

Foundations and Trends® in Human-Computer
Interaction
Vol. 8, No. 2-3 (2014) 73–272
© 2015 M. Billingham, A. Clark, and G. Lee
DOI: 10.1561/1100000049



A Survey of Augmented Reality

Mark Billingham, Adrian Clark, and Gun Lee

The Human Interface Technology Laboratory New Zealand
University of Canterbury
Christchurch, New Zealand
{mark.billingham, adrian.clark, gun.lee}@hitlabnz.org

Contents

1	Introduction	74
2	Definition and Taxonomy	77
3	History	85
4	AR Tracking Technology	103
4.1	Magnetic Tracking	104
4.2	Vision Based Tracking	105
4.3	Inertial Tracking	122
4.4	GPS Tracking	122
4.5	Hybrid Tracking	123
4.6	Summary	125
5	AR Display Technology	126
5.1	Combining Real and Virtual View Images	126
5.2	Eye-to-World spectrum	139
5.3	Other Sensory Displays	144
5.4	Summary	146
6	AR Development Tools	147
6.1	Low Level software libraries and frameworks	148

6.2	Rapid Prototyping/Development Tools	153
6.3	Plug-ins to Existing Developer Tools	156
6.4	Stand Alone AR Authoring Tools	158
6.5	Summary	163
7	AR Input and Interaction Technologies	165
7.1	AR Information Browsers	166
7.2	3D User Interfaces in AR	167
7.3	Tangible User Interface in AR	169
7.4	Natural User Interfaces in AR	172
7.5	Multimodal Interaction in AR	174
7.6	Other Interaction Methods	176
7.7	Summary	177
8	Design Guidelines and Interface Patterns	179
8.1	Case Study: levelHead	183
8.2	Using Design Patterns	184
8.3	Designing for Special Needs	187
8.4	Summary	188
9	Evaluation of AR Systems	189
9.1	Types of User Studies	191
9.2	Evaluation Methods	203
9.3	Summary	206
10	AR Applications Today	207
10.1	Education	207
10.2	Architecture	211
10.3	Marketing	217
11	Research Directions	219
11.1	Tracking	220
11.2	Interaction	223
11.3	Displays	229
11.4	Social Acceptance	231
11.5	Summary	235

iv

12 Conclusion	236
Acknowledgements	238
References	239

Abstract

This survey summarizes almost 50 years of research and development in the field of Augmented Reality (AR). From early research in the 1960's until widespread availability by the 2010's there has been steady progress towards the goal of being able to seamlessly combine real and virtual worlds. We provide an overview of the common definitions of AR, and show how AR fits into taxonomies of other related technologies. A history of important milestones in Augmented Reality is followed by sections on the key enabling technologies of tracking, display and input devices. We also review design guidelines and provide some examples of successful AR applications. Finally, we conclude with a summary of directions for future work and a review of some of the areas that are currently being researched.

M. Billinghurst, A. Clark, and G. Lee. *A Survey of Augmented Reality*. Foundations and Trends® in Human-Computer Interaction, vol. 8, no. 2-3, pp. 73–272, 2014.
DOI: 10.1561/1100000049.

1

Introduction

In 1977 many moviegoers were amazed as a small robot projected a three-dimensional image of a woman in mid air. With the words "Help me Obiwan-Kenobi, you're my only hope", a recording of Princess Leia delivered a message that would change Luke Skywalker's life forever. In this Star Wars¹ scene, special effects were used to create the magical impression that three-dimensional virtual content was appearing as part of the real world. The movie forecast a future where people could interact with computers as easily as interacting with the real world around them, with digital and physical objects existing in the same space.

Thirty years later, in the 2008 US presidential campaign, a version of technology was shown for real. During the CNN election coverage reporter Wolf Blitzer turned to an empty studio and suddenly a life sized three-dimensional virtual image of reporter Jessica Yellin appeared beamed in live from Chicago². Just like Princess Leia, she appeared to be part of the real world, but this time it was real and not through movie special effects. Wolf was able to talk to her as easily

¹<http://www.starwars.com>

²<http://edition.cnn.com/2008/TECH/11/06/hologram.yellin/>

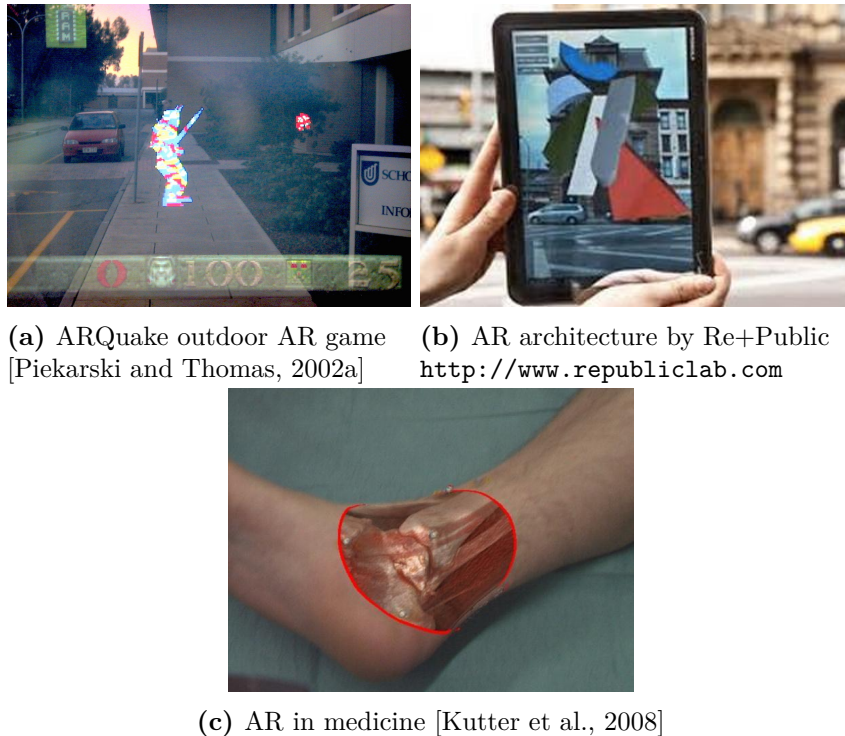
as if there was there face to face, even though she was thousands of miles away. It had taken only thirty years for the Star Wars fantasy to become reality.

The CNN experience is an example of technology known as Augmented Reality (AR), which aims to create the illusion that virtual images are seamlessly blended with the real world. AR is one of the most recent developments in human computer interaction technology. Ever since the creation of the first interactive computers there has been a drive to create intuitive interfaces. Beginning in the 1960's, computer input has changed from punch cards, to teletype, then mouse and keyboard, and beyond. One overarching goal is to make the computer interface invisible and make interacting with the computer as natural as interacting with real world objects, removing the separation between the digital and physical. Augmented Reality is one of the first technologies that makes this possible.

Star Wars and CNN showed how the technology could enhance communication and information presentation, but like many enabling technologies, AR can be used in a wide variety of application domains. Researchers have developed prototypes in medicine, entertainment, education and engineering, among others. For example, doctors can use AR to show medical data inside the patient body [Navab et al., 2007, Kutter et al., 2008], game players can fight virtual monsters in the real world [Piekarski and Thomas, 2002a], architects can see unfinished building [Thomas et al., 1999], and students can assemble virtual molecules in the real world [Fjeld and Voegtli, 2002]. Figure 1.1 shows a range of applications.

The potential of AR has just begun to be tapped and there is more opportunity than ever before to create compelling AR experiences. The software and hardware is becoming readily available as are tools that allow even non-programmers to build AR applications. However there are also important research goals that must be addressed before the full potential of AR is realized.

The goal of this survey is to provide an ideal starting point for those who want an overview of the technology and to undertake research and development in the field. This survey compliments the earlier surveys of



(a) ARQuake outdoor AR game
[Piekarski and Thomas, 2002a]

(b) AR architecture by Re+Public
<http://www.republiclab.com>

(c) AR in medicine [Kutter et al., 2008]

Figure 1.1: Typical AR applications.

Azuma [1997], Azuma et al. [2001], Van Krevelen and Poelman [2010] and Carmigniani et al. [2011] and the research survey of Zhou et al. [2008]. In the next section we provide a more formal definition of AR and related taxonomies, then a history of the AR development over the last 50 years. The rest of this survey gives an overview of key AR technologies such as Tracking, Display and Input Devices. We continue with sections on Development Tools, Interaction Design methods and Evaluation Techniques. Finally, we conclude with promising directions for AR research and future work.

7

AR Input and Interaction Technologies

In this section we review common input methods and interaction technologies for Augmented Reality. AR systems can incorporate various types of input methods, ranging from traditional 2D user interfaces (UI), such as keyboard, mouse and touch screen input, to 3D and multimodal interfaces such as handheld wands, speech, gesture, etc. Different input methods have been used for different types of AR application depending on the user interaction tasks required in the application, and so the type of interface used has evolved over time. There have been a number of AR interface types developed since the 1960's, including:

- 1) *Information Browsers*: Interfaces for showing AR information on the real world
- 2) *3D User Interfaces*: Using 3D interaction techniques to manipulate content in space
- 3) *Tangible User Interfaces*: Using real objects to interact with AR virtual content
- 4) *Natural User Interfaces*: Using natural body input such as free hand gestures



Figure 7.1: NaviCam AR browser. [Rekimoto and Nagao, 1995]

5) *Multimodal Interfaces:* Using combined speech and gesture input

We will review each of these interface types in turn and describe some interaction technologies used to implement them.

7.1 AR Information Browsers

AR information browsers are one of the representative types of AR applications where AR displays are considered as a window into an information space, and the main task of the user is to manipulate this window to browse the information. One of the earliest AR systems, NaviCam [Rekimoto and Nagao, 1995] is a typical example of AR information browser (see Figure 7.1), as well as many of the AR browser apps on smartphones such as Wikitude, Junaio, and Layar. This type of AR application requires the most basic interaction tasks: viewing the visualized AR scene and browsing the information provided.

Compared to VR systems where a navigational interface is usually needed to manoeuvre in the 3D virtual space, AR systems visualize virtual objects registered to the real world space. So users do not need additional interface object for navigation, but they can simply use their

natural skills to move through the physical environment. For instance, when using head mounted AR displays users simply have to turn their head to change their view, while with handheld AR displays holding and moving the device and pointing at the direction to view is sufficient.

Other types of interactivity common to AR information browsers includes choosing different information (e.g. channels) to view, filtering information shown, navigating into details of the information provided, and changing visualization style, etc. Most of these interactions can be accomplished using traditional 2D graphical user interfaces (GUI) and screen input. Users can use traditional input devices, such as keyboard, mouse, joystick, touch pad or touch screen, to interact with the GUI provided with the AR content.

AR information browsers are one of the important classes of AR interface that is widely used in AR navigation with wearable computers, AR simulation and training. The interaction method provided is simple and easy to learn as the users can use their knowledge of traditional mobile user interfaces. On the other hand, these types of applications have limitations in providing direct interaction with virtual objects.

7.2 3D User Interfaces in AR: Interacting with Virtual Objects through Controllers

One of the straightforward approaches for supporting direct interaction with virtual objects is to adopt traditional 3D user interfaces techniques. Various 3D UI technologies and interaction methods have been investigated to provide direct interaction with 3D objects in virtual environments [Bowman et al., 2004], and many of these techniques can be applied to AR as well.

Bowman [Bowman et al., 2004] summarizes various types of 3D interactions into three categories: (1) navigation, (2) selection, and (3) manipulation. While the 3D navigation techniques might not be directly applicable to AR applications, as navigation in AR environments is naturally achieved by the users' moving their body, selection and manipulation interaction techniques can be easily adopted in AR applications [Schmalstieg et al., 2000]. For example in Kiyokawa's VLeggo application users have their hands tracked in 3d space and can reach



Figure 7.2: Using 3D user interface in a collaborative AR system. [Kiyokawa et al., 2000]

out and directly pick up virtual blocks in the collaborative AR interface, in much the same way that they could interact in an immersive VR application (see Figure 7.2) [Kiyokawa et al., 1999].

In many cases, 3D UIs use input devices that can be used for 6 degree of freedom (DOF) manipulation (translation and rotation in 3D) of virtual objects. There are various types of devices invented and used in VR and 3D UI field, including 3D mouse or wand type pointing devices, 6 DOF joysticks, spaceballs, etc. Among them, 3D motion tracking sensors are one of the most widely used technologies. These allow tracking of various physical objects including the user's body motion, and let users to point at or manipulate virtual objects. VLEGO used 3D motion tracking magnetic sensors for manipulating virtual objects registered in the physical environment.

Haptic devices (e.g. the Phantom¹) are another type of traditional 3D UI that researchers investigated using in AR environments [Vallino and Brown, 1999, Adcock et al., 2003] (see Figure 7.3). Haptic interfaces not only work as a 3D pointing device but also provide force and tactile

¹<http://geomagic.com/en/products-landing-pages/haptic>

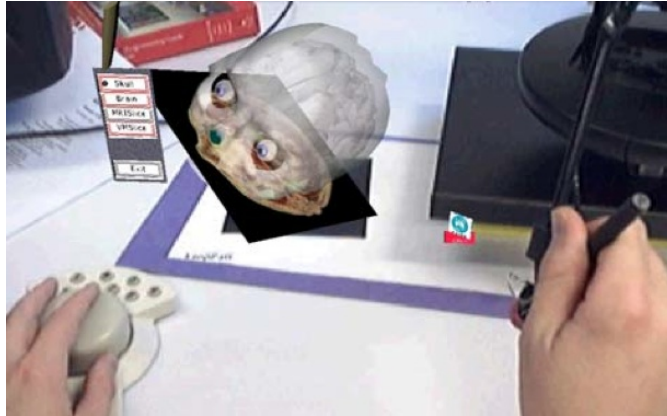


Figure 7.3: Using haptic interface in AR. [Adcock et al., 2003]

feedback which complements the visual experience by creating illusion of physical existence of a virtual object.

Traditional 3D UI can provide good interactivity in AR applications for entertainment, design, and training. With 3D UI, users can interact with 3D virtual objects everywhere in space in a natural and familiar way. While using traditional 3D UI for AR interactions benefit from previous work in VR and 3DUI fields, there are certain limitations with this approach as well. One of the most prominent problems is that the methods used for interacting with virtual objects are different from interacting with physical objects. With most of the traditional 3D UI, users will have to hold a control device and use it to point and manipulate virtual objects, while with physical objects, they mainly use their hands for direct manipulation.

7.3 Tangible User Interface in AR

Using physical objects as a medium for interacting with computers [Ishii and Ullmer, 1997] has been one of the main themes in modern human-computer interface technology research. The concept is referred to as Tangible User Interfaces (TUI). Tangible UI use physical objects for representing virtual entities and information, and to bridge between the physical and digital worlds.



Figure 7.4: Tangible AR interface with a physical book with virtual object overlaid on it.

While Tangible UI provide natural and intuitive interaction with digital information through manipulating physical objects, they can have limitations with display capabilities, either showing very limited information with different status of physical objects. Even when using a visual display in combination with Tangible UI, it still has problem of separation between the physical interaction space and the display where the virtual information is shown. For example, in a TUI digital content may be projected on a physical surface or shown on a screen while the user manipulates physical controllers some distance away creating a gap between the interaction space and presentation space.

To reduce the gap between the interaction methods for virtual and physical objects, Kato et al. [2000] proposed the concept of Tangible AR (TAR). TAR uses Tangible UI as input interaction metaphor while using AR for visualizing virtual information overlaid on the physical object used for interaction. Figure 7.4 shows a typical example of Tangible AR interface that uses a physical book as a Tangible UI while visualizing virtual objects overlaid on its pages. As can be seen in this case the interaction space and display space are seamlessly merged together.

Billinghurst et al. [2005] defines the characteristics of Tangible AR interfaces as 1) each virtual object being registered to a physical object,

and 2) the user interacting with virtual objects by manipulating physical objects. To use physical objects as input devices for interaction requires accurate tracking of the objects, and for this purpose many Tangible AR applications use computer vision based tracking software, such as ARToolkit². With accurate tracking, the computer system not only can recognize and identify different physical objects, but also can estimate the 3D motion of the objects so that the pose of the object or the motion gesture can be used for various interactions.

The basic goal of designing a Tangible AR interface is to map physical objects (input) with virtual objects (output) using an appropriate interaction metaphor. Lee et al. [2007] summarizes common interaction metaphors and methods used in various Tangible AR applications, and provides guidelines for designing and implementing interaction methods for Tangible AR applications. Providing a static mapping between the two is a very basic yet useful way for Tangible AR applications where the main purpose is to view 3D objects. However, in order to provide more interactivity, many Tangible AR applications also dynamically map them using a metaphor of tools for performing certain tasks.

Similar to how physical tools are designed, Tangible AR interfaces use two different design approaches, or a combination of both methods. One approach is space multiplexed interfaces where each physical tool is dedicated for one function, and the other is time multiplexed interfaces where a tool is used for many functions and purposes depending on the status and context. Space multiplexed interfaces are considered to be more intuitive to learn as each function is mapped to a single tool, while time multiplexed interfaces could have different function depending on the status and context hence requiring further steps of understanding and learning the functions. Figure 7.5 shows examples of Tangible AR applications that use time multiplexed and space multiplexed interfaces for interaction.

While Tangible AR interfaces provide an intuitive, natural and seamless way to interact with both physical and virtual objects in AR applications there is also a drawback of requiring physical objects for

²<http://www.hitl.washington.edu/artoolkit>



Figure 7.5: Examples of time multiplexed (left: VOMAR [Kato et al., 2000]) and space multiplexed (right: Tiles [Poupyrev et al., 2002]) Tangible AR interfaces.

interaction which might be inappropriate for mobile or wearable AR applications.

7.4 Natural User Interfaces in AR: Body Motion and Gesture

With traditional 3D user interfaces, body motion can be tracked and recognized using various types of motion tracking sensors that user's wear. Various motion sensors of different size and shape can also be applied in AR applications, ranging from a glove type device [Lee et al., 2010b] used for tracking hand gestures to full body motion tracking systems. An overview of various types of tracking technologies can be found in section 4.

With advance in computer vision technology, AR systems became capable of recognizing user's body motion and gesture in real time without requiring the user to wear any sensors. For example, Lee and Hollerer [2007] developed the HandyAR system capable of bare hand interaction using a standard web camera, although the supported gestures were limited (see Figure 7.6). Another work [Lee et al., 2008b] used a stereo camera to detect natural hand interaction. Most of this research focused on recognizing the human body through detecting features or skin colour Lee and Hollerer [2007], McDonald and Roth [2003] from video image that had limitations of accuracy of recognized hand postures yet requiring high speed processing power for running computer vision algorithms. Some research [Lee et al., 2004a] tried to

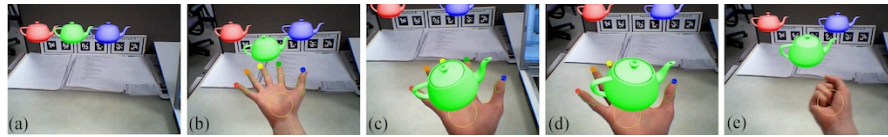


Figure 7.6: HandyAR free hand interaction with AR content. [Lee and Hollerer, 2007]

overcome this through reversing the focus from human body to objects of interest and used occlusions by the user's body as an input for interaction.

With recent technical advances and commercialization of depth cameras (e.g. Microsoft Kinect³), more accurate motion and gesture-based interaction became widely available for VR and AR applications (see Figure 7.7). Use of depth camera in AR applications enabled tracking dexterous hand motions for physical interaction with virtual objects using bare hands. Microsoft HoloDesk [Hilliges et al., 2012] demonstrated use of the Kinect camera to recognize the user's hands and other physical objects and allow them to interact with virtual objects shown on an optical see-through AR workbench. ARET [Corbett-Davies et al., 2013] shows a similar technique applied in AR based exposure treatment application. With growing interest in using hand gestures for interaction in AR applications, Piumsomboon et al. [2013] categorized a user-defined gesture set that can be applied to different tasks in AR applications.

Integrating hand motion and gesture based interaction with mobile or wearable AR systems is one of the topics actively investigated. Mistry and Maes [2009] demonstrated a wearable camera and projector system that recognizes user's hand gestures for interaction by detecting fingertips with color markers (see Figure 7.8). Using the camera on Google Glass, On the Go Platforms⁴ developed hand gesture recognition software though supported gestures are limited to simple ones. Along the efforts for improvement, mobile version of depth sensing

³<http://www.microsoft.com/en-us/kinectforwindows>

⁴<http://www.otgplatforms.com>

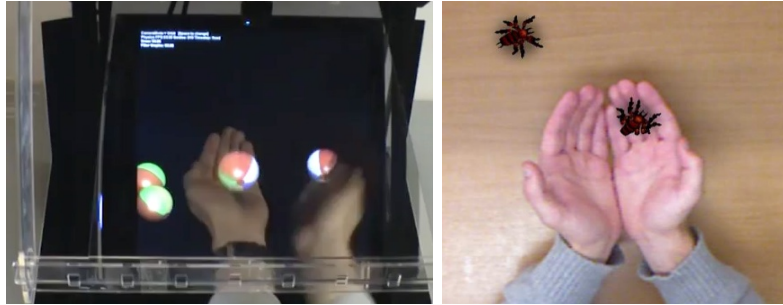


Figure 7.7: Hand gesture based physical interaction with virtual objects (left: HoloDesk [Hilliges et al., 2012], right: ARET [Corbett-Davies et al., 2013]).

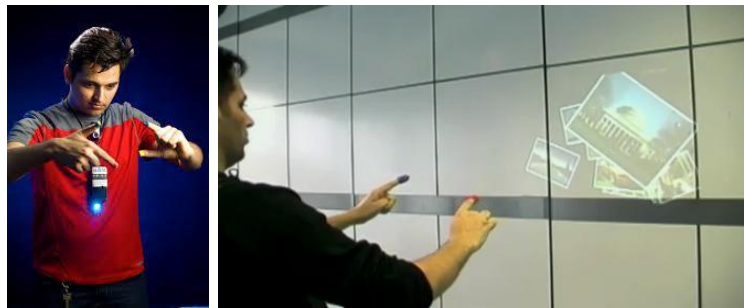


Figure 7.8: Tracking fingers positions in the Sixth Sense projection based AR interface. [Mistry and Maes, 2009]

cameras are also under development, such as SoftKinetic⁵, that would benefit from implementing more accurate hand gesture interaction in mobile and wearable AR systems. There are also other approaches to use biometric sensors for gesture interaction in mobile and wearable environment, such as the Myo⁶ gesture armband controller.

7.5 Multimodal Interaction in AR

To provide richer interactivity in AR applications, there have been efforts to combine different modalities of input. Among different combination of input modalities, speech and gesture recognition combined is

⁵<http://www.softkinetic.com>

⁶<https://www.thalmic.com/myo>

one of the most widely and actively researched combinations. Previous research has shown that multimodal interfaces (MMI) that combine speech and gesture input can be an intuitive way to interact with 2D interfaces and 3D graphics on the desktop [Bolt, 1980, Hauptmann, 1989, Cohen et al., 1997]. This is because the input modalities are complimentary, with speech being good for quantitative input, while gesture is ideal for qualitative input [Cohen et al., 1989].

A number of researchers have used speech and gesture input in Virtual Reality interfaces. For example, the Naval Research Laboratory's Dragon 3D VR system [Cohen et al., 1999] used a multimodal system to create digital content in a 3D topographical scene, allowing users to create and position objects by speaking while gesturing in 3D space. Ciger et al. [2003] presented a multimodal user interface that combined a magic wand with spell casting. The user could navigate in the virtual environment, grab and manipulate objects using a combination of speech and pointing with a physical wand. Similarly, LaViola Jr. [1999] developed an immersive VR interface that used whole-hand input captured with a data glove and speech input for interior design. The user could create virtual objects with speech and then perform object manipulation using hand gestures. These examples are just a few of many and show that MMI are also a natural way to interact in immersive virtual environments.

However, despite this earlier research there has relatively little research on the use of multimodal interaction for AR. One of the first systems was the SenseShapes work of Olwal et al. [2003] which used volumetric regions of interest that can be attached to the user, allowing the user to point at virtual objects in the real world and manipulate them with speech commands. Kaiser et al. [2003] extended Olwal's SenseShapes work by focusing on mutual disambiguation between input channels (speech and gesture) to improve interpretation robustness. Heidemann et al. [2004] presented an AR system designed for online acquisition of visual knowledge and retrieval of memorized objects. More recently, Irawati et al. [2006] added speech input to the earlier VOMAR [Kato et al., 2000] interface allowing people to create AR scenes using a combination of speech and tangible paddle gestures. In a user study

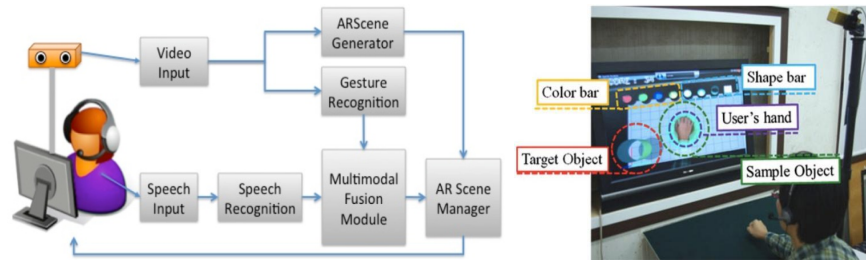


Figure 7.9: Natural hand gesture based multimodal AR system. [Lee et al., 2013b]

with her system, she found that people were able to complete a scene creation task over 35% faster with the interface using combined speech and paddle gesture input, than using gesture input on its own. The multimodal interface was also more accurate and was preferred by the users.

These multimodal systems all required the users to either wear gloves or hold a special input tool in their hands. More recently systems have been developed that support natural bare hand input. For example, Kolsch et al. [2006] create a multimodal interface for outdoor AR that used a wearable computer to perform hand tracking and recognize command gestures, enabling users to interact in a very intuitive manner with AR cues superimposed the real world. Piumsomboon et al. [2014] used a Wizard of Oz technique to classify the types of gesture that users would like to use in an AR multimodal interface, helping to inform the design of such systems. Finally, Lee et al. [2013b] developed a multimodal system that used a stereo camera to tracking users hand gestures and allow them to issue speech and gesture commands to manipulate virtual content in a desktop AR interface (see Figure 7.9. A user study with the system found that the MMI was 25% faster than using gesture only interaction and that users felt that the MMI was the most efficient, fastest, and accurate interface.

7.6 Other Interaction Methods

While speech recognition is the most actively investigate as audio input, there are also attempts to detect other types of sound for interaction

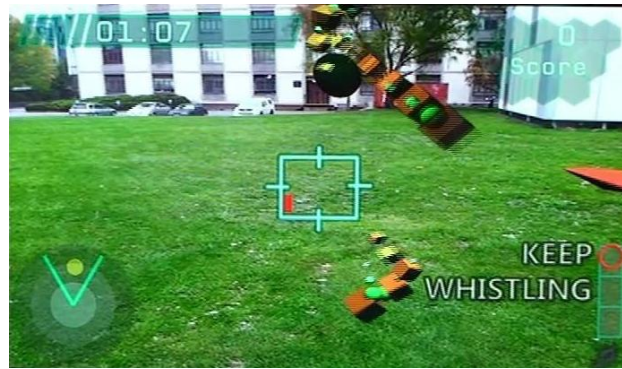


Figure 7.10: Whistle recognition in GeoBoids mobile AR game. [Lindeman et al., 2012]

in AR applications. For instance, Lindeman et al. [2012] used whistle recognition in a mobile AR game where users have to whistle for the right pitch and length to scare virtual creatures in the game (see Figure 7.10).

Another type of input modality gaining interest of AR researchers is Brain Computer Interfaces (BCI). While BCI is at the very early stage of investigation, studies show BCI and AR can be a good combination where AR can be efficient and powerful tool for enhancing and studying BCI technology [Kansaku et al., 2010, Lenhardt and Ritter, 2010, Scherer et al., 2010]. While traditional electroencephalogram (EEG) sensing equipments required cumbersome setup of wearing the sensors, recent development of the technology made a headset style sensor (e.g. NeuroSky⁷) that can be easily integrated with AR applications (See Figure 7.11).

7.7 Summary

In this section we have reviewed some of the common input techniques and interaction methods used in AR systems. As we have seen they have evolved from simple information browsing which support very little interaction with the virtual content, to complex interfaces that can

⁷<http://neurosky.com>



Figure 7.11: A mobile BCI headset from NeuroSky. <http://neurosky.com>

recognise and respond to natural gesture input and speech. However, even though there are now a wide range of different input methods available for AR interfaces, they still need to be carefully designed. In the next section we review research into design guideline and interface patterns for effective AR experiences.