors, *Computer-Aided Design of User Interfaces II*, pages 245–256. Springer Netherlands, 1999.

[42] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015.

[43] Mark Billinghurst, Raphael Grasset, and Julian Looser. Designing augmented reality interfaces. *ACM Siggraph Computer Graphics*, 39(1):17–22, 2005.

[44] Mark Billinghurst, Ivan Poupyrev, Hirokazu Kato, and Richard May. Mixing realities in shared space: An augmented reality interface for collaborative computing. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World*

of Multimedia (Cat. No. 00TH8532), volume 3, pages 1641–1644. IEEE, 2000.

[45] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds*. AK Peters/CRC Press, 2005.

[46] Oliver Bimber and Ramesh Raskar. Modern approaches to augmented reality. In *ACM SIGGRAPH 2006 Courses*, page 1. ACM, 2006.

[47] Frank Biocca, Charles Owen, Arthur Tang, and Corey Bohil. Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality. *Journal of Management Information Systems*, 23(4):163–184, 2007.

[48] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1115–1122. ACM, 2006.

[49] Jan O Borchers. A pattern approach to interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 369–378. ACM, 2000.

[50] Felix Bork, Christian Schnelzer, Ulrich Eck, and Nassir Navab. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE transactions on visualization and computer graphics*, 2018.

[51] Andrew C Boud, David J Haniff, Chris Baber, and SJ Steiner. Virtual reality and augmented reality as a training tool for assembly tasks. In *iv*, page 32. IEEE, 1999.

[52] Doug A Bowman, Larry F Hodges, Don Allison, and Jean Wineman. The educational value of an information-rich virtual environment. *Presence: Teleoperators & Virtual Environments*, 8(3):317–331, 1999.

[53] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators and virtual environments*, 10(1):96–108, 2001.

[54] Doug A Bowman, Chris North, Jian Chen, Nicholas F Polys, Pardha S Pyla, and Umur Yilmaz. Information-rich virtual environments: theory, tools, and research agenda. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 81–90. ACM, 2003.

[55] Andreas Brennecke and Reinhard Keil-Slawik. Notes on the alltagspraxis of hypermedia design. In *Educational Multimedia and Hypermedia Annual (Proc. of ED-MEDIA 95, Graz)*, pages 115–120, 1995.

[56] Ricardo Buettner. Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. In *Annual Conference on Artificial Intelligence*, pages 37–48. Springer, 2013.

[57] Sebastian Büttner, Markus Funk, Oliver Sand, and Carsten Röcker. Using head-mounted displays and in-situ projection for assistive systems: a comparison. In *Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments*, page 44. ACM, 2016.

[58] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. The design space of augmented and virtual reality applications for assistive environments in manufacturing: a visual approach. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 433–440. ACM, 2017.

[59] Brad Cain. A review of the mental workload literature. Technical report, Defence Research And Development Toronto (Canada), 2007.

[60] Ozan Cakmakci and Jannick Rolland. Head-worn displays: a review. *Journal of display technology*, 2(3):199–216, 2006.

[61] Mackinlay Card. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[62] Thomas P Caudell and David W Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, volume 2, pages 659–669. IEEE, 1992.

[63] Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, and Pan Hui. Mobile augmented reality survey: From where we are to where we go. *Ieee Access*, 5:6917–6950, 2017.

[64] Jian Chen, Pardha S Pyla, and Doug A Bowman. Testbed evaluation of navigation and text display techniques in an information-rich virtual environment. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 181–289. IEEE, 2004.

[65] Lawrence Chung, Brian A Nixon, Eric Yu, and John Mylopoulos. *Non-functional requirements in software engineering*, volume 5. Springer Science & Business Media, 2012.

[66] Andrew I Comport, Éric Marchand, and François Chaumette. A real-time tracker for markerless augmented reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 36. IEEE Computer Society, 2003.

[67] Lynn A Cooper. Mental rotation of random two-dimensional shapes. *Cognitive psychology*, 7(1):20–43, 1975.

[68] Sébastien Cuendet, Quentin Bonnard, Son Do-Lenh, and Pierre Dillenbourg. Designing augmented reality for the classroom. *Computers & Education*, 68:557–569, 2013.

[69] James E Cutting. Reconceiving perceptual space. *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space*, pages 215–238, 2003.

[70] James E Cutting and Peter M Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *Perception of Space and Motion*, pages 69–117, 1994.

[71] Rudolph P Darken, Joseph A Sullivan, and Mark Lennerton. A chromakey augmented virtual environment for deployable training. 2003.

[72] Ralf Dauenhauer. Nutzerbezogene untersuchungen zur visuellen informationsverknüpfung in augmented-reality-anwendungen, 8 2016.

[73] Ralf Dauenhauer and Tobias Müller. An evaluation of information connection in augmented reality for 3d scenes with occlusion. In *Proceedings of the 15th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 235–237. IEEE, 2016.

[74] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):1052–1067, 2007.

[75] Francesca De Crescenzio, Massimiliano Fantini, Franco Persiani, Luigi Di Stefano, Pietro Azzari, and Samuele Salti. Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, 31(1):96–101, 2011.

[76] José de Souza's. An illustrated review of how motion is represented in static instructional graphics. In *Proc. STC's 55th Conf*, pages 1–4. Citeseer, 2008.

[77] Dick De Waard. *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands, 1996.

[78] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE transactions on visualization and computer graphics*, 20(1):125–139, 2014.

[79] Catherine Diaz, Michael Walker, Danielle Albers Szafir, and Daniel Szafir. Designing for depth perceptions in augmented reality. In *2017 IEEE international symposium on mixed and augmented reality (ISMAR)*, pages 111–122. IEEE, 2017.

[80] ISO DIS. 9241-210: 2010. ergonomics of human system interaction - part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland*, 2009.

[81] ISO DIS. 9241-210: 2018. ergonomics of human-system interaction - part 125: Guidance on visual presentation of information. *International Standardization Organization (ISO). Switzerland*, 2018.

[82] ISO DIS. 9241-110: 2020. ergonomics of human-system interaction - part 110: Dialogue principles. *International Standardization Organization (ISO). Switzerland*, 2020.

[83] Stephen DiVerdi, Tobias Höllerer, and Richard Schreyer. Level of detail interfaces. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 300–301. IEEE Computer Society, 2004.

[84] Klaus Dorfmüller. Robust tracking for augmented reality using retroreflective markers. *Computers & Graphics*, 23(6):795–800, 1999.

[85] Ashish Doshi, Ross T Smith, Bruce H Thomas, and Con Bouras. Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *The International Journal of Advanced Manufacturing Technology*, 89(5-8):1279–1293, 2017.

[86] David Drascic and Paul Milgram. Perceptual issues in augmented reality. In *SPIE: Stereoscopic Displays and Virtual Reality Systems III*, volume 2653, pages 123–134, 1996.

[87] Matt Dunleavy. Design principles for augmented reality learning. *TechTrends*, 58(1):28–34, 2014.

[88] Matt Dunleavy and Chris Dede. Augmented reality teaching and learning. In *Handbook of research on educational communications and technology*, pages 735–745. Springer, 2014.

[89] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. Applying hci principles to ar systems design. pages 37–42. University of Canterbury. Human Interface Technology Laboratory., 2007.

[90] Elsa Eiriksdottir and Richard Catrambone. Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to

enhance performance, learning, and transfer. *Human factors*, 53(6):749–770, 2011.

[91] Stephen R Ellis, Francois Breant, B Manges, Richard Jacoby, and Bernard D Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, pages 138–145. IEEE, 1997.

[92] Stephen R Ellis, Anthony Wolfram, and Bernard D Adelstein. Three dimensional tracking in augmented environments: user performance trade-offs between system latency and update rate. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 46, pages 2149–2153. SAGE Publications Sage CA: Los Angeles, CA, 2002.

[93] Niklas Elmqvist, Ulf Assarsson, and Philippas Tsigas. Employing dynamic transparency for 3d occlusion management: Design issues and evaluation. *Human-Computer Interaction–INTERACT 2007*, pages 532–545, 2007.

[94] José Luis Encarnação and Didier Stricker. Augmented-reality für industrielle anwendungen in entwicklung, produktion und service am beispiel des leitprojekts arvika. In *Informatikforschung in Deutschland*, pages 269–326. Springer, 2008.

[95] Timo Engelke, Sabine Webel, and Nirit Gavish. Generating vision based lego augmented reality training and evaluation systems. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 223–224. IEEE, 2010.

[96] Manfred Fahle and Tom Troscianko. Computation of texture and stereoscopic depth in humans. Technical report, MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 1989.

[97] Mark D Fairchild. *Color appearance models*, chapter "Color Appearance Phenomena". John Wiley & Sons, 2013.

[98] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 145–155. ACM, 1993.

[99] Steven Feiner, Blair MacIntyre, and Tobias Höllerer. Wearing it out: First steps toward mobile augmented reality systems. In *First International Symposium on Mixed Reality (ISMR'99)*, 1999.

[100] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies*, 1(4):208–217, 1997.

[101] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. In *Wearable Computers, 1997. Digest of Papers., First International Symposium on*, pages 74–81. IEEE, 1997.

[102] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62, 1993.

[103] Yan Feng. Estimation of light source environment for illumination consistency of augmented reality. In *Image and Signal Processing, 2008. CISP'08. Congress on*, volume 3, pages 771–775. IEEE, 2008.

[104] Michele Fiorentino, Saverio Debernardis, Antonio E Uva, and Giuseppe Monno. Augmented reality text style readability with see-through head-mounted displays in industrial context. *Presence: Teleoperators and Virtual Environments*, 22(2):171–190, 2013.

[105] Jan Fischer. Visual consistency in augmented reality compositing. In *Fundamentals of Wearable Computers and Augmented Reality*, pages 474–501. CRC Press, 2015.

[106] Jan Fischer, Dirk Bartz, and Wolfgang Straber. Stylized augmented reality for improved immersion. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, pages 195–202. IEEE, 2005.

[107] Jan Fischer, Michael Haller, and Bruce H. Thomas. Stylized depiction in mixed reality. *The International Journal of Virtual Reality*, 7(4):71–79, 2008.

[108] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[109] Pierre Fite-Georgel. Is there a reality in industrial augmented reality? In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 201–210. IEEE, 2011.

[110] David R Flatla and Carl Gutwin. Individual models of color differentiation to improve interpretability of information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2563–2572. ACM, 2010.

[111] James D Foley, Foley Dan Van, Andries Van Dam, Steven K Feiner, John F Hughes, Edward Angel, and J Hughes. *Computer graphics: principles and practice*. Addison-Wesley Professional, 3 edition, 2013.

[112] Eric Foxlin, Yury Altshuler, Leonid Naimark, and Mike Harrington. Flight-tracker: A novel optical/inertial tracker for cockpit enhanced vision. In

Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, pages 212–221. IEEE Computer Society, 2004.

[113] Matthew Franklin. The lessons learned in the application of augmented reality. Technical report, QINETIQ LTD FARNBOROUGH (UNITED KINGDOM), 2006.

[114] Tycho K Fredericks, Sang D Choi, Jason Hart, Steven E Butt, and Anil Mital. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics*, 35(12):1097–1107, 2005.

[115] Wolfgang Friedrich, D Jahn, and L Schmidt. Arvika-augmented reality for development, production and service. In *ISMAR*, volume 2002, pages 3–4. Citeseer, 2002.

[116] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. Working with augmented reality?: a long-term analysis of in-situ instructions at the assembly workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 222–229. ACM, 2017.

[117] Markus Funk, Thomas Kosch, and Albrecht Schmidt. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 934–939. ACM, 2016.

[118] Chris Furmanski, Ronald Azuma, and Mike Daily. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Proceedings of the first International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 215–320. IEEE, 2002.

[119] Joseph L Gabbard, J Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In *Virtual Reality Conference, 2007. VR'07. IEEE*, pages 35–42. IEEE, 2007.

[120] Joseph L Gabbard, J Edward Swan, Deborah Hix, Robert S Schulman, John Lucas, and Divya Gupta. An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, pages 11–18. IEEE, 2005.

[121] Joseph L Gabbard, J Edward Swan, Jason Zedlitz, and Woodrow W Winchester. More than meets the eye: An engineering study to empirically examine

the blending of real and virtual color spaces. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 79–86. IEEE, 2010.

[122] Joseph L Gabbard, J Edward II Swan, and Deborah Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoperators and Virtual Environments*, 15(1):16–32, 2006.

[123] TB Garland and Christopher A Sanchez. Rotational perspective and learning procedural tasks from dynamic media. *Computers & Education*, 69:31–37, 2013.

[124] Nirit Gavish, Teresa Gutierrez, Sabine Webel, Jorge Rodriguez, and Franco Tecchia. Design guidelines for the development of virtual reality and augmented reality training systems for maintenance and assembly tasks. In *Proceedings of the The International Conference of the SKILLS European Project*, volume 1, page 29, 2011.

[125] Elizabeth Goodman, Mike Kuniavsky, and Andrea Moed. *Observing the user experience: A practitioner's guide to user research.* Elsevier, 2012.

[126] Michael M Granaas, Timothy D McKay, R Darrell Laham, Lance D Hurt, and James F Juola. Reading moving text on a crt screen. *Human Factors*, 26(1):97–104, 1984.

[127] Raphael Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. Image-driven view management for augmented reality browsers. In *Proceedings of the 11th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 177–186. IEEE, 2012.

[128] Jens Grubert, Yuta Itoh, Kenneth Moser, and J Edward Swan. A survey of calibration methods for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 24(9):2649–2662, 2018.

[129] Uwe Gruenefeld, Dag Ennenga, Abdallah El Ali, Wilko Heuten, and Susanne Boll. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 109–118. ACM, 2017.

[130] Mika Hakkarainen, Charles Woodward, and Mark Billinghurst. Augmented assembly using a mobile phone. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 167–168. IEEE Computer Society, 2008.

[131] Michael Haller. Photorealism or/and non-photorealism in augmented reality. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry*, pages 189–196. ACM, 2004.

[132] David J Haniff and Chris Baber. User evaluation of augmented reality systems. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 505–511. IEEE, 2003.

[133] Aki Harma, Julia Jakka, Miikka Tikander, Matti Karjalainen, Tapio Lokki, and Heli Nironen. Techniques and applications of wearable augmented reality audio. In *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[134] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[135] Steven Henderson and Steven Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE transactions on visualization and computer graphics*, 17(10):1355–1368, 2011.

[136] Steven J Henderson and Steven Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 135–144. IEEE, 2009.

[137] Steven J Henderson and Steven K Feiner. Augmented reality in the psychomotor phase of a procedural task. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 191–200. IEEE, 2011.

[138] Gentaro Hirota, David T Chen, William F Garrett, Mark A Livingston, et al. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 429–438. ACM, 1996.

[139] Ira J Hirsh and Carl E Sherrick Jr. Perceived order in different sense modalities. *Journal of experimental psychology*, 62(5):423, 1961.

[140] Tim N Höffler and Detlev Leutner. Instructional animation versus static pictures: A meta-analysis. *Learning and instruction*, 17(6):722–738, 2007.

[141] Richard L Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):413–432, 1997.

[142] Lei Hou and Xiangyu Wang. Using augmented reality to cognitively facilitate product assembly process. In *Augmented Reality*. IntechOpen, 2010.

[143] Lei Hou, Xiangyu Wang, Leonhard Bernold, and Peter ED Love. Using animated augmented reality to cognitively guide assembly. *Journal of Computing in Civil Engineering*, 27(5):439–451, 2013.

[144] Lei Hou, Xiangyu Wang, and Martijn Truijens. Using augmented reality to facilitate piping assembly: an experiment-based evaluation. *Journal of Computing in Civil Engineering*, 29(1):05014007, 2013.

[145] Hong Hua, Prasanna Krishnaswamy, and Jannick P Rolland. Video-based eyetracking methods and algorithms in head-mounted displays. *Optics Express*, 14(10):4328–4350, 2006.

[146] Charles E Hughes, Christopher B Stapleton, and Matthew R O'Connor. The evolution of a framework for mixed reality experiences. In *Emerging Technologies of Augmented Reality: Interfaces and Design*, pages 198–217. IGI Global, 2007.

[147] Yuta Itoh, Maksym Dzitsiuk, Toshiyuki Amano, and Gudrun Klinker. Semi-parametric color reproduction method for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 21(11):1269–1278, 2015.

[148] Yuta Itoh and Gudrun Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *2014 IEEE symposium on 3d user interfaces (3dui)*, pages 75–82. IEEE, 2014.

[149] Katrien Jacobs and Céline Loscos. Classification of illumination methods for mixed reality. In *Computer Graphics Forum*, volume 25, pages 29–51. Wiley Online Library, 2006.

[150] Christian Jerome and Bob Witmer. The perception and estimation of egocentric distance in real and augmented reality environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 2249–2252. SAGE Publications Sage CA: Los Angeles, CA, 2005.

[151] Laura G Johnson, Philip Edwards, and David Hawkes. Surface transparency makes stereo overlays unpredictable: the implications for augmented reality. *Studies in health technology and informatics*, pages 131–136, 2003.

[152] J Adam Jones, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 9–14. ACM, 2008.

[153] Simon Julier, Marco Lanzagorta, Yohan Baillot, Lawrence Rosenblum, Steven Feiner, Tobias Höllerer, and Sabrina Sestito. Information filtering for mobile augmented reality. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 3–11. IEEE, 2000.

[154] Svenja Kahn and Arjan Kuijper. Fusing real-time depth imaging with high precision pose estimation by a measurement arm. In *2012 International Conference on Cyberworlds*, pages 256–260. IEEE, 2012.

[155] Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg. Interactive focus and context visualization for augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–10. IEEE Computer Society, 2007.

[156] Ichiro Kasai, Yasushi Tanijiri, Takeshi Endo, and Hiroaki Ueda. A forgettable near eye display. In *Wearable Computers, The Fourth International Symposium on*, pages 115–118. IEEE, 2000.

[157] Reinhard Keil and Christian Schild. Gestaltungskonflikte in der softwareergonomie. In *Mensch & Computer*, pages 67–77, 2013.

[158] Kangsoo Kim, Mark Billinghurst, Gerd Bruder, Henry Been-Lirn Duh, and Gregory F Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE transactions on visualization and computer graphics*, 24(11):2947–2962, 2018.

[159] Gary R King, Wayne Piekarski, and Bruce H Thomas. Arvino-outdoor augmented reality visualisation of viticulture gis data. In *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*, pages 52–55. IEEE, 2005.

[160] Kiyoshi Kiyokawa, Mark Billinghurst, Bruce Campbell, and Eric Woods. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 133. IEEE Computer Society, 2003.

[161] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.

[162] Christian Knöpfle, Jens Weidenhausen, Laurent Chauvigné, and Ingo Stock. Template based authoring for ar based service scenarios. In *Proceedings of the 2005 IEEE Conference on Virtual Reality*, VR '05, pages 249–252, Bonn, Germany, 2005. IEEE Computer Society.

[163] Sang Min Ko, Won Suk Chang, and Yong Gu Ji. Usability principles for augmented reality applications in a smartphone environment. *International journal of human-computer interaction*, 29(8):501–515, 2013.

[164] Paul A Kolers, Robert L Duchnicky, and Dennis C Ferguson. Eye movement measurement of readability of crt displays. *Human Factors*, 23(5):517–527, 1981.

[165] Artesas Konsortium. Artesas–advanced augmented reality technologies for industrial service applications, 2008.

[166] Gerald Kotonya and Ian Sommerville. *Requirements engineering: processes and techniques*. Wiley Publishing, 1998.

[167] Moritz Krell. Evaluating an instrument to measure mental load and mental effort using item response theory. 2015.

[168] Ernst Kruijff, J Edward Swan, and Steven Feiner. Perceptual issues in augmented reality revisited. In *Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 3–12. IEEE, 2010.

[169] David M Krum, Evan A Suma, and Mark Bolas. Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing*, 16(1):17–26, 2012.

[170] Mikko Kytö, Aleksi Mäkinen, Jukka Häkkinen, and Pirkko Oittinen. Improving relative depth judgments in augmented reality with auxiliary augmentations. *ACM Transactions on Applied Perception (TAP)*, 10(1):1–21, 2013.

[171] Peter Lang, Albert Kusej, Axel Pinz, and Georg Brasseur. Inertial tracking for mobile augmented reality. In *Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE*, volume 2, pages 1583–1587. IEEE, 2002.

[172] F. Ledermann and D. Schmalstieg. April: a high-level framework for creating augmented reality presentations. In *Proceedings of IEEE Virtual Reality 2005*, VR '05, pages 187–194, March 2005.

[173] Gun A Lee, Ungyeon Yang, Yongwan Kim, Dongsik Jo, Ki-Hong Kim, Jae Ha Kim, and Jin Sung Choi. Freeze-set-go interaction method for handheld mobile augmented reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pages 143–146. ACM, 2009.

[174] Mirna Lerotic, Adrian J Chung, George Mylonas, and Guang-Zhong Yang. Pq-space based non-photorealistic rendering for augmented reality. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 102–109. Springer, 2007.

[175] Alex Leykin and Mihran Tuceryan. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 224–230. IEEE Computer Society, 2004.

[176] Mark A Livingston. Quantification of visual capabilities using augmented reality displays. In *Proceedings of the 2006 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 3–12. IEEE, 2006.

[177] Mark A Livingston, Zhuming Ai, and Jonathan W Decker. A user study towards understanding stereo perception in head-worn augmented reality displays. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 53–56. IEEE, 2009.

[178] Mark A Livingston, Jane H Barrow, and Ciara M Sibley. Quantification of contrast sensitivity and color perception using head-worn augmented reality displays. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 115–122. IEEE, 2009.

[179] Mark A Livingston, Joseph L Gabbard, J Edward Swan II, Ciara M Sibley, and Jane H Barrow. Basic perception in head-worn augmented reality displays. In *Human factors in augmented reality environments*, pages 35–65. Springer, 2013.

[180] Mark A Livingston, William F Garrett, Gentaro Hirota, Mary C Whitton, Etta D Pisano, Henry Fuchs, et al. Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In *Proceedings of the 23rd annual conference on computer graphics and interactive techniques*, pages 439–446. ACM, 1996.

[181] Mark A Livingston, J Edward Swan II, Joseph L Gabbard, Tobias H Höllerer, Deborah Hix, Simon J Julier, Yohan Baillot, and Dennis Brown. Resolving multiple occluded layers in augmented reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 56. IEEE Computer Society, 2003.

[182] Jack M Loomis, Joshua M Knapp, et al. Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments*, 11:21–46, 2003.

[183] Julian Looser, Raphael Grasset, and Mark Billinghurst. A 3d flexible and tangible magic lens in augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–4. IEEE Computer Society, 2007.

[184] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[185] Weiquan Lu, Been-Lirn Henry Duh, and Steven Feiner. Subtle cueing for visual search in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 161–166. IEEE, 2012.

[186] Weiquan Lu, Henry Been-Lirn Duh, Steven Feiner, and Qi Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE transactions on visualization and computer graphics*, 20(3):404–412, 2014.

[187] Stefan Maass, Markus Jobst, and Juergen Doellner. Depth cue of occlusion information as criterion for the quality of annotation placement in perspective views. In *The European information society*, pages 473–486. Springer, 2007.

[188] Blair MacIntyre and E Machado Coelho. Adapting to dynamic registration errors using level of error (loe) filtering. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 85–88. IEEE, 2000.

[189] Colin M MacLeod. Half a century of research on the stroop effect: an integrative review. *Psychological bulletin*, 109(2):163, 1991.

[190] Asa MacWilliams, Christian Sandor, Martin Wagner, Martin Bauer, Gudrun Klinker, and Bernd Bruegge. Herding sheep: live system for distributed augmented reality. In *Proceedings of the second International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 123–132. IEEE, 2003.

[191] Jacob Boesen Madsen, Markus Tatzqern, Claus B Madsen, Dieter Schmalstieg, and Denis Kalkofen. Temporal coherence strategies for augmented reality labeling. *IEEE transactions on visualization and computer graphics*, 22(4):1415–1423, 2016.

[192] Michael R Marner, Andrew Irlitti, and Bruce H Thomas. Improving procedural task performance with augmented reality annotations. In *Proceedings of the 12th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 39–48. IEEE, 2013.

[193] Erick Mendez, Steven Feiner, and Dieter Schmalstieg. Focus and context in mixed reality by modulating first order salient features. In *International Symposium on Smart Graphics*, pages 232–243. Springer, 2010.

[194] Maija Mikkola, Atanas Boev, and Atanas Gotchev. Relative importance of depth cues on portable autostereoscopic display. In *Proceedings of the 3rd workshop on Mobile video delivery*, pages 63–68, 2010.

[195] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–293. International Society for Optics and Photonics, 1995.

[196] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. Applications of augmented reality for human-robot communication. In *Intelligent*

Robots and Systems' 93, IROS'93. Proceedings of the 1993 IEEE/RSJ International Conference on, volume 3, pages 1467–1472. IEEE, 1993.

[197] Pranav Mistry and Pattie Maes. Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, page 11. ACM, 2009.

[198] Tobias Müller. Towards a framework for information presentation in augmented reality for the support of procedural tasks. In *International Conference on Augmented and Virtual Reality*, pages 490–497. Springer, 2015.

[199] Tobias Müller. Challenges in representing information with augmented reality to support manual procedural tasks. 2019.

[200] Tobias Müller and Ralf Dauenhauer. A taxonomy for information linking in augmented reality. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 368–387. Springer, 2016.

[201] Tobias Müller and Tim Rieger. Arpml: The augmented reality process modeling language. In *Proceedings of the 14th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 160–163. IEEE, 2015.

[202] Miwa Nakanishi, Mugihiko Ozeki, Toshiya Akasaka, and Yusaku Okada. Human factor requirements for applying augmented reality to manuals in actual work situations. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 2650–2655. IEEE, 2007.

[203] Miwa Nakanishi, Mugihiko Ozeki, Toshiya Akasaka, and Yusaku Okada. What conditions are required to effectively use augmented reality for manuals in actual work. *Journal of Multimedia*, 3(2), 2008.

[204] Nassir Navab. Developing killer apps for industrial augmented reality. *IEEE Computer Graphics and applications*, 24(3):16–20, 2004.

[205] Ulrich Neumann and Anthony Majoros. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 4–11. IEEE, 1998.

[206] Albert Ng, Julian Lepinski, Daniel Wigdor, Steven Sanders, and Paul Dietz. Designing for low-latency direct-touch input. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 453–464. ACM, 2012.

[207] Jakob Nielsen. 10 usability heuristics for user interface design. *Nielsen Norman Group*, 1(1), 1995.

[208] David Noton and Lawrence Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968):308–311, 1971.

[209] David Noton and Lawrence Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, 11(9):929–IN8, 1971.

[210] Alex Olwal, Jonny Gustafsson, and Christoffer Lindfors. Spatial augmented reality on industrial cnc-machines. In *The Engineering Reality of Virtual Reality 2008*, volume 6804, page 680409. International Society for Optics and Photonics, 2008.

[211] SK Ong, ML Yuan, and AYC Nee. Augmented reality applications in manufacturing: a survey. *International journal of production research*, 46(10):2707–2742, 2008.

[212] Antti Oulasvirta. User interface design with combinatorial optimization. *Computer*, 50(1):40–47, 2017.

[213] Antti Oulasvirta, Niraj Ramesh Dayama, Morteza Shiripour, Maximilian John, and Andreas Karrenbauer. Combinatorial optimization of graphical user interface designs. *Proceedings of the IEEE*, 108(3):434–464, 2020.

[214] Fred Paas, Juhani E Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1):63–71, 2003.

[215] Fred GWC Paas and Jeroen JG Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743, 1993.

[216] Riccardo Palmarini, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torabmostaedi. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49:215–228, 2018.

[217] Stephen Palmer and Irvin Rock. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic bulletin & review*, 1(1):29–55, 1994.

[218] George Papagiannakis, Gurminder Singh, and Nadia Magnenat-Thalmann. A survey of mobile and wireless technologies for augmented reality systems. *Computer Animation and Virtual Worlds*, 19(1):3–22, 2008.

[219] F. Paterno, C. Mancini, and S. Meniconi. ConcurTaskTrees: A diagrammatic notation for specifying task models. In Steve Howard, Judy Hammond, and Gitte Lindgaard, editors, *Human-Computer Interaction INTERACT '97*, IFIP — The International Federation for Information Processing, pages 362–369. Springer US, 1997.

[220] Stephen Peterson, Magnus Axholt, and Stephen R Ellis. Managing visual clutter: A generalized technique for label segregation using stereoscopic disparity. In *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 169–176. IEEE, 2008.

[221] Stephen D Peterson, Magnus Axholt, Matthew Cooper, and Stephen R Ellis. Visual clutter management in augmented reality: Effects of three label separation methods on spatial judgments. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, pages 111–118. IEEE, 2009.

[222] Stephen D Peterson, Magnus Axholt, and Stephen R Ellis. Label segregation by remapping stereoscopic depth in far-field augmented reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 143–152. IEEE Computer Society, 2008.

[223] Lisa Pfannmueller, Martina Kramer, Bernhard Senner, and Klaus Bengler. A comparison of display concepts for a navigation system in an automotive contact analog head-up display. *Procedia Manufacturing*, 3:2722–2729, 2015.

[224] Marina Plavšic, M Duschl, M Tönnis, H Bubb, and Gudrun Klinker. Ergonomic design and evaluation of augmented reality based cautionary warnings for driving assistance in urban environments. *Proceedings of Intl. Ergonomics Assoc*, 2009.

[225] Jarkko Polvi, Takafumi Taketomi, Atsunori Moteki, Toshiyuki Yoshitake, Toshiyuki Fukuoka, Goshiro Yamamoto, Christian Sandor, and Hirokazu Kato. Handheld guides in inspection tasks: Augmented reality versus picture. *IEEE transactions on visualization and computer graphics*, 24(7):2118–2128, 2018.

[226] Nicholas F Polys, Doug A Bowman, and Chris North. The role of depth and gestalt cues in information-rich virtual environments. *International journal of human-computer studies*, 69(1):30–51, 2011.

[227] Nicholas F Polys, Seonho Kim, and Doug A Bowman. Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. *Computer Animation and Virtual Worlds*, 18(1):19–38, 2007.

[228] Ernst Pöppel. A hierarchical model of temporal perception. *Trends in cognitive sciences*, 1(2):56–61, 1997.

[229] Homer L Pryor, Thomas A Furness III, and Erik Viirre III. The virtual retinal display: A new display technology using scanned laser light. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 1570–1574. SAGE Publications Sage CA: Los Angeles, CA, 1998.

[230] Philip T Quinlan and Glyn W Humphreys. Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & psychophysics*, 41(5):455–472, 1987.

[231] Rafael Radkowski, Jordan Herrema, and James Oliver. Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *International Journal of Human-Computer Interaction*, 31(5):337–349, 2015.

[232] Rafael Radkowski, Jordan Herrema, and James Oliver. Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *International Journal of Human-Computer Interaction*, 31(5):337–349, 2015.

[233] Ramesh Raskar, Jeroen Van Baar, Paul Beardsley, Thomas Willwacher, Srinivas Rao, and Clifton Forlines. ilamps: geometrically aware and self-configuring projectors. *ACM Transactions on Graphics (TOG)*, 22(3):809–818, 2003.

[234] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*, pages 89–102. Springer, 2001.

[235] Holger Regenbrecht, Gregory Baratoff, and Wilhelm Wilke. Augmented reality projects in the automotive and aerospace industries. *IEEE Computer Graphics and Applications*, 25(6):48–56, 2005.

[236] Gary B Reid and Thomas E Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*, volume 52, pages 185–218. Elsevier, 1988.

[237] Dirk Reiners, Didier Stricker, Gudrun Klinker, and Stefan Müller. Augmented reality for construction tasks: Doorlock assembly. *Proc. IEEE and ACM IWAR*, 98(1):31–46, 1998.

[238] Gerhard Reitmayr and Tom Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 109–118. IEEE Computer Society, 2006.

[239] Gerhard Reitmayr and Tom W Drummond. Initialisation for visual tracking in urban environments. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 161–172. IEEE, 2007.

[240] Jun Rekimoto and Yuji Ayatsuka. Cybercode: designing augmented reality environments with visual tags. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10. ACM, 2000.

[241] Helge Renkewitz, Verena Kinder, Mario Brandt, and Thomas Alexander. Optimal font size for head-mounted-displays in outdoor applications. In *Information Visualisation, 2008. IV'08. 12th International Conference*, pages 503–508. IEEE, 2008.

[242] Miguel Ribo, Axel Pinz, and Anton L Fuhrmann. A new optical tracking system for virtual and augmented reality applications. In *Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE*, volume 3, pages 1932–1936. IEEE, 2001.

[243] Tim Rieger. Konzeption und prototypische implementierung eines webbasierten editors für ein augmented-reality-prozessmodell, 8 2015.

[244] Cindy M Robertson, Blair MacIntyre, and Bruce N Walker. An evaluation of graphical context when the graphics are outside of the task area. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 73–76. IEEE Computer Society, 2008.

[245] Cindy M Robertson, Blair MacIntyre, and Bruce N Walker. An evaluation of graphical context as a means for ameliorating the effects of registration error. *IEEE Transactions on Visualization and Computer Graphics*, 15(2):179–192, 2009.

[246] Cindy Marie Robertson. *Using graphical context to reduce the effects of registration error in augmented reality*. PhD thesis, Georgia Institute of Technology, 2007.

[247] Irvin Rock and Stephen Palmer. The legacy of gestalt psychology. *Sci Am*, 263:84–90, 1990.

[248] Jorge Rodríguez, Teresa Gutiérrez, Emilio J Sánchez, Sara Casado, and Iker Aguinaga. Training of procedural tasks through the use of virtual reality and direct aids. *Virtual Reality and Environments*, pages 43–68, 2012.

[249] Jannick P Rolland and Ozan Cakmakci. The past, present, and future of head-mounted display designs. In *Optical Design and Testing II*, volume 5638, pages 368–378. International Society for Optics and Photonics, 2005.

[250] Daniel Röltgen and Roman Dumitrescu. Classification of industrial augmented reality use cases. *Procedia CIRP*, 91:93–100, 2020.

[251] Edward Rosten, Gerhard Reitmayr, and Tom Drummond. Real-time video annotations for augmented reality. *Advances in Visual Computing*, pages 294–302, 2005.

[252] Juha Sääski, Tapio Salonen, Marja Liinasuo, Jarkko Pakkanen, Mikko Vanhatalo, Asko Riitahuhta, et al. Augmented reality efficiency in manufacturing industry: a case study. In *DS 50: Proceedings of NordDesign 2008 Conference, Tallinn, Estonia, 21.-23.08. 2008*, pages 99–109, 2008.

[253] Christian Sandor, Andrew Cunningham, Arindam Dey, and Ville-Veikko Mattila. An augmented reality x-ray system based on visual saliency. In *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 27–36. IEEE, 2010.

[254] Christian Sandor, Arindam Dey, Andrew Cunningham, Sebastien Barbier, Ulrich Eck, Donald Urquhart, Michael R Marner, Graeme Jarvis, and Sang Rhee. Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 47–50. IEEE, 2010.

[255] Andrea Sanna, Federico Manuri, Fabrizio Lamberti, Gianluca Paravati, and Pietro Pezzolla. Using handheld devices to support augmented reality-based maintenance and assembly tasks. In *2015 IEEE international conference on consumer electronics (ICCE)*, pages 178–179. IEEE, 2015.

[256] Jeff Sauro and James R Lewis. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[257] Gerhard Schall, Daniel Wagner, Gerhard Reitmayr, Elise Taichmann, Manfred Wieser, Dieter Schmalstieg, and Bernhard Hofmann-Wellenhof. Global pose estimation using multi-sensor fusion for outdoor augmented reality. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 153–162. IEEE, 2009.

[258] Torben Schinke, Niels Henze, and Susanne Boll. Visualization of off-screen objects in mobile augmented reality. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 313–316. ACM, 2010.

[259] Dieter Schmalstieg and Tobias Hollerer. *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[260] Marc Aurel Schnabel. Framing mixed realities. In *Mixed Reality In Architecture, Design And Construction*, pages 3–11. Springer, 2009.

[261] Johannes Schöning, Michael Rohs, Sven Kratz, Markus Löchtefeld, and Antonio Krüger. Map torchlight: a mobile augmented reality camera projector unit. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3841–3846. ACM, 2009.

[262] Werner Schreiber and Peter Zimmermann. *Virtuelle Techniken im industriellen Umfeld: Das AVILUS-Projekt-Technologien und Anwendungen*. Springer-Verlag, 2011.

[263] Bjorn Schwerdtfeger and Gudrun Klinker. Supporting order picking with augmented reality. In *Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 91–94. IEEE, 2008.

[264] Björn Schwerdtfeger, Daniel Pustka, Andreas Hofhauser, and Gudrun Klinker. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 134–137. ACM, 2008.

[265] Roger N Shepard and Jacqueline Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971.

[266] Fumihisa Shibata, Hiroyuki Nakamoto, Ryoichi Sasaki, Asako Kimura, and Hideyuki Tamura. A view management method for mobile mixed reality systems. In *IPT/EGVE*, pages 17–24, 2008.

[267] Tobias Sielhorst, Christoph Bichlmeier, Sandro Michael Heining, and Nassir Navab. Depth perception–a major issue in medical ar: evaluation study by twenty surgeons. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 364–372. Springer, 2006.

[268] Harvey S Smallman and Robert M Boynton. On the usefulness of basic colour coding in an information display. *Displays*, 14(3):158–165, 1993.

[269] Sidney L Smith and Jane N Mosier. *Guidelines for designing user interface software*. Mitre Corporation, 1986.

[270] Ian Sommerville and Pete Sawyer. *Requirements engineering: a good practice guide*. John Wiley & Sons, Inc., 1997.

[271] Henry Sonnet, Sheelagh Carpendale, and Thomas Strothotte. Integration of 3d data and text: the effects of text positioning, connectivity, and visual hints on comprehension. *Human-Computer Interaction-INTERACT 2005*, pages 615–628, 2005.

[272] Robert Spence. *Information visualization*, volume 1. Springer, 2001.

[273] Darko Stanimirovic, Nina Damasky, Sabine Webel, Dirk Koriath, Andrea Spillner, and Daniel Kurz. A mobile augmented reality system to assist auto mechanics. In *Proceedings of the 13th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 305–306. IEEE, 2014.

[274] Ingo Stock and Michael Weber. Authoring technical documentation using a generic document model. In *Proceedings of the 24th Annual ACM International Conference on Design of Communication*, SIGDOC '06, pages 172–179, New York, NY, USA, 2006. ACM.

[275] Natsuki Sugano, Hirokazu Kato, and Keihachiro Tachibana. The effects of shadow representation of virtual objects in augmented reality. In *Proceedings of the second IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 76–83. IEEE, 2003.

[276] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[277] J Edward Swan, Adam Jones, Eric Kolstad, Mark A Livingston, and Harvey S Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE transactions on visualization and computer graphics*, 13(3):429–442, 2007.

[278] J Edward Swan, Mark A Livingston, Harvey S Smallman, Dennis Brown, Yohan Baillot, Joseph L Gabbard, and Deborah Hix. A perceptual matching technique for depth judgments in optical, see-through augmented reality. In *IEEE Virtual Reality Conference (VR 2006)*, pages 19–26. IEEE, 2006.

[279] Anna Syberfeldt, Oscar Danielsson, Magnus Holm, and Lihui Wang. Visual assembling guidance using augmented reality. *Procedia Manufacturing*, 1:98–109, 2015.

[280] Akinari Takagi, Shoichi Yamazaki, Yoshihiro Saito, and Naosato Taniguchi. Development of a stereo video see-through hmd for ar systems. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 68–77. IEEE, 2000.

[281] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80. ACM, 2003.

[282] Vineet Thanedar and Tobias Höllerer. Semi-automated placement of annotations in videos. *Univ. California, Santa Barbara, CA, USA, Tech. Rep*, 11, 2004.

[283] Bruce Thomas, Victor Demczuk, Wayne Piekarski, David Hepworth, and Bernard Gunther. A wearable computer system with augmented reality to support terrestrial navigation. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pages 168–171. IEEE, 1998.

[284] Jenifer Tidwell. Common ground: A pattern language for human-computer interface design. *URL: http://www. mit. edu/˜ jtidwell/common_ground. html*, 1999.

[285] Makoto Tomioka, Sei Ikeda, and Kosuke Sato. Approximated user-perspective rendering in tablet-based augmented reality. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 21–28. IEEE, 2013.

[286] Elaine G Toms. Information interaction: Providing a framework for information architecture. *Journal of the American Society for Information Science and Technology*, 53(10):855–862, 2002.

[287] Marcus Tönnis, David A Plecher, and Gudrun Klinker. Representing information–classifying the augmented reality presentation space. *Computers & Graphics*, 37(8):997–1011, 2013.

[288] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[289] Mihran Tuceryan, Yakup Genc, and Nassir Navab. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators & Virtual Environments*, 11(3):259–276, 2002.

[290] Johannes Tümler. Untersuchungen zu nutzerbezogenen und technischen aspekten beim langzeiteinsatz mobiler augmented reality systeme in industriellen anwendungen (dissertation). 2009.

[291] Luca Vacchetti, Vincent Lepetit, and Pascal Fua. Fusing online and offline information for stable 3d tracking in real-time. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–241. IEEE, 2003.

[292] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1, 2010.

[293] Thomas Vincent, Laurence Nigay, and Takeshi Kurata. Classifying handheld augmented reality: Three categories linked by spatial mappings. In *Workshop on Classifying the AR Presentation Space at ISMAR 2012*, 2012.

[294] Thomas Vincent, Laurence Nigay, and Takeshi Kurata. Handheld augmented reality: spatial relationships and frames of reference. In *Workshop Designing Mobile Augmented Reality at MobileHCI*, volume 2013, 2013.

[295] Benjamin Volmer, James Baumeister, Stewart Von Itzstein, Ina Bornkessel-Schlesewsky, Matthias Schlesewsky, Mark Billinghurst, and Bruce H Thomas. A comparison of predictive spatial augmented reality cues for procedural tasks. *IEEE transactions on visualization and computer graphics*, 24(11):2846–2856, 2018.

[296] Benjamin Walther-Franks and Rainer Malaka. Evaluation of an augmented photograph-based pedestrian navigation system. In *International Symposium on Smart Graphics*, pages 94–105. Springer, 2008.

[297] Jih-fang Wang, Ronald T Azuma, Gary Bishop, Vernon Chi, John Eyles, and Henry Fuchs. Tracking a head-mounted display in a room-sized environment with head-mounted cameras. In *Helmet-Mounted Displays II*, volume 1290, pages 47–58. International Society for Optics and Photonics, 1990.

[298] Wei Wang, Yue Qi, and QingXing Wang. An augmented reality application framework for complex equipment collaborative maintenance. In *International Conference on Cooperative Design, Visualization and Engineering*, pages 154–161. Springer, 2011.

[299] Xin Wang, Soh K Ong, and Andrew YC Nee. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4(1):1–22, 2016.

[300] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012.

[301] Sabine Webel, Uli Bockholt, Timo Engelke, Nirit Gavish, Manuel Olbrich, and Carsten Preusche. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4):398–403, 2013.

[302] Sabine Webel, Uli Bockholt, Timo Engelke, Matteo Peveri, Manuel Olbrich, and Carsten Preusche. Augmented reality training for assembly and maintenance skills. In *BIO web of conferences*, volume 1, page 00097. EDP Sciences, 2011.

[303] Sabine Webel, Ulrich Bockholt, Timo Engelke, Nirit Gavish, and Franco Tecchia. Design recommendations for augmented reality based training of maintenance skills. In *Recent trends of mobile collaborative augmented reality systems*, pages 69–82. Springer, 2011.

[304] Sabine Webel, Ulrich Bockholt, and Jens Keil. Design criteria for ar-based training of maintenance and assembly tasks. *Virtual and Mixed Reality-New Trends*, pages 123–132, 2011.

[305] Stefan Wiedenmaier, Olaf Oehme, Ludger Schmidt, and Holger Luczak. Augmented reality (ar) for assembly processes design and experimental evaluation. *International journal of Human-Computer interaction*, 16(3):497–514, 2003.

[306] Jason Wither, Stephen DiVerdi, and Tobias Höllerer. Annotation in outdoor augmented reality. *Computers & Graphics*, 33(6):679–689, 2009.

[307] Jason Wither and Tobias Hollerer. Pictorial depth cues for outdoor augmented reality. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 92–99. IEEE, 2005.

[308] Matthias M Wloka and Brian G Anderson. Resolving occlusion in augmented reality. In *Proceedings of the 1995 symposium on Interactive 3D graphics*, pages 5–12, 1995.

[309] Jeremy M Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238, 1994.

[310] Harald Wuest, Florent Vial, and Didier Stricker. Adaptive line tracking with multiple hypotheses for augmented reality. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 62–69. IEEE Computer Society, 2005.

[311] Michelle Yeh and Christopher D Wickens. Attention and trust biases in the design of augmented reality displays. Technical report, Aviation Research Lab, University of Illinois, Urbana-Champaign, Savoy, Illinois, 2000.

[312] Suya You and Ulrich Neumann. Fusion of vision and gyro tracking for robust augmented reality registration. In *Virtual Reality, 2001. Proceedings. IEEE*, pages 71–78. IEEE, 2001.

[313] Suya You, Ulrich Neumann, and Ronald Azuma. Hybrid inertial and vision tracking for augmented reality registration. In *vr*, page 260. IEEE, 1999.

[314] Jürgen Zauner, Michael Haller, Alexander Brandl, and Werner Hartmann. Authoring of a mixed reality assembly instructor for hierarchical structures. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 237. IEEE Computer Society, 2003.

[315] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE, 2008.

[316] Ferdinand Rudolf Hendrikus Zijlstra. Efficiency in work behaviour: A design approach for modern tools (dissertation). 1995.