Achieving total immersion:
Technology trends behind Augmented Reality - A survey

GABOR SZIEBIG
Narvik University College
Department of Industrial Engineering
Lodve Langes gate 2., 8514 Narvik
NORWAY
gabor.sziebig@hin.no

Abstract: Virtual Environments allow users to interact with virtual worlds, but usually these interactions are in front of a monitor or a projected wall. Augmented Reality brings the feeling of reality to these interactions. The user can combine the virtual world with the virtual world. Interact and feel the objects, together with artefacts that are generated by a computer. This article provides an overview and introduction to the problems of Augmented Reality systems and shows the state of the art in current technological developments. The usability bottlenecks of existing systems and probable solutions are also discussed.

Key–Words: Augmented Reality, Usability, Head-mounted displays, Trackers, Immersion

1 Introduction

Augmented Reality (AR) is a special type of Virtual Reality. It combines the real-world with computer generated virtual data [1]. This enables the programmers to create such a mixed environment, where virtual objects can be attached to real things or placed into an otherwise real environment. It is a special type of human-machine interaction, which utilizes the user’s perceptual-motor skills in the real world. This is due to the tangible interface, which can be virtual or real. Simulation, training, assistance, telemanipulation and communication can be performed in a whole new way [2]. AR systems exist in many sizes: from a simple handheld device, that can show virtual objects superposed on real objects using a camera and the display of the device, up to world sized environments, where the user is tracked by global positioning system (GPS) and the real world and virtual data is shown overlaid on a head mounted display (HMD) based on the position of the user and the scene that can be seen from the HMD [3]. Each size has different level of reality feeling and immersion level. The scope of this paper is limited to technologies that let humans enter a virtual world without mayor restrictions and tabletop [4], handheld [5] or desktop [6] based AR systems are left out.

Until now true AR systems (that utilizes each feature of AR) are not available for commercial use. This is mainly, because of the different needs (system flexibility, type of user-system interaction, level of immersion, hardware and software elements) of each AR application and the lack of general framework and the developed AR systems were mainly a result of a research projects which focused on specified problems. AR is mainly a gathering name for the technologies that involves the following techniques: haptics, stereoscopic visualization, human-machine inter-action, motion tracking, virtual reality. This already shows how widespread the field of AR systems is. Although there are successful demonstrations in different application fields and a huge interest in research fields, the big commercial breakthrough of AR systems is still waiting. Some of the successful applications are in the following domains: entertainment [7], maintenance [8], [9], manufacturing [10]-[14], medical [15], [16], military training [17], [18] and telerobotics [19].

In this article a typical setup of AR systems is discussed. In general, such a system consists of two layers: hardware and software. These will be introduced in the following subsections: in Section 2.1 the needed hardware elements (HMD, trackers, haptics) of an AR system are introduced, which is followed by the necessary software solutions in Section 2.2. In Section 3, these hardware and software trends and developments are summarized.

2 System components

As already mentioned in the introduction, the AR field stands on widespread technological foundation. It utilizes different hardware components (haptic devices,
HMDs, monitors, motion capture equipment, etc.) with software components (image processing, object recognition, virtual word rendering, etc.). In the past ten years research focused on the general bottlenecks of AR systems: tracking of human position, tracking of real-world objects, synchronizing the virtual- and real-world, elimination of lags, jitters and drift caused by the hardware elements. Not only the technology advanced, but also the AR software interfaces [20]. The most widespread AR software platform is the ARToolkit [21], which will be introduced in Section 2.2, and the common hardware elements are cameras connected to video displays (see Section 2.1.1). In the next subsections the AR systems different components will be introduced.

2.1 Hardware

2.1.1 Head-mounted displays

There are two basic types of HMDs: optical- and video- see-through. In case of the optical see-through system the real-world and the computer generated image are mixed in an optical combiner (transparent display) in front of the user eyes. In the other case (video see-though) the real-world is captured by digital video cameras and the camera images are overlapped with computer generated images and displayed on monitors in front of the user’s eyes.

Both technologies suffer from drawbacks and have a lot of technological difficulties. It is much easier to create AR system with a video see-through system, because the user only see, what the system generates based on the video cameras images, but it much harder to achieve the feeling of reality than in the optical see-through system. Also registration of the head position in both cases highly defines the overall performance of the system (this problem is also referred as Tracking of head motion and will be introduced in Section 2.1.2).

10 years ago, when augmented reality was just starting to spread around the world, there were only a few commercially available systems (MicroOptical, Minolta, Sony, MicroVision, etc.) and these systems have mainly disappeared by now [1]. As technology advanced, HMDs got more and more functions and better and better displays. Now the market consists of more than 38 different products, starting from entry level (low resolution, cheap) to experimental (high FOV, expensive) high-end systems. In general there are more video see-through (over 80%) than optical see-through systems, this is because of the expensive and technologically challenging optical combiner. The average resolution is 800*600 and the FOV is around 35 degrees, which is far less than is required for full immersion. Natural human vision has a field of view of up to 80 degrees without eye movement, although only 1 degree provides sharp sight. Humans have an oculomotor horizontal range of 55 degrees without moving the head [22].

Manufacturers are continually trying to make their equipment more user friendly, but not with much success. The reasons of this are the following:

- The average weight of high-end HMD is 700 grams, while normal reading glasses weight around 100 grams.
- The HMD needs fixed and known position relative to head, which can be only achieved through a helmet or a belt, this is usually uncomfortable.
- There are cables that run to the HMD, is usually not wireless, so the movement of the user is limited within the range of the wires.
- The optical see-through HMDs are more convenient to use and give higher feeling of immersion, but they do not provide the same FOV as the video see-through system.
- Display resolution is not high enough.

Until the above mentioned problems are not solved, the commercialization of the HMDs will be limited only to researchers or task specific applications.

2.1.2 Trackers

In order to match the virtual world with the real-world, the movement of the objects in both worlds must be tracked. In the virtual case everything is created in an artificial way, where positions are readily defined. It is much harder to follow changes in the real-world. For this reason different type of trackers are used: electromagnetic, mechanical, inertial, vision-based, ultrasound and hybrid systems (combination of two or more). Each type of tracker has different operation conditions and can be used for different type of tasks. The overall goal of each tracker is to provide high accuracy, low latency, low jitter and robustness. These are crucial for a successful AR application, as it depends on these hardware elements performances in calculating position and orientation for virtual objects and for the HMD virtual scene. Sometimes it is needed to track the whole human body motion, even including the fingers. As this type of tracking may involve several types of tracking technique it is not included in the comparison. Also the mechanical types are left out, as these are detailed in the next subsection.

Inertia tracker systems are typically compact systems where rate gyroscopes and accelerometers are
mounted on an electrical circuit. Given the initial position (calibration position), velocity and orientation, the rate gyroscopes measure the angular velocities, from which the mathematical integral gives the angular displacements or orientation. The only problem is that, it can only serve with 3 degree of freedom information and does not react to very slow movement. Acoustic tracker, on the other hand, uses the time of flight of ultrasonic pulses for tracking. It uses the echo scheme or direct transmission and typically requires multiple sensors to triangulate the 3D position of the tracked object. The time of flight is very effective solution in case of no disturbing electrical equipment and no blocking objects. Electromagnetic tracker includes using infrared, visible (optical) and radio waves. Two configurations are possible, outside-looking-in, where the sensors are mounted in the environment tracking emitters on the user and inside-looking-out, where sensors are mounted on the user. The former scheme is typically used for motion capture in the entertainment industry while the latter is used in mobile applications including AR and machine vision in robotics. Magnetic tracker uses magnetic sensors to determine the intensity and direction of the magnetic field. Optical tracker uses one or more cameras and the computer vision technology to detect targets in the images and compute their position and orientation based on this information [3].

The trackers are usually mounted inside the HMD or glued on surfaces of objects or belted to the user. The ultimate solution would be some kind of hybrid system, but until now there is no general solution for tracking. Each AR application requires specific trackers and creates new solutions for this problem.

2.1.3 Haptics

AR systems without feedback are worth nothing. HMDs serve as visualization for the virtual world, but without having the capability of touching and feeling the weight of these virtual objects, there will never be realistic environment. For this purpose haptic devices serve as key elements in the feedback process. These haptics can feedback: force, heat, vibration. The 3D virtual objects will also have weight and it is transferred to the user through some kind of mechanical solution. It is known that our skin hosts four types of tactile receptors (mechanoreceptors). When these receptors are activated they emit small electrical discharges, which are detected by the brain. Two of the mechanoreceptors are slow-adapting and two are fast-adapting. The slow-adapting receptors respond to static force applied on the skin. The fast-adapting ones respond to vibrations and accelerations. Density of the receptors determines the spatial resolution of the skin. The fingertips, where the density of receptors is the highest, can discriminate two contacts much closer than, for example, the palm can. The minimum distance value for the fingertips is 2.5 millimetres apart, for the palm 11 millimetres apart. Contacts points closer than these values are perceived as a single contact point. Spatial resolution is complemented by temporal resolution when two contacts occur on the skin close in time. To be sensed as two successive ones, the contacts must not occur closer in time than 5 milliseconds. Heat sensing is realized by thermoreceptors sensitive to cold and warm. Their spatial resolution is less than the mechanoreceptors [23].

Force feedback is usually achieved with an arm stage, that also records the movement of arm and if needed can apply inverse forces, for slowing or acceleration of the arm. Heat is transferred to the skin with small heat pads, which can cool and warm the pads to a specified temperature. It is quite convenient to use these pads, as it has small size factor and easy integration to human skin. To use vibration (or tactile sensing) as a feedback, is a little bit more challenging task. It is mainly an alternative to force feedback, as it is much easier to wear a vibration glove that have small vibrating motors on each finger, than to connect the arm to a big mechanical stage. The hard part is to teach the brain that specific levels of vibration intensity refer to specific weights. As the weight in a normal environment is detected by different types of receptors, this results in a long learning period.

The usability of the arm stages in mobile AR systems (which are fixed to specific positions) is questionable and there is no mobile, lightweight and commercially available force feedback system, which can be mounted on humans. The only alternative is heat or vibration (tactile) feedback systems. There are already successful experiments, where small vibration motors are used as guidance for navigation [24] or for mapping different colours to specific vibrations [25] and these experiments show that humans can learn basic interactions very fast and efficiently, but the effectiveness of these vibration gloves is not promising [26]. One possible explanation for the failure of the haptics in AR could be that usually AR system developers use a top-down approach in application development: they look for efficient interaction interfaces for the given application rather than matching the physical and mental capabilities of the given interaction (e.g. touch) to the application. Also objects in Virtual Reality are usually too simplified and don’t have enough properties for accurate haptic interaction and the lack of resistance in tactile feedback makes things feel unreal [27].
2.2 Software

As already seen at the hardware level of an AR system, there is no general solution for any kind of problem. Luckily on the software level, generalization is available. However, until now programs are mainly created from scratch and for specific applications. There are some pioneering activities that tried to generalize parts of an AR system. In order to use any kind of tracker, an open-source software library was developed called OpenTracker [28]. SenseGraphics, a haptic interface reseller, created an open-source haptic library, called H3D API [29]. The API is more than a simple haptic interface reader and writer library, it already includes some low level visualization commands.

Also several research groups have developed well-known AR software platforms that allow AR researchers to further develop specific-purpose AR applications. These solutions mainly build on existing open-source libraries (as mentioned above), or creates a brand new solution for the specific problem. The main contributors are the following: ARToolkit [21], Designer’s Augmented Reality Toolkit (DART) [30] and Studierstube [31].

ARToolkit is a marker-based platform and it is the most well-known tool that has been extensively used for AR applications. It is an open-source and free platform containing video tracking libraries, which calculate the real camera position and orientation relative to physical markers in real time. This enables the easy development of a wide range of AR applications [21].

DART is based on Macromedia Director [32] and provides low-level support for the management of trackers, sensors, and cameras. DART allows designers to specify complex relationships between the physical and virtual worlds, and supports 3D animatic actors (informal, sketch-based content) in addition to more polished content. Designers can capture and replay synchronized video and sensor data, allowing them to work off-site and to test specific parts of their experience more effectively [30].

Studierstube is based on the OpenInventor [33] real-time rendering framework and OpenTracker is used to support a wide range of software and hardware, including various HMDs and trackers. This system allows the user to combine multiple approaches - augmented reality, projection displays, and ubiquitous computing - to the interface as needed. The environment is controlled by the Personal Interaction Panel, a two handed, pen-and-pad interface that has versatile uses for interacting with the virtual environment. Studierstube also borrows elements from the desktop, such as multitasking and multi-windowing. The resulting software architecture is a user interface management system for complex augmented reality applications [31].

Beside the standalone applications there are also mobile frameworks. One of these frameworks is the Distributed Wearable Augmented Reality Framework (DWARF). The application is shielded from the low-level services, such as for user interface or tracking hardware, and accesses these at a higher level of abstraction using the various DWARF services. It includes a special service which provides bootstrapping functionality and "glue logic". This provides the other services with models of the world and of tasks the user wishes to perform [34].

One another mobile framework is the Morgan AR/VR Framework. This supports the development and the usage of distributed multi-modal multi-user AR/VR applications and their customization to the individual requirements of the users and application scenarios. It was developed to provide full support for heterogeneous distributed environments and a large variety of individual input and output devices [35].

The above mentioned frameworks fill in the gap between the hardware and software level, making AR application development easy, but these are only visualization tools for creating an AR environment and beside a few initiatives (e.g. Studierstube Personal Interaction Panel [31]) not much focus is given on the interaction (system control) with the AR environment. Techniques which work well in 2D environments (e.g. menus, windows, text fields) mainly fail in direct 3D implementation [36]. The main reason for failure is that the spatial input is more complex (physical constraints, lack of feedback) than using a pointing device in 2D. Such a simple task, as selecting a menu item from a menu list, is not easy in 3D [37].

3 Conclusion

There is a huge potential in Augmented Reality systems, as the already successful demonstrations show, but the everyday use of this type of systems is far away. Within years the display technologies (like OLED) will become so advanced that a HMD will look like normal reading glasses and will have built in batteries and trackers. The communication with the virtual world will be wireless and the environment surrounding the human will carry the intelligence. Joining an AR world will be as easy as putting on glasses and interact with the virtual world with the person’s own hands without any trackers on the body / hands. The technology development in the last 10 years already shows that this is achievable in the near future.

Due to advances in personal computing and rapid growth of high bandwidth internet, virtual worlds will
reach everyone [38]. As already seen in 2003, the virtual world of Second Life has achieved remarkable success. Multinational enterprises and companies have invested millions of US dollars to participate in a virtual galaxy [39]. Until now Second Life is only a 3D application running on normal PCs using keyboard and mouse as interaction interface, with a few initiatives to extend its functionalities with haptics [40]. Next breakthrough could be in Augmented Reality. Already in Japan AR is promoted in different TV shows (e.g. Dennou Coil (Cyber Coil: Circle of Children)). The effect on everyday life is unimaginable: old, disabled people can live without constraints or there is no need to travel, just meet in the AR world, in your own living room. The impact of this type of virtual worlds in education is already under investigation [41], [42].

The first step toward this world is a standardized interface to AR worlds, a general approach to interaction [41], [42]. In your own living room. The impact of this type of virtual worlds in education is already under investigation [41], [42].

Acknowledgements: The research was supported by HUNOROB project (HU0045), a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Hungarian Focal Point.

References:


