

# Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR

Feng Zhou<sup>1</sup>, Henry Been-Lirn Duh<sup>2</sup>, Mark Billinghurst<sup>3</sup>

<sup>1</sup>Center for Human Factors and Ergonomics, Nanyang Technological University, Singapore

<sup>2</sup>Department of Electrical and Computer Engineering/Interactive and Digital Media Institute, National University of Singapore

<sup>3</sup>The HIT Lab NZ, University of Canterbury, New Zealand

## ABSTRACT

Although Augmented Reality technology was first developed over forty years ago, there has been little survey work giving an overview of recent research in the field. This paper reviews the ten-year development of the work presented at the ISMAR conference and its predecessors with a particular focus on tracking, interaction and display research. It provides a roadmap for future augmented reality research which will be of great value to this relatively young field, and also for helping researchers decide which topics should be explored when they are beginning their own studies in the area.

**CR Categories and Subject Descriptors:** H5.1. [Information System]: Multimedia information systems—Artificial, augmented, and virtual realities

**Additional Keywords:** Augmented reality, tracking, interaction, calibration and registration, AR application, AR display

## 1 INTRODUCTION

Augmented Reality (AR) is a technology which allows computer generated virtual imagery to exactly overlay physical objects in real time. Unlike virtual reality (VR), where the user is completely immersed in a virtual environment, AR allows the user to interact with the virtual images using real objects in a seamless way. Azuma [3] provides a commonly accepted definition of AR as a technology which (1) combines real and virtual imagery, (2) is interactive in real time, and (3) registers the virtual imagery with the real world. As such there are many possible domains that could benefit from the use of AR technology such as engineering, entertainment and education.

The first AR interface was developed by Sutherland in the 1960's [86] but it has been only 10 years since the first AR conference was held; the International Workshop on Augmented Reality '98 (IWAR 98) in San Francisco, October 1998. This paper reviews the decade of research presented since then at the ISMAR (International Symposium on Mixed and Augmented Reality) conferences, and the earlier IWAR, International Symposium on Mixed Reality (ISMR) and International Symposium on Augmented Reality (ISAR) conferences. Naturally, these conferences are not the only venue for presenting AR research. However, they are the premier conferences for the AR field and so tracking research trends through them provide an interesting history of the evolution of AR research and helps identify areas for future research. Broader surveys of the AR field as a whole can be found in the original 1997 work of Azuma et al. [3] and the more recent 2001 Azuma et al. [4] article.

In the rest of the paper, we first present several research topics

presented at these conferences. Next, we provide a comprehensive review of the number of AR papers published in each of these topic areas. Then, in Sections 5, 6 and 7 respectively, we focus specifically on the important topics of AR tracking, interaction and display technology, discussing research developments, the main problems explored in the field and current and future AR research directions. We come to conclude in Section 8.

## 2 METHOD

The main method used in this research is to review previously published conference papers and other related material from the conference proceedings of ISMAR'02 to ISMAR'07 and its forerunner events, IWAR'98, IWAR'99, ISMR '99, ISMR '01, ISAR'00, and ISAR '01. There are 276 full and short papers contained in these proceedings, providing an interesting snapshot of emerging research trends in AR/MR over the last ten years. We exclude posters which are typically shorter and not normally reviewed as rigorously.

Our analysis of the collected research was specifically guided by the following three questions. (1) Which areas have been explored in AR? (2) What are the developments and key problems in these areas? And (3) what are important future trends for AR research?

In addition to analyzing paper topics, we also measured their relative impact by calculating the citation rate of the papers. This was found by taking the total number of citations as reported on Google Scholar, and dividing by number of years since publication. There are issues with the absolute accuracy of citation data from Google Scholar, such as reporting citations from non-scholarly sources, but it at least provides some indication about the relative importance of the papers. For example, searching for the term "Augmented Reality" on Google Scholar finds that there are 56 papers from 1998 to 2007 with an average citation rate of more than 10 citations per year. Of these, 15 were from the target conferences, showing that approximately 27% of the most cited AR papers from 1998 to 2007 came from the ISMAR conference and its predecessors. This percentage is increasing, demonstrating a growing influence of the ISMAR conference; considering just the papers published from 1998-2002, there are 9 papers from 36 or 25% of all the most cited AR papers. However, considering work from 2003, there are 20 papers which have an average of 10 citations per year or more, 35% of which were presented at the ISMAR conference.

## 3 AR RESEARCH TOPICS

From our own research experience and by studying the existing technology surveys [3][4], it is clear that to provide an effective AR experience there are a number of factors that must be developed, including:

(a) Graphics rendering hardware and software that can create the virtual content for overlaying the real world

(b) Tracking techniques so that the changes in the viewer's position can be properly reflected in the rendered graphics

<sup>1</sup>Email: zhou0099@ntu.edu.sg

<sup>2</sup>Email: eledbl@nus.edu.sg

<sup>3</sup>Email: mark.billinghurst@hitlabnz.org

(c) Tracker calibration and registration tools for precisely aligning the real and virtual views when the user view is fixed

(d) Display hardware for merging virtual images with views of the real world

(e) Computer processing hardware for running AR simulation code and supporting input and output devices

(f) Interaction techniques specifying how the user can manipulate the AR virtual content.

In addition there are a number of secondary topics that may be important depending on the AR application being explored, such as usability evaluation, mobile/wearable systems, AR authoring tools, visualization techniques, multimodal AR input, and novel rendering methods, software architecture, etc.

Developing complete AR applications typically involve system integration of hardware, software and techniques from topics (a) – (f). However, researchers will typically focus on one or more of these topics to explore them in more depth. In the next section we show how the conference papers cover these research areas.

#### 4 RESULTS OF REVIEW OF ISMAR PROCEEDINGS

During the evolution of AR, a variety of related research topics have been developed and discussed extensively. In our work, we grouped past AR research into the eleven categories shown in Table 1 below. These categories are based on the key topics described in the previous section. This table shows the number of papers published in ISMAR, ISMR, ISAR and IWAR in each of these categories over time, and the final total percentage break down of the papers. Note that some papers discuss several topics and are not limited to one category.

From this total set of papers it is apparent that the papers published can be broken down into two groups. The first group contains the five main research areas of:

- (1) Tracking techniques (20.1%, i.e. 63/313)
- (2) Interaction techniques (14.7%)
- (3) Calibration and registration (14.1%)
- (4) AR applications (14.4%)
- (5) Display techniques (11.8%)

These are the five topics that papers are published on most frequently, which is interesting because they are also the core AR technology areas needed to deliver an AR application. The second group of topics reflects more emerging research interests, including:

- (6) Evaluation/testing (5.8%);
- (7) Mobile/wearable AR (6.1%);
- (8) AR authoring (3.8%);
- (9) Visualization (4.8%);
- (10) Multimodal AR (2.6%);
- (11) Rendering (1.9%).

These results largely match the citation rate of most heavily cited papers presented in these conferences. We considered all ISMAR/ISAR/ISMR/IWAR papers with an average citation rate of 5.0 or more per year. Table 2 shows the proportion of papers in these areas with a citation rate of 5.0 or better. As can be seen, the percentage of citations of the most popular papers from each of these areas matches the overall publication rate, except for the cases of Display, Evaluation and Authoring. For example, even though authoring papers make up only 3.8% of the total papers published they make up 11.5% of the most cited papers, showing that there are a small number of well regarded papers. Conversely, display papers make up almost 12% of the papers, but only 6 % of the most cited papers.

These results largely match the citation rate of most heavily cited papers presented in these conferences. We considered all ISMAR/ISAR/ISMR/IWAR papers with an average citation rate of 5.0 or more per year. Table 2 shows the proportion of papers in these areas with a citation of 5.0 or better. As can be seen, the percentage of citations of the most popular papers from each of these

areas matches the overall publication rate, except for the cases of Display, Evaluation and Authoring. For example, even though authoring papers make up only 3.8% of the total papers published they make up 11.5% of the most cited papers, showing that there are a small number of well regarded papers. Conversely, display papers make up almost 12% of the papers, but only 6 % of the most cited papers.

Table 1 Papers in the literature set

Year	98	99	00	01	02	03	04	05	06	07		
Category											Total	%
Tracking	6	6	2	7	7	5	9	5	8	9	63	20.1
Interaction	2	9	2	6	3	1	3	8	9	7	46	14.7
Calibration	5	6	4	5	6	3	2	1	3	6	44	14.1
AR App.	6	7	2	9	5	8	2	2	1	4	45	14.4
Display	0	4	5	7	2	3	3	4	1	8	37	11.8
Evaluations	0	4	1	3	2	2	0	3	5	4	18	5.8
Mobile AR	1	0	1	1	0	1	1	1	3	4	19	6.1
Authoring	0	0	0	1	2	3	3	2	0	1	12	3.8
Visualization	0	0	0	2	1	3	0	2	3	5	15	4.8
Multimodal AR	0	2	0	0	0	0	1	0	3	2	8	2.6
Rendering	0	2	1	2	0	1	0	0	0	0	6	1.9
Total	20	40	18	43	28	30	24	28	35	47	313	100

We also explored research trends over time. From Table 1, for tracking and display papers there have been no trends emerging in the proportion of papers in these areas published in the conferences over the last 10 years. However, it appears that the proportions of calibration and application papers are decreasing over time, while the proportion of the interaction papers is increasing, perhaps reflecting the growing maturity of the field. There are also a number of areas (Visualization, Multimodal AR, and Authoring) in which there were no publications at the early conferences, but which have more recent research.

Table 2 Proportion of Highly Cited Papers

Topics	% Papers	% Citations
Tracking	20.1	32.1%
Interaction	14.7	12.5%
Calibration	14.1	12.5%
AR App.	14.4	12.5%
Display	11.8	5.4%
Evaluations	5.8	1.8%
Mobile AR	6.1	7.1%
Authoring	3.8	8.9%
Visualization	4.8	5.4%
Multimodal	2.6	0.0%
Rendering	1.9	1.8%

Three key findings can be found from analyzing the literature set. First, the majority of papers have been published and in the core AR technology areas (topics 1 to 5), with tracking papers being the largest group of papers published, and the most influential (highest proportion of heavily cited papers). This is not surprising since tracking is one of the fundamental enabling technologies for AR, and is still an unsolved problem with many fertile areas for research. Second, papers on interaction techniques, mobile AR, and multimodal AR are becoming more popular. This reflects a change in research from exploring fundamental technologies to applying those technologies in real world applications Third, some peripheral topics, like visualization and authoring are underrated since they have far fewer papers presented compared to other topics.

## 5 THE DEVELOPMENT OF AR RESEARCH TOPICS

As can be seen from the previous section, the research and publication areas of ISMAR have developed over time, both in the number of publications in each area, and in the research topics covered by these publications. In the remainder of this paper we focus specifically on the topics of Tracking, Interaction, and Display and how the research presented has developed in these areas. We chose to focus on these topics because they are some of the ISMAR/ISAR/ISMR/IWAR AR research areas which are the most mature and have shown the greatest development.

### 5.1 Tracking Techniques

In ten-year development of the ISMAR/ISAR/ISMR and IWAR conferences, tracking has been the most popular topic for research. The most cited paper is a tracking paper [42] with over 50 citations per year since publication and five of the top ten ISMAR papers with the highest citation rates are tracking papers. This is probably because it is one of the fundamental enabling technologies for AR (along with display and graphics). In this section we review the evolution of tracking research presented at ISMAR, focusing on sensor-based, vision-based, and hybrid tracking techniques.

#### 5.1.1 Sensor-Based Tracking Techniques

Sensor-based tracking techniques are based on sensors such as magnetic, acoustic, inertial, optical and/or mechanical sensors. They all have their respective advantages and disadvantages. For example, magnetic sensors have a high update rate and are light, but they can be distorted by any nearby metallic substance that disturbs the magnetic field. See [73] for a good review of sensor-based tracking.

Sensor tracking technologies were well developed by the time of the first IWAR 98 conference, in part due to their use in virtual reality research. Thus in these conferences there have been very little research published on tracking using only non-camera based sensors. One of the exceptions is the work of Newman et al. [57] showing how ultrasonic sensors can be used to provide wide area indoor tracking. Researchers are also exploring how to combine different types of sensors to provide robust tracking with dynamic sensor hand over between tracking technologies. For example, Klinker et al. [50] describe combining body worn local tracking systems with fixed global tracking sensors. More recently, Newman et al. [56] extend this concept to larger networks of sensors which support ubiquitous tracking and dynamic data fusion.

#### 5.1.2 Vision-Based Tracking Techniques

Sensor-based tracking systems are analogous to open loop systems whose output is perceived to have error. Vision-based tracking techniques can use image processing methods to calculate the camera pose relative to real world objects and so are analogous to closed loop systems which correct errors dynamically [6]. This is the most active area of tracking research in ISMAR with over 80% of past tracking papers describing computer vision methods.

In computer vision, most of the available tracking techniques can be divided into two classes: feature-based and model-based [66]. The rationale underlying feature-based methods is to find a correspondence between 2D image features and their 3D world frame coordinates. The camera pose can then be found from projecting the 3D coordinates of the feature into the observed 2D image coordinates and minimizing the distance to their corresponding 2D features [96].

Early computer vision papers at IWAR described marker tracking methods that could be used to calculate camera pose in real time from artificial markers. The popular ARToolKit library [42] was first presented at IWAR '99, while other papers described efficient methods for line finding [84] or minimizing marker tracker errors [18]. For example, Stricker et al. [84] describe a method

for finding the 3D coordinates of the 4 corners of a square marker, while Park et al. [62] describe an algorithm for calculating camera pose from known features.

Although square markers were the dominant techniques used in these early conference papers, they were not introduced first at IWAR. Rekimoto [71] had previously shown the key approach of combining pattern recognition and pose calculation in earlier work published elsewhere. By 2002 marker based tracking had been so thoroughly researched that Zhang et al. [97] were able to produce a paper comparing several leading approaches. After this date, there were no new general marker based systems presented, although some researchers explored tracking from LEDs [55].

Other researchers explored tracking from non-square visual markers. Cho [17] used ring shaped fiducial markers while Vogt et al. [89] designed circular shaped marker clusters with various parameters, i.e., number of markers, height, and radius with single camera tracking from its topology. Naimark and Foxlin [54] proposed a circular 2D bar-coded fiducial system for vision-inertial tracking. It offered a high information density and sub-pixel accuracy of centroid location.

Rather than using fiducial markers, camera pose can also be determined from naturally occurring features, such as points, lines, edges or textures. This research direction was also explored from the first AR conference with Park et al. [63] presenting a paper at IWAR 98 showing how natural features can be used to extend tracking beyond artificial features. After calculating camera pose from known visual features, their system dynamically acquired additional natural features and used them to continuously update the pose calculation. In this way they could provide robust tracking even when the original fiducials were no longer in view.

Since that time a range of different nature feature techniques have been applied to AR such as [18][68] and [88] (these features are often used to construct models, see below). This has been the most active area of computer vision tracking research. Of the 12 vision tracking papers with the highest citation rate, 6 present natural feature tracking methods, and only 3 present marker based approaches.

The most recent trend in computer vision tracking techniques presented at ISMAR is model-based tracking methods. These are techniques that explicitly use a model of the features of tracked objects such as a CAD model or 2D template of the object based on the distinguishable features. The first model based result presented at ISMAR was the work of Comport et al. [18] in 2003 who used a visual serving approach adapted from robotics to calculate camera pose from a range of model features (lines, circles, cylinders and spheres). They found that that knowledge about the scene improved tracking robustness and performance by being able to predict hidden movement of the object and reduce the effects of outlier data. Since that time model based tracking has quickly become one of the dominant vision methods presented at ISMAR.

Model-based trackers usually construct their models based on lines or edges in the model. Edges are the most frequently used features as they are computationally efficient to find and robust to changes in lighting. A popular approach is to look for strong gradients in the image around a first estimation of the object pose, without explicitly extracting the contours [27]. Comport et al. [18] used a CAD model which was created by hand and a piecewise parametric representation of complex objects such as straight lines, spheres, and cylinders. Wuest et al. [96] presented a real time model-based tracking approach where an adaptive system was adopted to improve the robustness and efficiency.

Texture is another useful feature for constructing models. Reitmayr and Drummond [68] presented a textured 3D model-based hybrid tracking system combined with edge information, dynamically determined at runtime by performing edge detection. Vacchetti et al. [88] combined edge information and feature points which let the tracker handle both textured and untextured objects,

and was more stable and less prone to drift. Likewise, Pressigout and Marchand [66] proposed a model-based hybrid monocular vision system, combining edge extraction and texture analysis to obtain a more robust and accurate pose computation.

As can be seen, computer vision-based tracking papers presented at ISMAR and earlier conferences began by exploring marker based methods and quickly moved from this to natural feature and model-based techniques.

### 5.1.3 Hybrid Tracking Techniques

For some AR applications computer vision alone cannot provide a robust tracking solution and so hybrid methods have been developed which combine several sensing technologies. For example, Azuma et al. [5] proposed that developing AR systems that work outdoors required a tracking system based on GPS, inertial and computer vision sensing.

In the early stages of ISMAR, hybrid methods usually relied on markers [2][41], with some exceptions (e.g. [24]). Later there was a growing consensus to combine inertial and computer vision technologies to provide closed-loop-type tracking. Vision-based tracking [51] has low jitter and no drift, but it is slow, and outliers can occur. Furthermore, drastic motions often lead to tracking failure and recovery is time-consuming with a temporary loss of real-time tracking abilities. Compared to vision-based tracking methods, inertial tracking offers attractive complementary features [51][65]. It is fast and robust and can be used for motion prediction when rapid changes occur. Moreover, object pose can be recovered based on measurements of acceleration and rotational speed, but inertial trackers tend to drift due to noise accumulation.

Klein and Drummond [48] presented a hybrid visual tracking system, based on a CAD model of edge and a serving system. It uses rate gyroscopes to track the rapid camera rotations. In [26] hybrid optical and inertial tracker that used miniature MEMS (Micro Electro-Mechanical Systems) sensors was developed for cockpit helmet tracking. The drift differential of the inertial tracking was corrected with tilt sensors and a compass. Another two papers [75][76] described hybrid head tracking using a bird eye's view and a gyroscope which reduced the number of parameters to be estimated. Reitmayr and Drummond [68] presented a model-based hybrid tracking outdoor system. Bleser et al. [14] presented a hybrid approach combining SFM (structure from motion), SLAM (Simultaneous Localisation and Mapping) and model-based tracking. A CAD model was first used for initialization and then 3D structure was recovered during the tracking allowing the camera to move away and explore unknown parts of the scene.

## 5.2 Interaction Techniques and User Interfaces

While it may be some time before AR technology becomes mature, there are many issues, both technical and social, that should be pursued in the meantime. One of the important aspects is creating appropriate interaction techniques for AR applications that allow end users to interact with virtual content in an intuitive way.

### 5.2.1 Tangible AR

AR bridges the real and virtual world and so objects in the real world can be used as AR interface elements, and their physical manipulation can provide a very intuitive way to interact with the virtual content. Previously Ishii developed the concept of tangible user interfaces (TUIs), which are interfaces where users can manipulate digital information with physical objects [40]. Tangible interfaces are powerful because the physical objects used have familiar properties, physical constraints, and affordances, making them easy to use. (Affordance refers to the actionable properties which suggests how the object should be used [30][60]). This same TUI metaphor can be applied to AR interfaces where the intuitiveness of the physical input devices can be combined with the enhanced display possibilities provided by virtual image overlays.

This new interaction metaphor is known as Tangible Augmented Reality and since its introduction at ISAR 2000 has become one of the most frequently used AR input methods [43].

A good example of the power of Tangible AR interfaces was shown in the VOMAR application developed by Kato et al. [43]. In this case a person uses a real paddle to select and arrange virtual furniture in an AR living room design application. Paddle motions are mapped to gesture based commands, such as tilting the paddle to place a virtual model in the scene and hitting a model to delete it. The intuitive mapping of application commands to the physical affordances of the paddle enable users to easily create 3D virtual scenes with no prior experience with AR technology.

Other researchers have provided more recent examples of the power of tangible AR techniques. For example, Reitmayr et al. [67] combined the physical and tangible qualities of real maps with the possibilities of dynamic and selective visualization of digital media. A set of devices were used to interact with information referenced by locations on the map. A rectangular image browser interface allowed users to access to images that were related to places. A second interaction device provided control over entities referenced to at map locations, and pushed distributed user interfaces onto remote devices such as PDA. The universal media book [32] is a mixed-reality interface for accessing information that combined projected information and a physical book that acted as the display surface. The book pages were not marked with fiducials so the users experience in interacting with data in a tactile or visually stimulating way was very natural.

One challenge of Tangible AR interfaces is to show users how to map the motions of the real objects to commands. White et al. [95] provide a nice solution by presenting AR visual hints on the real object showing how it should be moved to initiate commands. In this way they provide graphical representations for potential actions and their consequences in the augmented physical world.

Tangible AR interaction naturally leads to combining real object input with gesture and voice interaction, which often leads to multimodal interfaces. Hand gesture recognition is one of the most natural ways to interact with an AR environment [52]. A finger tracker was presented in [20] that allowed gestural interaction. Malik et al. [52] proposed a method for detecting a hand on top of a vision-based tracking pattern and to render the hand over the attached virtual object. Their method was robust against partial occlusions of the patterns. Irawati et al. [39] showed an extension of Kato's VOMAR interface that fused speech with paddle-based gesture input using time-based and semantic techniques. The overall goal of these new interaction techniques is to enable manipulating AR content to be as easy as interacting with real objects.

### 5.2.2 Collaborative AR

Although single user AR applications were studied for decades it was not until the mid-nineties that the first collaborative AR applications were developed. The Studiersube [78] and Shared Space projects [9] showed that AR can support remote and co-located activities in ways that would otherwise be impossible [72]. Since that time there have been a number of excellent examples of collaborative AR interfaces presented at the ISMAR/ISAR/IWAR and ISMR conferences.

For co-located collaboration, AR can be used to enhance a shared physical workspace and create an interface for 3D CSCW [8]. In testing with the Shared Space application, users found the interface very intuitive and conducive to real world collaboration, because unlike other interfaces, the groupware support can be kept simple and mostly left to social protocols [8].

More recently researchers have begun exploring how mobile AR platforms can be used to enhance face-to-face collaboration. Reitmayr and Schmalstieg [70] used wearable computers to enable shared object viewing and game play in the same physical space.

The Invisible Train project allowed up to 8 users to play an AR train game together on PDAs [90] and Henrysson et al. [33] demonstrated the first face-to-face collaborative AR application on mobile phones—AR Tennis. In AR Tennis two players pointed their camera phones at ARToolKit markers and saw a virtual tennis court model superimposed over the real world. They could then move their phones to hit a virtual ball to each other, hearing a sound and feeling a vibration when the virtual ball hits the real phone. A small pilot user study found that users preferred AR gaming over non AR face to face gaming and that multi-sensory feedback enhanced the game playing experience.

For remote collaboration, AR is able to seamlessly integrate multiple users with display devices in multiple contexts, enhancing telepresence. For example, an AR conferencing system was developed [42] that allowed virtual images of remote collaborators to be overlaid on multiple users' real environments. Its virtual video windows and users' eye point were placed properly to support spatial and gaze cues for natural collaboration. Stafford et al. [81] proposed a new interaction metaphor—godlike interaction that used full 3D capture and reconstruction to facilitate communication of situational and navigational information between indoor users equipped with tabletop displays, and outdoor users equipped with mobile AR systems. The outdoor AR view showed the indoor user's hand appearing from the sky, and pointed to a location in the distance like a God-like interaction.

Although there have been a number of collaborate AR prototypes created, few of them have been evaluated in formal user studies. Kiyokawa et al. [46] provide a good example of how to conduct a user study of a collaborative interface. In this case they explored how different AR display affordances may change the nature of the collaboration and how the location of the AR task space affects verbal and non-verbal communication behaviours

### 5.2.3 Hybrid AR Interface

A hybrid user interface combines a variety of different, but complementary interfaces [15]. Thus a flexible infrastructure for hybrid user interfaces should automatically accommodate a changing set of input and output devices and the interaction techniques which use them. As a result, the operations the system supports will be more than conventional ones and can even be extended to allow users to specify new operations at run time.

Butz et al. [15] described a prototype of the hybrid interface EMMIE in which a shared 3D environment manager was distributed across a heterogeneous set of computers, displays, and devices, complementing each other and supporting general information exchange mechanisms between the different displays. Likewise, Sandor et al. [74] configured a hybrid user interface where an AR overlay was presented on a head-tracked, see-through, head-worn display providing overlaid visual and auditory feedback. Their AR system could then be used to help support end users in assigning physical interaction devices to operations and virtual objects on which to perform those operations, and in documenting these mappings as the user interacts with the system. Benko et al. [7] presented the VITA collaborative mixed reality system for archaeological excavation where see-through HMDs were combined with a multi-user, multi-touch, projected table surface, a large screen display, and a tracked hand-held display. Users could use a tracked glove, speech commands, and the multi-touch sensitive surface to interact multimodally with the system and collaborate by jointly navigating, searching, and viewing data.

## 5.3 Display Techniques

Analyzing the conference proceedings, the display technologies mainly focus on three types: see-through head-mounted displays, projection-based displays and handheld displays.

### 5.3.1 See-through HMDs

See-through HMDs are mostly employed to allow the user to see the real world with virtual objects superimposed on it by optical or video technologies. They may be fundamentally divided into two categories: optical see-through (OST) and video see-through (VST) HMDs. Optical see-through displays are those that allow the user to see the real world with their natural eyes and which overlay graphics onto the users view by using a holographic optical element, half silvered mirror or similar technology. The main advantage of OST displays is that they offer a superior view of the real world including a natural, instantaneous view of the real scene, seamlessness (strictly speaking, zero-parallax, as most (but not all) of what are available today have some sort of frame surrounding the optics) between aided and periphery views as well as simple and lightweight structures. Video see through displays are those in which the user has a video view of the real world with graphics overlaid on it. The advantages of VST HMDs include consistency between real and synthetic views, and availability of a variety of image processing techniques like correction of intensity and tint, blending ratio control [44]. Thus, VST displays can handle occlusion problems more easily compared to OST displays due to various image processing techniques.

At the ISMAR conference, researchers use both OST HMDs [10][45][47][61] and VST HMDs [82]. For OST HMDs, one of the main challenges is the generation of correct occlusion effects between virtual and real objects [3]. Kiyokawa et al. [47] first presented a novel optics design for OST displays (ELMO-1) that can present mutual occlusion of real and virtual environments. Later they showed another occlusion-capable OST HMD (ELMO-4) for supporting co-located collaboration [47]. Compared to their previous work, ELMO-4 was greatly improved in terms of its contrast, resolution, pixel quality, effective FOV, and response time, etc.

Bimber and Fröhlich [10] showed a method that produced correct occlusion effects by projecting shadows onto real objects located behind virtual ones using projector-based illumination. Olwal et al. [61] presented a novel auto-stereoscopic OST system, where a transparent holographic optical element (HOE) was used to separate the views from two digital projectors. The HOE can be incorporated into other surfaces and users needn't wear special glasses. Thus it provided minimal intrusiveness and great flexibility.

For VST HMD, State et al. [82] built one with zero eye-offset from commercial components and a mount fabricated via rapid prototyping. The orthoscopic HMD's layout was created and optimized with a software simulator. Using simulation and rapid prototyping, one can quickly design and build a parallax-free, orthoscopic VST HMD suitable for demanding medical AR work. This device is likely the most sophisticated VST HMD ever constructed.

Head-mounted projective displays (HMPD) [36] are an alternative to HMDs. They typically use a pair of miniature projectors mounted on the head that project images onto retro-reflective material in the environment which is then reflected into the users eyes. The main advantages of HMPDs compared to HMDs is that they can support a large field of view (up to 90°), allow easier corrections of optical distortions, and provides the ability to project undistorted images onto curved surfaces. However, light from the HMPD needs to pass through the display optics several times which can cause a reduction in image brightness. Outside of ISMAR, the idea of HMPDs was first presented by [25]; other related work can be referred to in [37].

### 5.3.2 Projection-Based Displays

A projection-based display is a good option for applications that do not require several users to wear anything, providing minimal intrusiveness. There are a variety of projection-based display techniques proposed for displaying graphical information directly on real objects or even daily surfaces in everyday life. Bimber and Fröhlich [10] presented projector illumination techniques for creat-

ing correct occlusion effects with view-dependent OST displays. In this way, they combined the merits of projectors and OST displays. Bimber et al. [13] showed a view-dependent stereoscopic visualization on ordinary surfaces, enabled by four main components: geometric warping, radiometric compensation, multi-focal projection and multi-projector contributions.

Many researchers have seen possibilities from the fact that video projectors and cameras can be operated simultaneously [11][19], although conflicting lighting requirements made such systems very difficult to realize. For example, Bimber et al. [11] used video projectors and cameras to create a consistent illumination between real and virtual objects inside the Virtual Showcase. Cotting et al. [19] introduced a method to imperceptibly embed arbitrary binary patterns into ordinary color images for simultaneous acquisition and display, though their system lacked scalability and a working stereo projection. Ehnes et al. [22] extended the earlier work of Pinhanez [64] on using video projection to display virtual images on real objects directly. Unlike Pinhanez's work they combined video tracking technologies for tracking real object motion and keeping the augmented image on the object while it was moving.

### 5.3.3 Handheld Displays

Handheld displays are a good alternative to HMD and HMPD systems for AR applications, particularly because they are minimally intrusive, socially acceptable, readily available and highly mobile. Currently, there are several types of handheld devices which can be used for a mobile AR platform: Tablet PCs, UMPCS, and phones (cell phones, smart phones and PDAs). Most of relatively early wearable prototypes, like "Touring Machine" [24] and MARS [35], were based on tablet PCs, notebooks or custom PC hardware, and usually had hardware in a large and heavy backpack. They provided greater computational power and input options than PDAs or phones, but were bulky and considerably more expensive.

Möhring et al. [53] presented the first self-contained AR system running on a consumer cell phone. It supported optical tracking of passive paper markers and correct integration of 2D/3D graphics into the live video-stream at interactive rates. However, the drawbacks were low video stream resolution, simple graphics and memory capabilities, as well as a slow processor. Compared to cell phones, PDAs have bigger screens and faster processors. The AR-PDA setup [28] consisted of a PDA and a camera. The PDA was connected to an AR-server using WLAN and sent a video stream to the server for markerless tracking and rendering. However, this thin-client approach requires a dedicated server infrastructure and limited mobility. Recent smart phones, with integrated cameras, GPS, and fast processors and dedicated graphics hardware will be more preferred.

Wagner et al. [91] ported the ARToolKit tracking library to the Pocket PC and developed the first self-contained PDA AR application. They also developed a library called Klimt for computer graphics researchers that wrapped an underlying OpenGL ES implementation and added the most important missing OpenGL and WGL features [92].

At ISMAR, there have been a number of handheld AR applications presented, such as the AR-Kanji collaborative game [93] which allows people to use a handheld AR system for teaching Japanese Kanji characters. Schmalstieg and Wagner [79] present a framework for the development of handheld (e.g. smartphones with Windows CE) Augmented Reality, Studierstube ES. A client-server architecture for multi-user applications, and a game engine for location based museum games is constructed based on this platform.

## 6 LIMITATIONS OF AR

Despite the considerable advances made in each of these areas, there are still limitations with the technology that needs to be over-

come. In this section we describe some of the obstacles to be addressed.

### 6.1 Tracking Techniques

The main difficulty of real time 3D tracking lies in the complexity of the scene and the motion of target objects, including the degrees of freedom of individual objects and their representation. For example, moving objects may separate and merge due to (partial) occlusion or image noise, and object appearance may also change due to different illuminations. Vision-based tracking aims to associate target locations in consecutive video frames, especially when the objects are moving fast relative to the frame rate. In some cases, this can be mitigated when markers such as LEDs are allowed placed in the environment, especially in small-scale applications. Although marker-based tracking can enhance robustness and reduce computational requirements, they need certain maintenance and often suffer from limited range and intermittent errors because they provide location information only when markers are in sight. Therefore, marker-based methods are not scalable to handle large scale navigation which may be required in an outdoor environment.

Model-based methods can capitalize on the natural features existing in the environment and thus extend the range of the tracking area using natural features which are relatively invariant to illuminations. For example, edges are the most frequently used features as they are relatively easy to find and robust to changes in lighting. Often model-based tracking methods use correspondence analysis which is in turn supported by prediction. For example, Kalman filters [87] are often deployed to estimate the uncertainties of prediction. Although this is computationally efficient, its representation power is restricted because Kalman filters only apply to (approximately) linear systems that can be described by unimodal distributions. It is often not the case with a cluttered background with complex dynamics where the distribution is likely to be multimodal. Despite the fact that particle filters [87] can be applied to arbitrary, non-linear systems, they are less efficient than Kalman filters. Model-based methods also usually require the cumbersome process of modelling, especially when creating detailed models for large a cluttered environment.

Another challenge lies in how to accurately find distinguishable objects for "markers" outdoors, since they are not always available. This often leads to tracking failure. For example, the limitation of [14] is that distant features cannot be reliably reconstructed and so are not effectively used in the pose computation.

Thus hybrid tracking systems are more promising, although one often overlooked issue is the ergonomics of such systems, including size, weight, power, and ruggedness, etc.

### 6.2 Interaction Techniques and User Interfaces

AR enables users to manipulate digital objects by tangible physical tools, even directly by hand. However, there are some limitations currently, including (1) it is relatively difficult to tell the state of the digital data associated with physical tools [31], (2) the visual cues conveyed by tangible interfaces are also sparse [80], and (3) three dimensional imagery in a tangible setting can be problematic as it is dependent on a physical display surface. For example, the universal media book [32], employing markerless tracking, depends on the texture of the projected content. If the content to be rendered does not have enough texture needed for detection of feature points, the system will not be able to estimate the geometry of the page. Markers can mitigate this problem. The finger tracker in [20] is based on a special glove with retro-reflective markers. Nevertheless, as mentioned before marker-based object tracking might fail by hand-occlusion or become unstable in slanted views.

From a human factors point of view, there are also plenty of issues to be considered. Physically, the design of the system is often cumbersome, leading to discomfort for the user. For example, in

the wearable system of [77] a whole set of heavy equipment is needed to be carried for a long time. Likewise, users complained about the discomfort of wearing the interaction devices for an extended period of time and tethered cables often becoming tangled for the VITA collaborative mixed reality system [7]. In this sense, AR interfaces based on handheld devices are thus suitable, despite the fact that they have small keypads, small screens, limited resolution, bandwidth, and computational resources. However, they do provide uniqueness and attractiveness, such as mobility, light weight, and a personal gadget for social interaction, etc.

Cognitively, the complex design of the system often makes it hard to use. One good example is hybrid AR interfaces, where more complex operations tend to make it harder for users to maneuver the system and thus may reduce user satisfaction. Seamlessness may be more complicated to achieve among different computer platforms, display devices of different sizes and dimensionalities, and among different (local or remote) users. According to Norman [60], the problem of incompatibility between the user's mental model and the designer's model arises because the designer does not talk directly with the user. The designer can only talk to the user through the "system image" - the designer's materialized mental model. Usability expert Nielsen [59] gives five quality components of usability goals, which are: learnability, efficiency, memorability, error, and subjective satisfaction.

Some might argue that it is not necessary to tackle this problem while many other technical problems are still pending. It should never be too early to consider this. We even have to consider more than these two (physical and cognitive) dimensions. In the context of human-computer interaction (HCI), we need approaches to understand the radical contingency of meanings in human activities, since the experience of using an interactive artifact is not only physically, cognitively but also socially, emotionally, and culturally situated.

### 6.3 AR Displays

According to Eitoku et al. [23], display design for AR should have four guidelines: (1) Virtual information coexists with the real world; (2) Supports collaborative work; (3) Does not burden the users with special apparatus; and (4) Supports the need to naturally display three dimensional images. HMDs are the most direct and flexible way of presenting augmentations of the environment to users. However, users have to wear them continuously which might be cumbersome and uncomfortable after a long time. Meanwhile, HMDs usually have a limited FOV, which cannot effectively support collaborative work without non-verbal cues like eye contact. Some HMDs may also distort 3D images. For example, the prototype display in [47] had problems with light attenuation caused by the LCD panel and transparency; the system in [61] suffered from the image distortions inherent in all horizontal-parallax-only displays, violating guideline 4. Also, HMDs may deal not well with supporting multiple users. As an example, the shadow generated in the real surface especially for observer A might also be viewed by observer B if this real surface can be visible from multiple perspectives [10].

Projector-based displays can make the augmentation visible to several users. Furthermore, they provide high resolution, improved consistency of eye accommodation and convergence, reduced motion sickness, and the possibility of integration into common working environments [10]. Projector-based displays can combine the advantages of spatial virtual reality and spatial augmented reality [12]. However, one main disadvantage is that they usually lack mobility, because the setups for most projection-based displays are fixed. A single projector usually is set up for a single viewer and often causes self-occlusion; not all portions of the real content can be lit by a single light projector [10]. However if the user is head-tracked, then arbitrary viewpoints can be supported, and if the multiplexing is fast enough, then multiple observers/viewpoints can be

supported at the same time. Multi-projector displays can provide a much larger display area, alleviate self-occlusion problems, and seem more ubiquitous. Unlike handheld displays, they usually don't support mobile AR applications as they are fixed to the surface for which they are installed, unless they can be embedded into mobile devices.

Handheld displays can be more flexible and support mobile AR applications. The main limitation may be the tracking. Currently, most handheld displays based AR systems employ ARToolKit or similar tracking with markers, which limit the work range. However GPS, physical sensors and RFID readers are integrated into many 3G cell phones, and could be used in future hybrid tracking solutions.

## 7 RECENT TRENDS AND FUTURE DIRECTIONS

Upon reviewing work published in the past ten years in ISMAR, recent trends and several possible future directions are speculated for further research.

### 7.1 Tracking Techniques

Analyzing various tracking methods, three possible tracking research directions are identified that allow researchers to effectively capitalize on knowledge in video frames, or integrate vision-based methods with other sensors in a novel way.

The first is to incorporate a recognition system to acquire a reference representation of the real world. Tracking method are usually based on a two-stage process [29][83]; a learning or feature extraction stage, and a tracking stage. During the learning stage, some key features relatively invariant to the environment are extracted such that it is mostly a bottom-up process. Typically the computational complexity for these algorithms is low, since most of the computational load is in the training phase. It can enable re-initialization of the tracking system to recover when it fails by using automatic recognition. However, recovering camera positions in each frame independently typically result in jitter. Thus, temporal continuity relations across frames can also be considered when constructing automatic recognition systems. With great success to various state estimation problems in tracking, Rao-Blackwellized particle filters (RBPf) leverage the efficiency of Kalman filters and the representational power of particle filters might be applied for estimation of uncertainties [21].

However, the method mentioned above still heavily relies on features of the scene which the learning stage provides. If features invariant to the scene are comprehensive, it can facilitate registration. However, comprehensive features are often not readily available. Thus, tracking without these features (prior knowledge, or maps, model-free) is a more general method and might be another trend for tracking research. One of the methods is SLAM which constructs edge models from image sequences without any prior knowledge of the world [58]. Klein and Murray [49] developed an alternative to the SLAM methods without any prior model of the scene. Their method produced detailed maps with thousands of landmarks tracked at frame-rate, with an accuracy and robustness comparable to state-of-the-art model-based systems. However, this method has various tracking failure modes and mapping inadequacies, and is constrained to a small AR workspace. Further research on this direction could provide promising results, but it is mostly a top-down process and hard to deal with object dynamics, and evaluation of different hypotheses. The computational complexity for these algorithms is usually much higher. Thus, the above mentioned methods might be combined with recursive tracking to some extent such that minimal prior knowledge is needed while object dynamics can be easily handled in complex environments.

As ubiquitous AR arises from the convergence of AR and ubiquitous computing, ubiquitous tracking is on the horizon. It extends hybrid tracking one step further, where data from widespread

and diverse heterogeneous tracking sensors can be automatically and dynamically fused and then transparently provided to applications, and thus make AR systems more pervasive [38][56]. Other than using one sensor (or a set of a few sensors) that is combined with an algorithm usually specific to that particular sensor, a large number of small and simple sensors can be employed. The combination of many simple sensors (forming a denser sensor network) that individually gives information on just a small aspect of the context, results in a total picture that might be richer than the one-smart-sensor approach. Thus, tracking failure from one sensor (e.g. partial, even total occlusion, drastically changes, illumination) can be recovered from others. Besides traditional sensors, new candidate sensors might be GPS [69], RFID, etc. Fusion of measurements from different (mobile and stationary) sensors improves overall tracking quality and robustness. For example, although GPS has low accuracy, the 2D GPS position together with average user height can be used as an initial estimate for the visual tracking [69].

The third challenge is to construct a pervasive middleware to support the AR system. Since there coexists diverse heterogeneous tracking sensors, it might be rather difficult to model complex setups and issues like time in processing sensor information arises. Moreover, to fully make use of the diverse sensor data, effective information fusion algorithms must be employed, whether it's at sensor data level, feature level, or decision level.

## 7.2 Interaction Techniques and User Interfaces

AR technology creates opportunities for exploring new ways to interact between the physical and virtual world, which is a very important area for future research. Three important research paradigms—ubiquitous computing [94], tangible bits [40], and sociological reasoning to problems of interaction [85] can be explored to develop new interaction techniques. Ubiquitous computing deals with computing in the environment and with activities in the context of the environment [94]. Tangible User Interfaces augment the real physical world by coupling digital information to everyday physical objects and environments [40]. Social, psychological, cultural, organizational, and interactional context can be found in AR systems integrating social computing by making socially produced information available to their users [85]. The first two paradigms change the world itself into an interface, taking advantage of natural physical affordances [30][60] to achieve a heightened legibility and seamlessness of interaction between users and information [40].

However, in terms of design and usability, whether these physical affordances, constraints, and properties are made perceived, visible, or actionable is of great importance. In AR, most work makes use of visual cues (e.g. [67][95]) to enhance usability and user experience. Other cues, such as haptic [1], or audio [34] are less obvious, and some affordances are of no use. For example, if a screen is not touch-sensitive, though it still affords touching, it has no result on the computer system. Ubiquitous computing enables information processing capability to be present in everyday objects. Thus, it essentially extends AR interfaces to daily objects. Currently, various challenges in ubiquitous tracking make this interaction paradigm in its infancy.

Moreover, the third paradigm is to understand the social, cultural and psychological phenomena behind AR. Historically, human knowledge, experience and emotion are expressed and communicated in words and pictures. Given the advances in interface and data capturing technology, knowledge, experience and emotion might now be presented in the form of AR content. As far as we know, almost no work published in the venue of ISMAR is toward affective (emotional) user interface. However, capturing, representing and processing human affective (emotional) and cognitive (behavioral) knowledge about the real and virtual space are fundamental issues of human-computer interaction (HCI) in the

context of interactive digital media (IDM). How to leverage psychological and social impacts of the virtual and real environments across demographic user profiles is expected to shed light on the understanding of the AR development as IDM products.

## 7.3 AR Displays

During the past ten years, HMDs were commonly employed in AR systems and will still be important for some years to come. Ideally, they are tailored for each specific application. Requirements of each application will guide the design process and then determine the appropriate HMD parameters [16]. We are expecting more sophisticated HMDs with higher resolution, wide FOV, high speed and flexible head mobility; the size and weight of HMDs are designed ergonomically. Unfortunately, slow, though steady, progress is being made towards low cost and ergonomic HMDs. Researchers and system designers have an important role not only to extend the cutting-edge AR technology, but also to enlighten the public to the AR technology and foster its potential market [44].

Projection-based displays have a bright future. Future projectors can be capable of undistorted projection, and multiple viewing angles with compact size, with low power. Thus multiple projectors can be integrated into the user's daily environment to enhance the overall image quality and even support mobile or spatial AR configurations with minimal intrusion.

Recently handheld Pico Projectors have been rolled out by companies like Microvision, Foxconn and DisplayTECH. For example, the Foxconn Pico Projector<sup>1</sup> uses a 0.3-inch Texas Instruments DLP chip and projects a resolution of 854x480 pixels from a small package roughly the size of a matchbox (just 65 grams) while using less than 100 mw of power. Future Pico Projectors can be embedded in or attached as an accessory cell phones, PDAs, cameras, or stand-alone video projectors, etc.

Handheld displays are also a promising platform for mobile AR applications, especially cell phones which have a huge number of users. Furthermore, they provide an interesting research area for AR interaction metaphors and techniques, which are quite different from HMDs.

## 8 CONCLUSIONS

This paper has reviewed the development in AR research presented over the last ten years at the ISMAR, ISMR, ISAR and IWAR conferences, with a particular focus on Tracking, Interaction and Display technologies. Other topics, such as rendering techniques, calibration and registration also need to be considered to ensure that AR technology reaches a high level of maturity. Currently, to bring AR research from laboratories to industry and widespread use is still challenging, but both academia and industry believe that there is huge potential for AR technology in a wide range of areas. Fortunately, more researchers are paying attention to these areas, and it is becoming easier than ever before to be involved in AR research.

## 9 ACKNOWLEDGEMENTS

We would like to appreciate the anonymous reviewers for their constructive suggestions for enhancing this paper. We also thank other reviewers, including Ronald Azuma, Reinhold Behringer, Oliver Bimber, Steve Fiener, Gudrun Klinker, David Mizell, Ulrich Neumann, and Bruce Thomas for their invaluable criticism and advice for further improving this paper.

<sup>1</sup> <http://gizmodo.com/tag/pico-projectors/>



## REFERENCES

- [1] M. Anabuki and H. Ishii. AR-Jig: A Handheld Tangible User Interface for Modification of 3D Digital Form via 2D Physical Curve. In *ISMAR '07*, pp. 55-66, 2007.
- [2] T. Auer and A. Pinz. Building a hybrid tracking system: Integration of optical and magnetic tracking. In *IWAR '99*, pp.13-22, 1999.
- [3] R.T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6:4, 355-385, 1997.
- [4] R.T. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics & Applications*, 21:6, 34-47, 2001.
- [5] R.T. Azuma, B.R. Hoff, I. Howard E. Neely, R. Sarfaty, M. J. Daily, G. Bishop, L. Vicci, G. Welch, U. Neumann, S. You, R. Nichols, and J. Cannon. Making augmented reality work outdoors requires hybrid tracking. In *IWAR '98*, pp. 219-224, 1998.
- [6] M. Bajura and N. Ulrich. Dynamic registration correction in video-based augmented reality systems. *IEEE Computer Graphics and Applications*, 15:5 52-60, September 1995.
- [7] H. Benko, E.W. Ishak and S. Feiner. Collaborative mixed reality visualization of an archaeological excavation. In *ISMAR '04*, pp. 132-140, 2004.
- [8] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45:7, 64-70, 2002.
- [9] M. Billinghurst, S. Weghorst and T. Furness III. Shared Space: An augmented reality interface for computer supported collaborative work. In *CVE '96*, 1996.
- [10] O. Bimber and B. Fröhlich. Occlusion shadows: using projected light to generate realistic occlusion effects for view-dependent optical see-through displays. In *ISMAR '02*, pp. 186-319, 2002.
- [11] O. Bimber, A. Grundhofer, G. Wetzstein and S. Knodel. Consistent illumination within optical see-through augmented environments. In *ISMAR '03*, pp. 198-207, 2003.
- [12] O. Bimber and R. Raskar. Spatial augmented reality: Merging real and virtual worlds. A K Peters LTD, 2005.
- [13] O. Bimber and G. Wetzstein, A. Emmerling, and C. Nitschke. Enabling view-dependent stereoscopic projection in real environments. In *ISMAR '05*, pp. 14-23, 2005.
- [14] G. Bleser, H. Wuest and D. Stricker. Online camera pose estimation in partially known and dynamic scenes. In *ISMAR '06*, pp. 56-65, 2006.
- [15] A. Butz, T. Hollerer, S. Feiner, B. MacIntyre and C. Beshers. Enveloping users and computers in a collaborative 3D augmented reality. In *IWAR '99*, pp. 35-44, 1999.
- [16] O. Cakmakci and J. Rolland. Head-Worn Displays: A Review. *IEEE/OSA J. Display Technology* 2:3, 199-216, 2006.
- [17] Y. Cho, J. Lee and U. Neumann. A multi-ring fiducial system and an intensity-invariant detection method for scalable augmented reality. In *IWAR '98*, pp. 147-156, 1998.
- [18] A. Comport, E. Marchand and F. Chaumette. A real-time tracker for markerless augmented reality. In *ISMAR '03*, pp. 36-45, 2003.
- [19] D. Cotting, M. Naef, M. Gross and H. Fuchs. Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In *ISMAR '04*, pp. 100-109, 2004.
- [20] K. Dorfmueller-Ulhaas and D. Schmalstieg. Finger tracking for interaction in augmented environments. In *ISAR '01*, pp. 55-64, 2001.
- [21] A. Doucet, N. De Freitas and N. Gordon, editors. *Sequential Monte Carlo in Practice*. Springer-Verlag, New York, 2001.
- [22] J. Ehnes, K. Hirota and M. Hirose. Projected augmentation-augmented reality using rotatable video projectors. In *ISMAR '04*, pp. 26-35, 2004.
- [23] S. Eitoku, T. Tanikawa and M. Hirose. Display composed of water drops for filling space with materialized virtual three-dimensional objects. In *IEEE VR '06*, pp. 159-166, 2006.
- [24] S. Feiner, B. MacIntyre, T. Höllerer and T. Webster. A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In *ISWC '97*, pp. 74-81, 1997.
- [25] J. Fergason. Optical system for head mounted display using retro-reflector and method of displaying an image. U.S. patent 5,621,572, April 15, 1997.
- [26] E. Foxlin, Y. Altshuler, L. Naimark, and M. Harrington. Flight Tracker: A novel optical/inertial tracker for cockpit enhanced vision. In *ISMAR '04*, pp. 212-221, 2004.
- [27] P. Fua and V. Lepetit. Vision based 3D tracking and pose estimation for mixed reality. In M. Haller, M. Billinghurst, B. H. Thomas Eds, *Emerging Technologies of Augmented Reality Interfaces and Design*, pp. 43-63, Idea Group, Hershey, 2007.
- [28] C. Geiger, B. Kleinjohann, C. Reimann and D. Stichling. Mobile AR4ALL. In *ISAR '01*, pp. 181-182, 2001.
- [29] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar and N. Navab. Marker-less tracking for AR: A learning-based approach. In *ISMAR '02*, pp. 295-304, 2002.
- [30] J. J. Gibson. The theory of affordances. in R. E. Shaw and J. Bransford, Eds, *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.
- [31] M. Gorbet, M. Orth and H. Ishii. Triangles: tangible interface for manipulation and exploration of digital information topography. In *CHI '98*, pp. 49-56, 1998.
- [32] S. Gupta and C.O. Jaynes. The universal media book: tracking and augmenting moving surfaces with projected information. In *ISMAR '06*, page 177-180, 2006.
- [33] A. Henrysson, M. Billinghurst and M. Ollila. Face to face collaborative AR on mobile phones. In *ISMAR '05*, pp. 80-89, 2005.
- [34] K. Higa, T. Nishiura, A. Kimura, F. Shibata and H. Tamura. A two-by-two mixed reality system that merges real and virtual worlds in both audio and visual senses. In *ISMAR '07*, pp. 203-206, 2007.
- [35] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid and D. Hallaway. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers and Graphics*, 23:6, 779-785, 1999.
- [36] H. Hua, C. Gao and L.D. Brown, J.P. Rolland. Using a head-mounted projective display in interactive augmented environments. In *ISAR '01*, pp. 217-223, 2001.
- [37] H. Hua, A. Girardot, C. Gao and J.P. Rolland. Engineering of head-mounted projective displays. *Applied Optics*, 39:22, 3814-3824, 2000.
- [38] M. Huber, D. Pustka, P. Keitler, F. Echtler and G. Klinker. A system architecture for ubiquitous tracking environments. In *ISMAR '07*, pp. 211-214, 2007.
- [39] S. Irawati, S. Green, M. Billinghurst, A. Duenser and K. Heedong. "Move the couch where?" developing an augmented reality multimodal interface. In *ISMAR '06*, pp. 183-186, 2006.
- [40] H. Ishii, B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *ACM CHI'97*, pp. 234-241, 1997.
- [41] M. Kanbara, H. Fujii, H. Takemura and N. Yokoya. A stereo vision-based augmented reality system with an inertial sensor. In *ISAR '00*, pp. 97- 100, 2000.
- [42] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *IWAR '99*, pp. 85-94, 1999.
- [43] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto and K. Tachibana. Virtual object manipulation on a table-top AR environment. In *ISAR '00*, pp. 111-119, 2000.
- [44] K. Kiyokawa. An introduction to head mounted displays for augmented reality. In M. Haller, M. Billinghurst, B. H. Thomas Eds, *Emerging Technologies of Augmented Reality Interfaces and Design*, pp. 43-63, Idea Group, Hershey, 2007.
- [45] K. Kiyokawa, M. Billinghurst, B. Campbell and E. Woods. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *ISMAR '03*, pp. 133-141, 2003.
- [46] K. Kiyokawa, M. Billinghurst, S. E. Hayes, A. Gupta, Y. Sannohe and H. Kato. Communication behaviors of co-located users in collaborative AR interfaces. In *ISMAR '02*, pp. 139-148, 2002.

- [47] K. Kiyokawa, Y. Kurata and H. Ohno, An optical see-through display for mutual occlusion of real and virtual environments. In *ISAR '00*, pp. 60-67, 2000.
- [48] G. Klein and T. Drummond. Robust visual tracking for non-instrumented augmented reality. In *ISMAR '03*, pp. 113-122, 2003.
- [49] G. Klein and D. Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *ISMAR '07*, pp. 225-234, 2007.
- [50] G. Klinker, R. Reicher and B. Brugge. Distributed user tracking concepts for augmented reality applications. *ISAR '00*, pp. 37-44, 2000.
- [51] P. Lang, A. Kusej, A. Pinz and G. Brasseur, Inertial tracking for mobile augmented reality. In *IEEE IMTC*, Vol.2, pp. 1583-1587, 2002.
- [52] S. Malik, C. McDonald and G. Roth. Tracking for interactive pattern-based augmented reality. In *ISMAR '02*, pp. 117-126, 2002.
- [53] M. Möhring, C. Lessig and O. Bimber. Video see-through AR on consumer cell-phones. In *ISMAR '04*, pp. 252-253, 2004.
- [54] L. Naimark and E. Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *ISMAR '02*, pp. 27-36, 2002.
- [55] L. Naimark, E. Foxlin. Encoded LED system for optical trackers. In *ISMAR '05*, pp. 150-153, 2005.
- [56] J. Newman, M. Wagner, M. Bauer, A. MacWilliams, T. Pintaric, D. Beyer, D. Pustka, F. Strasser, D. Schmalstieg and G. Klinker. Ubiquitous tracking for augmented reality. In *ISMAR '04*, pp. 192-201, 2004.
- [57] J. Newman, D. Ingram and A. Hopper. Augmented reality in a wide area sentient environment. In *ISMAR '01*, pp. 77-86, 2001.
- [58] J. Neubert, J. Pretlove and T. Drummond. Semi-autonomous generation of appearance-based edge models from image sequences. In *ISMAR '07*, pp. 79-89, 2007.
- [59] J. Nielsen. (1994). *Usability Engineering* (1st ed.). San Francisco: Morgan Kaufmann.
- [60] D. A. Norman. *The Design of Everyday Things*. New York: Doubleday Business, 1988.
- [61] A. Olwal, C. Lindfors, J. Gustafsson, T. Kjellberg and L. Mattsson. ASTOR: An auto-stereoscopic optical see-through augmented reality system. In *ISMAR '05*, pp. 24-27, 2005.
- [62] J. Park, B. Jiang and U. Neumann. Vision-based pose computation: robust and accurate augmented reality tracking. In *IWAR '99*, pp. 3-12, 1999.
- [63] J. Park, S. You and U. Neumann. Natural feature tracking for extendible robust augmented realities. In *IWAR '98*, 1998.
- [64] C. Pinhanez. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces, Proc. Ubiquitous Computing 2001 (UbiComp '01), Atlanta, Georgia, September 2001, 315-331. Springer Lecture Notes In Computer Science, v. 2201.
- [65] A. Pinz, M. Brandner, H. Ganster, A. Kusej, P. Lang and M. Ribo. Hybrid tracking for augmented reality. *ÖGAI Journal*, 21:1 17-24, 2002.
- [66] M. Pressigout and É. Marchand. Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes. In *ISMAR '06*, pp. 52-55, 2006.
- [67] G. Reitmayr E. Eade and T.W. Drummond. Localisation and interaction for augmented maps. In *ISMAR '05*, pp. 120-129, 2005.
- [68] G. Reitmayr and T. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *ISMAR '06*, pp. 109-118, 2006.
- [69] G. Reitmayr and T. Drummond. Initialisation for Visual Tracking in Urban Environments, In *ISMAR '07*, pp. 161-172, 2007.
- [70] G. Reitmayr and D. Schmalstieg. Mobile collaborative augmented reality. In *ISAR '01*, pp. 114-123, 2001.
- [71] J. Rekimoto. Matrix: A realtime object identification and registration method for augmented reality. *APCHI '98*, pp.63-68, 1998.
- [72] L.S. Rodrigo Silva. Introduction to augmented reality. <http://virtual.lncc.br/~rodrigo/links/AR/node19.html>. 2003.
- [73] J. P. Rolland, L. Davis and Y. Baillet., A survey of tracking technology for virtual environments, In *Fundamentals of Wearable Computers and Augmented Reality*, 1st ed, W. Barfield and T. Caudell, Eds. Mahwah, NJ: CRC, 2001, pp. 67-112.
- [74] C. Sandor, A. Olwal, B. Bell and S. Feiner. Immersive mixed-reality configuration of hybrid user interfaces. In *ISMAR '05*, pp. 110-113, 2005.
- [75] K. Satoh, S. Uchiyama and H. Yamamoto. A head tracking method using bird's-eye view camera and gyroscope. In *ISMAR '04*, pp. 202-211, 2004.
- [76] K. Satoh, S. Uchiyama, H. Yamamoto and H. Tamura. Robust vision-based registration utilizing bird's-eye view with user's view. In *ISMAR '03*, pp. 46-55, 2003.
- [77] D. Schmalstieg, A. Fuhrmann and G. Hesina. Bridging multiple user interface dimensions with augmented reality. In *ISAR '00*, pp. 20-29, 2000.
- [78] D. Schmalstieg, A. Fuhrmann, Z. Szalavari and M. Gervautz. 1996. Studierstube-An environment for collaboration in augmented reality. In *CVE '96*, 1996.
- [79] D. Schmalstieg and D. Wagner. Experiences with Handheld Augmented Reality. In *ISMAR '07*, pp.3-18, 2007.
- [80] A. Singer, D. Hindus, L. Stifelman and S. Whitel. Tangible progress: less is more in somewire audio spaces. In *CHI*, pp. 104-111, 1999.
- [81] A. Stafford, W. Piekarski and B.H. Thomas. Implementation of god like interaction techniques for supporting collaboration between outdoor AR and indoor tabletop users. In *ISMAR '06*, pp. 165-172, 2006.
- [82] A. State, K. Keller and H. Fuchs. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *ISMAR '05*, pp. 28-31, 2005.
- [83] I. Skrypnik and D.G. Lowe. Scene modeling, recognition and tracking with invariant image features. In *ISMAR '04*, pp. 110-119, 2004.
- [84] D. Stricker, G.Klinker and D. Reiners, A fast and robust line-based optical tracker for augmented reality applications. In *IWAR '98*, pp. 31-46, 1998.
- [85] L. Suchman. Plans and situated actions: The problem of human machine communication, Cambridge: Cambridge University Press, 1987.
- [86] I. Suteland, The ultimate display. In *IFIP '65*, pp. 506-508, 1965.
- [87] S. Thrun, W. Burgard and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.
- [88] L. Vacchetti, V. Lepetit and P. Fua. Combining edge and texture information for real-time accurate 3d camera tracking. In *ISMAR '04*, pp. 48-57, 2004.
- [89] S. Vogt, A. Khamene, F. Sauer and H. Niemann. Single camera tracking of marker clusters: Multiparameter cluster optimization and experimental verification. In *ISMAR '02*, pp. 127-136, 2002.
- [90] D. Wagner, T. Pintaric, F. Ledermann and D. Schmalstieg. Towards Massively Multi-user Augmented Reality on Handheld Devices. In *Pervasive '05*, pp.208-219, 2005.
- [91] D. Wagner and D. Schmalstieg. Artoolkit on the pocketpc platform. Tech. Rep. TR-188-2-2003-23, Technical Univ. of Vienna, 2003.
- [92] D.Wagner and D.Schmalstieg. Handheld augmented reality displays. In *IEEE Virtual Reality Conference*, p. 321-321, 2006.
- [93] D.Wagner and I. Barakonyi. Augmented reality Kanji learning. In *ISMAR '03*, pp. 335-336, 2003.
- [94] M. Weiser. The computer for the twenty-first century. *Scientific American*, 265:3, 94-104, 1991.
- [95] S. White, L. Lister, and S. Feiner. Visual hints for tangible gestures in augmented reality. In *ISMAR '07*, pp. 47-50, 2007.
- [96] H. Wuest, F. Vial and D. Stricker. Adaptive line tracking with multiple hypotheses for augmented reality. In *ISMAR '05*, pp. 62-69, 2005.
- [97] X. Zhang, S. Fronz and N. Navab. Visual marker detection and decoding in AR systems: a comparative study. In *ISMAR '02*, pp. 97-106, 2002.