T U M

I N S T I T U T F Ü R I N F O R M A T I K

Presentation Principles in Augmented Reality - Classification and Categorization Guidelines

Marcus Toennis and David A. Plecher

TUM-I11
Juni 11

T E C H N I S C H E U N I V E R S I T Ä T M Ü N C H E N

TUM-INFO-06-I11-0/1.-FI

Druck: Institut für Informatik der
Technischen Universität München

Presentation Principles in Augmented Reality – Classification and Categorization Guidelines Version 1.0

Marcus Tönnis, David A. Plecher
Fachgebiet Augmented Reality
Technische Universität München, Fakultät für Informatik
Boltzmannstraße 3, 85748 Garching b. München, Germany
toennis@in.tum.de, plecher@in.tum.de

Abstract

Classifying the presentation space generated through the paradigm of Augmented Reality requires incorporation of different factors. Information not only can be presented in a 3D space, Augmented Reality also puts virtual information in relation to real objects, locations or events.

This document presents the current state of efforts towards such a classification of the Augmented Reality presentation space. Six classes of presentation principles are introduced and their definitions are presented. Where appropriate and necessary deeper discussions for possible future extensions are given and excurses discuss special appearances.

These classes can serve for different purposes. One purpose is the application on existing Augmented Reality systems. To facilitate such categorization tasks, restrictions on the number of entities are defined to keep the number at addressable levels. Abbreviated codewords for all entities are introduced to support illustration of presentation principles at useful length. Furthermore, examples are presented to illustrate, how applications are categorized.

1 Introduction

While presentation principles have been investigated for desktop environments for a long time and WIMP-based approaches have been established, presentation concepts for 3D environments have not been categorized in that extent. In Virtual Reality the focus often lies in immersion and extra presentation principles are mainly used for system control tasks. When stepping into Augmented Reality (AR), the demand on the user interface increases. Then no longer only one 3D world exists, rather two worlds exist where elements of the virtual world are embedded in the real world. The variety of principles on which information can be presented reaches a high level.

To know about different presentation principles is, among other things, crucial when conducting user studies. To eliminate confounding factors and to determine where the results of the study can be attributed to, only one factor may be changed at a time. In earlier work [16] we published an initial set of definitions of classes with the focus on the automotive domain. There we showed that may applications, from which the majority went into user studies, often mixed changes in multiple classes.

By getting a deeper insight into wide variety of applications and systems that have been published so far, we refined the different classes of presentation principles and spread the focus to AR in general. This report gives a survey of the new *definitions of the classes*, discussing in more detail and presenting excurses that show, why the classes are generic and allow for categorization of every presentation principle in AR. The second part of the report focuses on the application of the classes. When categorizing a wider range of applications for statistical analysis, the fine grained definition of the first part not necessarily generates valuable results as there is not yet a high number of AR systems that have been built. The definition of the *Categorization Guidelines* aggregates specific entities of classes to a categorization that provides a smaller range of entries in each class.

By categorizing concepts for new AR systems at the stage of development, confounding factors can be eliminated to study effects of changes in a presentation principle in more detail. Such an approach is necessary to get a deeper insight into the not yet fully explored presentation space of AR. With this classification at hand, the presentation space can be subdivided and new combinations that might not yet be thought of gain the potential to be investigated.

2 Definition of Classes

The following sections define six classes of presentation principles. To ease understanding of the following sections, we use an exact ontology, which in summary is as follows. This section, the definition of classes *classifies* different properties of presentation principles. Instances of these classes are called *entities*. Applications with their presentation principles are *categorized* to entities of all classes. Entities of different classes can be compound to *combinations*.

Each presentation used in an application always can be categorized to entities of all classes.

The classes were categorized on the foundation of previous work [16] which aimed on raising awareness about different presentation modalities when conducting user studies. As user studies often not only aim on testing an application on a high level, but also shall investigate effects leading to different effect, the initial set of classes was developed.

To derive a set of classes that puts its focus on presentation principles in general, a two stepped approach was executed. First, all classes were examined w.r.t cross-relationships in-between. Here, for instance, the initial class for glance behavior was discarded because most elements merged into the class categorizing mounting properties. Other elements merged into the class categorizing referencing properties. In the second step, nearly 40 publications over the last years of ISMAR, mainly taken from application sessions, were evaluated and categorized to search for exceptions. Flaws in exact categorization, finding a class at all, or missing categories were determined and added to the set of classes.

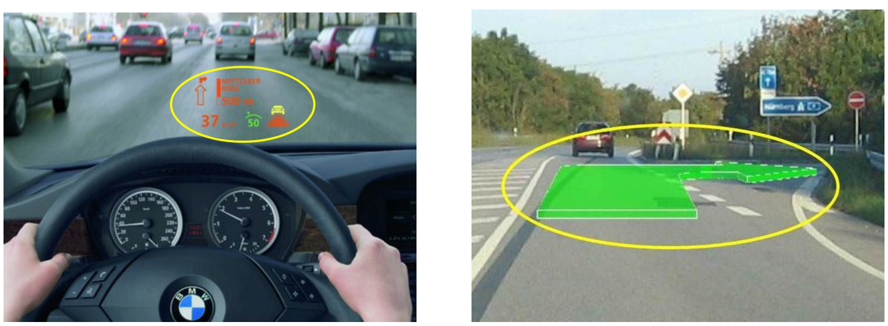
The following sections illustrate all six refined classes and provide excurses where appropriate or necessary.

2.1 Temporality

The first class categorizes the temporal property of a presentation. The class is subdivided into continuous and discrete information presentation (see Fig.1).

A permanently displayed information is called *continuous*. Such a presentation is, for instance, used in automotive head-up displays to indicate status data such as the current speed or the remaining distance to the destination.

Discrete displays are depending on certain events like completing a step of a shown construction phase or reaching a waypoint. Display duration of discrete information can differ. Short periods, for instance, can be walking by pedestrians being highlighted, long periods could be a pictogram of a gas station being shown until the tank of the car is refilled.



(a) Continuous [14]

(b) Discrete [9]

Figure 1: Types of time-dependent information presentation

Excursus: Lifetime of Object Visibility Basically can be said, that discrete and continuous information presentation are differing in the lifetime of the object visibility. A closer look at the visualization of navigational information – see Fig. 1(b) – can illustrate an conceptual issue where application design might complicate categorization.

If the driver is following the route calculated by a navigation system, he could get the impression of an continuous display. However one of the main functions of a navigational software is to change the displayed elements for guidance or to recalculate if the calculated route was left, e.g., due to a redirection of a building site. As the application is changing its state on certain events and the virtual information is depending on these such a presentation is categorized as discrete.

2.2 Dimensionality

The second class is represented by the continuum between 2D and 3D information presentation. A *2D* presentation is often used for symbolic information presentation but also can use text. In contrast, a *3D* presentation allows integrating virtual objects into the real world with a more realistic and naturalistic appearance.

In between entities such as *2.5D* may exist but are rarely used.

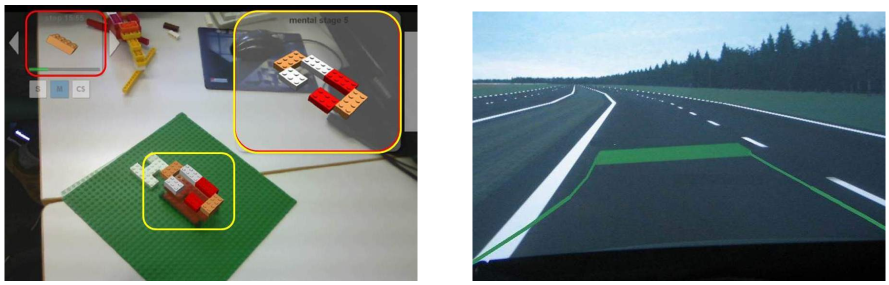
2.3 Registration

The third class deals with the registration of virtual objects to the environment. Objects can be unregistered, registered and can be presented in a so-called contact-analog manner.

An information presentation, such as a symbol, without alignment to the 3D world is called *unregistered*. Figure 2(a) illustrates such an unregistered presentation marked in the red box. In the shown application, the next structural elements to be placed is always shown in the upper left corner independent of the environment. Therefore this part is counted among the unregistered information presentation.

The second entity in the registration class is *registered* presentation. Azuma [2] already mentioned registration in his definition of AR. We deepen this by distinguishing, to what extent an object is registered. If the object is shown as if it embeds into the environment by having the same perspective and by appearing on the correct 2D position on the screen (or on the two screens in case of stereoscopic displays), we follow Azuma's definition. The group of structural elements marked yellow in Fig. 2(a) are registered with the 3D environment. The larger yellow box just enlarges what is seen in the smaller one. The user gains the impression that those objects are on the position independent from the observation pose.

Finally, *contact-analog* as an extension to registered content, is defined as follows: "contact-analog AR has a strong dependency to the physical state and behavior of the environment. AR schemes smoothly integrate into the environment." [14]. In addition to the information being respectively registered on the 2D screen, contact-analog presentation displays the objects in the correct focal depth. Figure 2(b) shows an AR application that is used in cars to visualize the breaking distance by a virtual breaking bar. This bar is displayed in the correct focal depth by use of an appropriate display. The underlying concept foresees such a depth-congruent presentation due to the aim of having no temporal dead-time for focal accommodation. The braking bar thus is fulfilling the requirements for contact-analog registration.



(a) Registered (yellow), Unregistered (red) [4]

(b) Contact-Analog [15]

Figure 2: Types of Registration

Excursus: Unregistered and Screen-fixed Presentation Registration differentiates to what extent an object requires exact positioning in the real world. While contact-analog presentation requires alignment in all three spatial dimensions, registered presentation only requires perspective rendering and alignment along the 2D screen of the display. Unregistered information does not require any alignment and thus might not require any tracking facilities. Any information displayed in conventional WIMP metaphors does not require spatial alignment and thus is unregistered.

In AR, often the term ,,screen-fixed“ is used to explain information displayed without any spatial registration. This term could lead to the assumption that the information is registered to the screen, but according the the definition it is categorized as unregistered.

2.4 Frame of Reference

The fourth class covers the frame of reference in which information is presented. W.r.t. AR already Milgram, Kishino and Colquhoun [8, 6] introduced a continuum classifying the centricity of a display. The continuum spans between egocentric and exocentric presentation.

An *egocentric* presentation uses the same point of view from which the user perceives the real scenery to place the virtual camera for object rendering.The object is seen in the frame of reference of the user.

In contrast, an object shown from another point of view, such as a mini-map, uses an *exocentric* frame of reference. Here, the virtual camera often does have no direct relationship to the user.

If the camera has a relationship to the user, e.g., a mini-map not showing the environment in a north-up manner but rather in a so-called face-up manner, we use the term *egomotion* [7]. The relation to the user can range from a single degree of freedom to several. Car racing games, for instance, often add a tethered transformation between the driven car and the eye-point.

When building AR systems, not only head-mounted displays are used but also hand-held displays. Such displays often use the camera mounted to the display to generate the superimposed image. After long discussion we decided to add two extra entities to the frame of reference class as shown in Fig. 3. The following item list discusses all entities.

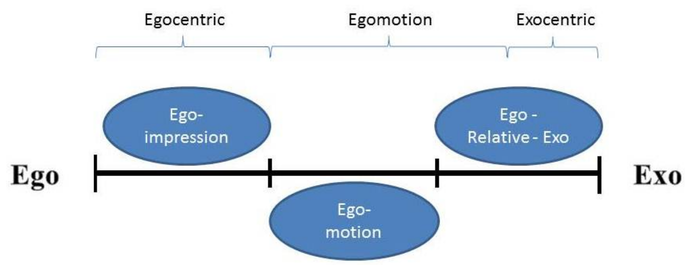


Figure 3: Ego-Exo-Continuum [10]

Egocentric The egocentric extrema shows information from the user's point of view. In consequence, the viewport of the user's eyes or of the real camera – depending on the used type of HMD – has to be the same as the viewport of the virtual camera rendering the scene.

Ego-impression This entity applies to displays differentiating between the user's eye and the position and orientation from which the scenery of the world is captured, i.e. an additional camera mounted to video see-through displays. While using a video-see through HMD the viewport of the user's eyes and the real camera are different. However, in the range of a low discrepancy the user is able to interpret and to act as with an egocentric presentation.

Egomotion The viewport of the user's eyes or the real cam is different from the viewport of the virtual camera. However the user is able to understand the displayed scene without higher mental workload. This phenomenon familiar from the side mirrors of cars is the mentioned egomotion. It is crucial that the transformation from the user's coordinate system to the coordinate system of the scene can be done very fast. So the user's location in the scene has to be known which mostly is achieved by use of an avatar.

Ego-Relative-Exo The scene shown in an exocentric frame of reference is indicating the user's position. Mental workload is increasing to transform between the coordinate systems. Expectantly, after a period of training it should be possible to interpret the given information faster. Figure 4 illustrates an application [13] for logistic purposes. The position of the user is indicated in white and of the position of the target location is shown in red. However the information is presented from another point of view, the user is able to absolve the order picking process.

Exocentric The user's viewpoint is totally different and independent from the point of view of the virtual camera. The user's location is not necessarily displayed. Interpreting the spatial relationships is left to the user.

Figure 4: Example for differentiation of the Ego-Relative-Exo entity: The exocentric view from [13] shows both positions, the position of the user and of the target object

2.5 Referencing

The fifth class deals with the relation of the object of concern w.r.t. visibility to the user of the AR system. A differentiation of such properties has been introduced in the automotive context [14] where other traffic participants may either be directly visible, concealed or outside the field of view. Broadened to AR in general, this class distinguishes between objects that are directly shown, information about the existence of concealed objects, often using indirect visualization, and guiding references to objects outside the field of view that might be visible if the user would look towards that direction.

In the first entity of the referencing class puts the object of interest in the user's field of view having it not occluded by any means. Instead of a virtual object, also any kind of information might be displayed directly. Occurrences of this type are called *direct referencing* or *direct overlay*. A prime example would be the well-known teapot that is presented unoccluded in the user's field of view.

More complex is the definition of concealed objects as it eventually comes in combination with a direct overlay. So-called *indirect referencing* or *overlay* occurs when a real object or a point of interest is not visible but lies in the field of view. Rendering the object then inverts depth but generates the indirect overlay. The object is not superimposed directly, rather the concealing object is the background for the overlay. Then, often a second, direct overlay is added to the AR scene, putting a hole into the concealing object to generate the illusion that the user is looking through the concealing object. This second object of course is a direct overlay. Milgram, Kishino and Colquhoun [8, 6] also introduced a continuum classifying the extent knowledge about the real world. To generate this second overlay, some extra knowledge about the world needs to be at hand to model the virtual world according to the real. Figure 5 illustrates an application of Avery et al. [1] representing the entity of indirect referencing, because it is visualizing an occluded scenery. More precisely, the occluding wall is superimposed by the virtual information of the background. The real background (concrete wall and yard) is indirectly superimposed on top of the real brick wall. To enhance understanding of the spatial relationships, the edges of the foreground wall are again superimposed in a direct manner.

(a) (b)

Figure 5: Indirect Referencing: X-Ray [1]

Finally, if a virtual object is pointing to a real object or location that is not in the field of view, the virtual object is called a *referencing object*. The AR object then only can give a hint about the location, specifically the direction, of the object of concern. Fig. 6 shows an AR application [17] used in cars. A virtual arrow is presenting an alert to the driver to guide his attention towards that direction, maybe to a car attending to pass the own car.

Figure 6: Referenced [17]

2.6 Mounting

The sixth and final class differentiates where a virtual object or information is mounted in the real world. This class has an infinite number of entities. Objects, for instance, can be hand-mounted, head-mounted, connected to another real object or they can lie in the world. Fig. 7 gives two examples of different mountings. The sheep in this case is hand-mounted and the colored bars are head-mounted.

Figure 7: Hand-mounting (sheep) and head-mounting (bars) [11]

Objects furthermore can have more than one mounting point. Often guidance systems such as the order-picking system by Schwerdtfeger and Klinker [12] have two mounting points, one near to the user's head, the second at the target location. Systems with more mounting points also can exist when, for instance, a third mounting point is added along the straight line of sight of the user to keep the visualized path constantly visible to the user.

3 Categorization Guidelines

The definitions of the classes provide a more or less generic view on six different presentation principles. To apply these classes to existing applications for further analysis of underlying principles the general definitions at some points require a delimitation of the entities to which visualization schemes can be categorized. Also some illustrations, how existing applications are categorized can facilitate the application of the definitions. The following sections illustrate both aspects.

3.1 Restricting the Number of Entities

The upcoming subsections restrict the numbers of entities to useful numbers if larger numbers are possible. To allow dealing with long identifiers these guidelines introduce abbreviations for each entity. Every entity of a class is associated with a special character or short word to summarize the whole categorization in one short codeword. Table 1 gives a short overview about these abbreviations beforehand. The following sections illustrate, how presentation principles are categorized.

Class	
Entity	**Code**
I. Temporalitay	
Continuous	*c*
Discrete	*d*
II. Dimensionality	
2D	*2*
3D	*3*
III. Registration	
Unregistered	*un*
Registered	*reg*
Contact-Analog	*ca*
IV. Frame of Reference	
Egocentric	*ego*
Egomotion	*m*
Exocentric	*exo*
V. Referencing	
Direct	*dir*
Indirect	*ind*
Referenced	*ref*
VI. Mounting	
Human	*hum*
Environment	*env*
World	*w*

Table 1: Abbreviation Codes

3.1.1 Temporality

Can be categorized as described.

Abbreviations:
Continuous – **c**
Discrete – **d**

3.1.2 Dimensionality

Can be categorized as described.

Abbreviations:
2D – **2**
3D – **3**

3.1.3 Registration

Can be categorized as described.

Abbreviations:
registered – **reg**
contact-analog – **ca**
unregistered – **unreg**

3.1.4 Frame of Reference

The definition of the fourth class describes the ego-exo-continuum including its five entities. Without additional information it is very complicated to categorize an application according to these, because the boundaries between egocentric and ego-impression and also ego-relative-exo and exocentric have not yet been defined on a fine-grained level. In order to get significant results after the categorization of applications that have been published, we decided to reduce the number of entities to the following three, that have been already described:

Abbreviations:
Egocentric – **ego**
Egomotion – **m**
Exocentric – **exo**

3.1.5 Reference

Can be categorized as described.

Abbreviations:
direct – **dir**
indirect – **ind**
referenced – **ref**

3.1.6 Mounting

The class categorizing mounting allows placement at almost every instance of the human body and allows for multiple mounting points. We therefore defined restrictions to groups of entities result in a suitable number of entries for each entity of each class to result in data sets that provide sufficiently meaningful data.

All mounting points at the human body, e.g. head, hand, etc. are summarized as human-mounted. If the position of an virtual object is depending on another real object in the vicinity, it will be categorized as environment-mounted.

This, for instance, applies to cars because the vehicle and its surrounding generate their own environment. Information presented on the head-up display such as a braking bar [15] thus categorizes to environmental mounting. Using GPS and other large scale tracking systems, it is possible to position virtual objects depending on the underlying coordinate system somewhere in the world. Therefore this type of mounting is called world-mounted.

Abbreviations:

Human – hum

Environment – env

World – w

3.2 Examples

Some selected examples of applications shall explain how the guidelines can be applied to categorize all appearing visualization schemes to matching class entities.

3.2.1 Virtual Mirror

The medical application "Virtual Mirror" – see Fig. 8 – introduced by Bichlmeier et al. [3] uses three different types of displays; in particular the virtual spine, the virtual mirror device and its mirror image.

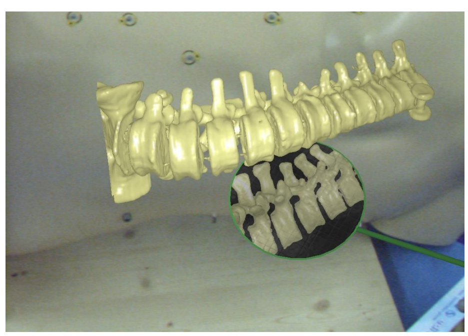


Figure 8: Virtual Mirror [3]

All of them are displayed *continuously* and in *3D*.

According to the design concept, the virtual spine is located in the correct focal depth depending on the patient's position, because that is an essential condition for medical surgeries. The presentation thus is *contact-analog* and mounted to the *human*, or more precisely, to the real spine. In this way the hidden information about the real spine, that is occluded by the patient's body, is getting visible in the surgeon's *egocentric* field of view by *indirect referencing*. In sum this display will be categorized as *c3caegoindenv*.

The categorization of the virtual mirror itself is much easier. It is a *registered* and *direct referenced egocentric* display. The virtual mirror superimposes a marker in the doctor's hand, so it is *human-mounted*, more precisely hand-mounted. That leads to the codeword *c3regegodirhum*.

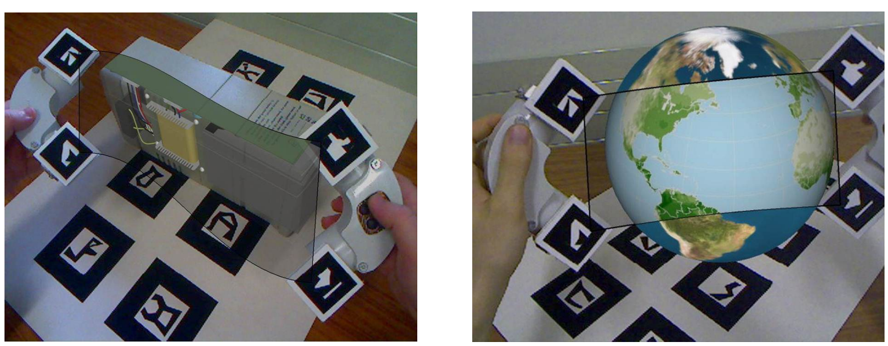
The mirror image differs from the virtual mirror only in the frame of reference and the type referencing. The image shows the spine from another point of view, therefore the viewports of the virtual and the real camera are different. Nevertheless the doctor is able to interpret the given information, because he is controlling the virtual cam with his hands. After a short adaption phase he will interact like it would be shown in an egocentric frame of reference. So the requirements of *egomotion* are fulfilled. Furthermore this type of display visualizes occluded information by *indirect referencing*. In total it will be categorized as *c3regmindhum*.

3.2.2 Magic Lens

The virtual lens [5] is located between two markers. All virtual objects are displayed *continuously* and in *3D* in the user's personal *egocentric* frame of reference. All objects are *registered* to the 3D world.

The application has two types of referencing. First, information can be represented by *indirect referencing* – see Figure 9(a). Second, the application also supports *direct referencing* – see Figure 9(b). The position is depending on the location of the lens, more exactly on the location of the user's hands. We have a dual *mounting*, both at the two hand's of the user.

That leads to the codewords *c3regegodirhum* and *c3regegoindhum*. The virtual globe is presented in the classical way and will be categorized as *c3caegodirenv*.



(a) (b)

Figure 9: Magic Lens [5]

3.2.3 X-Ray

The X-Ray application has already been described in section 2.5 when the referencing class had been introduced. The application has three presentation schemes, the scenery behind the wall, the overlay of the wall structure and the menu. As already stated in the definition of the referencing class, the scenery behind the wall has the codeword *c3regegoindenv* and the edges of the concealing wall are categorized as *c3regegodirenv*.

Now we will have a closer look at the displayed menus – see Figure 5 – along the edges of the HMD. In contrast to most other presentations in AR, the menus are displayed in *2D*. The menus do not require any kind of tracking, they are *unregistered*. To deepen understanding of the definition of unregistered presentation, we recapitulate the categorization. The display plane of the HMD has a certain focal distance to the user and appears to float in the space in front of the user's head. The presentation on the display thus is not registered to the user's head, it is unregistered. The position of the menus always stays in some respect to the orientation of the user's head, but might change if the display is shifted or recalibrated. We thus define the menus to be independent of the user's location. The menus are floating in the space near to the user, they thus get categorized as mounted to the *environment*. Finally, the menus are shown *directly*. In summary the menus are categorized as *c2unegodirenv*.

4 Summary

Classifying the presentation space of AR is an ongoing task. At the current state of work, six refined classes have been classified and illustrated. This document collects all definitions, further information and excurses.

In addition to the general definition of the classes, guidelines are presented explaining how design concepts or applications are categorized. Examples illustrate such a task.

Depending on future experiences and findings, some classes and guidelines may change or get extended. Then, updates for this document may be published under new version numbers.

References

[1] Ben Avery, Christian Sandor, and Bruce H. Thomas. Improving spatial perception for augmented reality x-ray vision. In *VR '09: Proceedings of the IEEE Conference on Virtual Reality*, pages 79–82, Lafayette, Louisiana, USA, March 2009.

[2] R.T. Azuma. A Survey of Augmented Reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.

[3] Christoph Bichlmeier, Sandro Michael Heining, Mohammad Rustaee, and Nassir Navab. Virtually Extended Surgical Drilling Device: Virtual Mirror

for Navigated Spine Surgery. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007, 10th International Conference*, pages 434–441, Brisbane, Australia, October/November 2007.

[4] T. Engelke, S. Webel, and N. 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.

[5] J. Looser, R. Grasset, and M. Billinghurst. A 3d flexible and tangible magic lens in augmented reality. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 51–54. IEEE, 2008.

[6] P. Milgram and H. Colquhoun. A taxonomy of real and virtual world display integration. *Mixed reality-merging real and virtual worlds*, pages 1–16, 1999.

[7] P. Milgram and H. Colquhoun. *Mixed Reality-Merging Real and Virtual Worlds*, chapter A Taxonomy of Real and Virtual World Display Integration, pages 5–30. Chiyodaku, Tokyo & New York: Ohmsha, Ltd. & Springer-Verlag, 1999.

[8] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems E series D*, 77:1321–1321, 1994.

[9] Marina Plavšić, Markus Duschl, Marcus Tönnis, Heiner Bubb, and Gudrun Klinker. Ergonomic Design and Evaluation of Augmented Reality Based Cautionary Warnings for Driving Assistance in Urban Environments. In *Proceedings of the International Ergonomics Association (IEA)*, August 2009. CD-ROM Proceedings.

[10] David A. Plecher. Klassifizierung von Anzeigeprinzipien in AR-Anwendungen anhand von definierten Paradigmata. Diplomarbeit, Technische Universität München, 2011.

[11] C. Sandor, A. MacWilliams, M. Wagner, M. Bauer, and G. Klinker. Sheep: The shared environment entertainment pasture. In *IEEE and ACM International Symposium on Mixed and Augmented Reality ISMAR*, volume 2002, pages 1–5, 2002.

[12] Björn Schwerdtfeger, , and Gudrun Klinker. Supporting Order Picking with Augmented Reality. In *Proceedings of the Seventh IEEE and ACM International Symposium on Mixed and Augmented Reality*, September 2008.

[13] Björn Schwerdtfeger, Troels Frimor, Daniel Pustka, and Gudrun Klinker. Mobile information presentation schemes for logistics applications. In *Proc. 16th International Conference on Artificial Reality and Telexistence (ICAT 2006)*, November 2006.

[14] Marcus Tönnis. *Towards Automotive Augmented Reality*. Dissertation, Technische Universität München, Nov 2008.

[15] Marcus Tönnis, Christian Lange, and Gudrun Klinker. Visual Longitudinal and Lateral Driving Assistance in the Head-Up Display of Cars. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, November 2007.

[16] Marcus Tönnis, Marina Plavšić, and Gudrun Klinker. Survey and Classification of Head-Up Display Presentation Principles. In *Proceedings of the International Ergonomics Association (IEA)*, August 2009. CD-ROM Proceedings.

[17] Marcus Tönnis, Christian Sandor, Christian Lange, Gudrun Klinker, and Heiner Bubb. Experimental Evaluation of an Augmented Reality Visualization for Directing a Car Driver's Attention. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, October 2005.

Contents