Trends in AR towards Surgical Training

Philip Feldman
University of Maryland
1000 Hilltop Rd, Baltimore,
MD 21250
feld1@umbc.edu

ABSTRACT

Augmented Reality (AR) has had a long history that has incorporated the most sophisticated graphics, presentation systems and tracking hardware of the day. As the capabilities of the hardware has changed and improved, so has the understanding of the perceptual issues in AR. This paper explores the history of AR and looks at the emergent field of tangible AR, particularly with respect to difficult tasks such as hands-on surgical simulators.

Author Keywords

Augmented Reality; history; human perception; tbd.

ACM Classification Keywords

I.3.0 [Computer Graphics]: General. H.5.0 [Information Interfaces and Presentation]: General. H.5.2 [Interaction Styles]: General. J.0 [Computer Applications]: General.

INTRODUCTION

Some time in 1967-68, Ivan Sutherland wore helmet adorned with tracking systems and display hardware and for the first time ever, the real world and the digital shared the same space[1]. Only a few years earlier, Gordon Moore realized that the number of transistors that could be crammed onto an integrated circuit was doubling every 24 months[2]. These two points are the headwaters for modern (circa 2013) Augmented Reality.

Beyond the technical and rendering considerations, AR is about placing or co-locating computers "in the world" This idea of embodiment, where the real world is used as the medium for interaction has a long and deep history, which is covered in detail by Dourish[3]. The types of interactions that AR can provide has depended throughout its history on the type of connection the AR system has with the world. In the first, initial efforts those connections to the world were mechanical, literally linkages and encoders. As time progressed and electronics improved, these linkages changed, first to non-mechanical tracking systems and most recently to image analysis. At the same time, the capabilities of the rendering and display systems were evolving. This paper will explore some of the perceptual and cognitive ramifications of these technological changes and attempt to look forward in how these trends may play out in the future. We start this exploration with two of the seminal papers that helped to define the field, and then discuss how increasing computational horsepower has changed the way that issues in AR can be addressed and how that in turn affects the areas of human-computer interaction that AR can be applied to.

Paper: "A head mounted three dimensional display" Sutherland [1]

There are few papers that can be called "visionary", but this is one of them. In this paper, Ivan Sutherland describes the design, construction and evaluation of a display system who's goal was "been to surround the user with displayed three-dimensional information ". To achieve this, hardware and software had to be integrated in unique ways:

- The position and orientation of each eye had to be accurately determined in real time. Initially this was done with mechanical linkages, since they were easier to work with. However, the weight if the entire assembly was uncomfortable, so a second means of tracking using ultrasonic interferometry was developed.
- Correctly aligned imagery had to be the user. In the case of the system described, compact CRTs attached to the headset presented stereoscopic imagery though partially silvered mirrors.
- Graphics had to be delivered to the user with correct perspective, without extraneous information.

Not surprisingly for such a novel system, there were perceptual issues. Even though the imagery was delivered in stereo for full 3-dimensionality, users still had difficulty determining spatial relationships in the wireframe environment. Users also had difficulty in perceiving unfamiliar shapes, for example a molecule of cyclo-hexane. Although dormant for many years, these exact issues would arise when this technology re-entered the field with the label of Augmented Reality in the early '90s.

Paper: "Knowledge-based augmented reality" - Feiner, Macintyre and Seligmann [4]

The Intel 4004, arguably the first CPU on a chip, had a transistor count of 2,300 in 1971[4]. The custom-built hardware used Sutherland almost certainly had substantially less. By the early 1990's when "Knowledge-based augmented reality" was written the Intel 80486 microprocessor had been introduced, with over one million transistors[6].

This increase in capability allowed Feiner, Macintyre and Seligmann to use off-the shelf hardware and software for tracking and rendering. In fact, the only piece of custom hardware that had to be built was a mount for the display device (A Reflection Technologies Private Eye monocular display) so that the imagery could be presented on a partially-silvered mirror that the user could see through.

The goal of this research was to provide assistance with complex tasks that take place in the world, such as assembling a wiring harness, or repairing an engine. In the case of this study, the task was to replace a tray in a laser printer, but the issues for all these "situated" tasks are the same:

1. Can I use the system comfortably? This means that the imagery needs to be aligned, and not suffer from distracting artifacts, such as motion lag, inappropriate focus, and general wearablity.
2. Is the system able to effectively show me what I need to do? Does the system show me hidden parts in a meaningful way? Is there too much information, so that I become confused?
3. What should be shown? Even if an object is occluded by a physical object, how should the (invisible) item that is/are my goal(s) be displayed in the augmented image.
4. What about other modalities? Hearing and touch should be considered as well.

Although discussed in the paper, items 1 and 4 are not the goal of the study. The physical display was somewhat ad-hoc, but was sufficient for developing the knowledge-based software elements, which were the focus of the study. To address these items (2 and 3), Feiner developed KARMA - Knowledge based Augmented Reality for Maintenance Assistance. This was a testbed to the automatic creation of AR information in a way that would complement the user's view of the real world. Using data stored about the printer, KARMA used a rule-based system for determining what illustrations to overlay, and how to render them. Their system distinguished between the design of an illustration (e.g. what and how much to draw) and the style of an illustration (e.g. how to render the graphics) . For example, deign goals might be to show, locate and identify a toner cartridge and paper tray. If there was no inter-object occlusion, then the style would be "visible" and solid lines would be used throughout. If one object were to occlude the other, then the style would compensate by drawing the occluded elements using a dashed line.

The results of the study did show that knowledge-based generation of instruction and graphics could be contextually generated, and provided reasonably effectively to the user through a monocular display. And although many of the recommendations for future work are particularly focused on the algorithmic "mechanics" of drawing the appropriate item in the appropriate location, a particularly prescient statement was of the need for "the development of a formal model of how a user's performance will be affected by different decisions made in designing 3D illustrations, taking into account the purpose for which the illustration is generated". This issue of determining what to render taking into account the user's information needs is still an issue being researched.

1994 - 2007 - AR develops and matures

Over this period, transistor counts on CPUs went just over 1,000,000 in Intel 80486 to over 220 million transistors with the Intel Itanium 2. This improvement in computational power allowed AR to develop from a process that depended on precise tracking of all aspects of the image recording chain to a process that could be accomplished entirely by image processing. By the end of this period, AR was beginning to become a commodity technology, available in many systems at no extra charge.

Following Feiner's paper, AR began to develop in many different directions. Azuma describes this arc in [7 and 8]. He also defines the components necessary for an Augmented Reality system. Rather than defining it in the context of head-mounted systems and as such limiting to particular technologies, he decided that AR systems should have the following three characteristics:

1) Combines real and virtual
2) Interactive in real time
3) Registered in 3-D

Using this as a basis Azuma's 2001 paper[7] describes new AR systems that involve handheld, flat panel LCD's with an integrated camera that behaves as an AR "window" of a real environment with AR overlays. He also describes the use of locked-down and head mounted projectors for adding augmented content to the real environment. As importantly, he summarizes research for how to interact with virtual information. Many of these interactions involve tangible elements, such as pages with registration marks printed on them that AR avatars can be referenced to.

The paper also lays out human factors considerations that need to be considered in every AR technology. These include latency (a major cause of registration errors), depth perception, adaptation to (and re-adjustment from) worn displays, and fatigue/eye strain.

Large scale outdoor systems are developed. The Columbia University Touring Machine is discussed, as well as research that discusses the issues that need to be considered when people in augmented environments interact. One of the more interesting possibilities of AR described is the ability to provide personalized information privately.

The fist successful application of AR - it's use in sports also occurs at this time. Starting with the FoxTrax hockey system[8], many sports ranging from motor racing to football took advantage of this technology during this period and rapidly became ubiquitous. These systems were able to take advantage of the fixed position of the cameras combined with high quality lenses to calculate a screen position for a given object. For some items, such as football's line of scrimmage, the integration is relatively straightforward. For an object that is moving during the shot, such as a puck or race car, the addition of a locating transponder was required.

2007 - Present - Mobile technology dominates

In 2012, an Intel Xeon Westmere, containing 2.5 billion transistors, could be bought commercially for about $1,000 [9]. Accounting for clock speed, that is over six billion times faster than the computer that powered Ivan Sutherland's head-mounted display, and 110,000 times faster than Feiner's hardware. This is what is predicted by Moore's law. However, more interesting as far as this paper is the Apple A7 chip used in the iPhone 5S. It has 1 billion transistors and a clock speed of 1.3 GhZ, making it about 14,000 times faster than Feiner's hardware. In fact, it is 3 times faster than the Intel Itanium, which was capable of fast enough image processing to be able to calculate camera position and orientation in unprepared environments. When combined with integrated GPS, north seekers, accelerometers, cameras and hardware accelerated graphics, the capabilities of a modern smartphone are quite astounding. At this point, AR has been un-tethered by the need to be registered by some tracking system to the environment[11]. If there is enough detail in the scene that a camera can detect, and that camera's relationship with a display is known, then AR, according to the requirements described by Azuma[8], can be achieved. The issues have now become mostly perceptual and cognitive.

Kruijff, Swan and Feiner. [12] in 2010, lay out the perceptual areas that are involved in AR systems, and discuss the consequences of problems that can occur. These areas are:

- Environment. Perceptual issues related to the environment itself, which can result in additional problems caused by the interplay between the environment and the augmentations.
- Capturing. Issues related to digitizing the environment in video see-through systems, and optical and illumination problems in both video see-through and optical see-through systems.
- Augmentation. Issues related to the design, layout, and registration of augmentations.
- Display device. Technical issues associated with the display device.
- User. Issues associated with the user perceiving the content.

They then go into detail about issues with each of the areas. Of particular interest is that the issue of clutter, first pointed out by Feiner's earlier paper [4] is still a significant area of research. On the other hand, and perhaps counterintuitively, the issues of registration and frame rate, particularly with hand-held AR devices seem to be less of an impediment to a user's situational awareness, though it may affect cognition[12].

New issues that derive from new technologies such as tablets are also discussed. For example, tablet AR applications typically have the user holding the tablet lower, but with the camera pointing forward. This means that the image is not really a "window" into a scene as traditional AR would expect. Whether there are cognitive issues with respect to these new configurations is an area that needs research. Also, individual user differences are just beginning to be addressed.

With respect to clutter, Kalkofen, et al.[13] posit the use of interactive Focus plus Context as a possible solution. Using a combination of vision, modeling and filtering techniques, they are able to programmatically determine an uncluttered yet effective "X-Ray" view that provides the item in a scene that should be the user's focus and embed it in an augmented context that is informative without being distracting, particularly with respect to providing simple, clear contextual depth cues. They implement a Magic Lens for X-Ray visualization that considers the hidden structures as "information in the focus of attention", while treating the occluding objects as context. By applying different rendering, and eliminating all non-occluding structures that are not currently context, they are able to present a general purpose algorithm for producing this type of AR imagery.

OPTIONS FOR HANDS-ON MEDICAL TRAINING WITH AUGMENTED REALITY

With this context in mind, let's consider how these advances in Augmented Reality might be used to support a complex, hands-on task such as "open" surgery training.

Simulators have been shown to be effective, including flight systems[14] to minimally invasive surgical trainers[15]. Open surgery is different however, in that the actions of the surgeon is often not mediated at all, or by the simplest of technologies - knives, needles, suturing material, etc. Given that open surgery is still a significant proportion of all surgeries performed[22], the need to increase the efficacy and access to effective forms of training are highly desired.

The ideal open surgical simulator would be close enough to the actual procedure that the student can get maximum training value for the time spent. The tools that would be used would be the same tools as the actual, and the behavior of the patient at the surgical site would be of high fidelity. As shown previously in this paper, the ways to portray this AR environment could vary, possibly using HMDs, large monitors placed between the surgeon and the surgical site, or even projected displays.

Augmented reality has made tremendous progress in the ability to provide detailed, registered imagery to the user in real time, However the ability for users to interact tangibly with augmented materials in the scene started later and has proceeded at a slower pace.

Over the same period, Augmented reality has become a regular feature in surgical studies and in the development of some products. In " Advanced medical displays: A literature review of augmented reality"[16], Sielhorst, et al provide an analysis of approaches ranging from a 1938 mechanical system for registering fluoroscopes to helmet mounted systems and augmented monitors. Significantly, they find that although the registration of the real and virtual image need to be quite precise (within millimeters for brain surgery), the placement of the augmented scene does not have to be aligned with the physical environment. In much the same way that minimally invasive procedures involve physicians looking at screens that are generally not positioned in line with the patient, with AR screens that are used in open surgery, it is sufficient to have the position of the screen to be approximately in the "in line" position.

In most respects, perceptual issues in surgical AR are the same as those discussed in the previous section - lag, clutter, dept perception, etc. Since the OR is a comparatively small, constrained environment, tracking of the cameras, headsets, monitors, etc is straightforward and need not rely on image processing techniques for registration.

A significant issue however, is the maintenance of registration with deformable human tissues.

Baumheimer and Matthias[17] discuss ways that that image guided endoscopic surgery can adapt to navigation through soft tissue. These are :

1. *Intraoperative Imaging*. In this case, additional imaging devices, such as ultrasound are used to provide updated imagery during the procedure. These tools have the benefit of being easy to use, but they suffer from poor image quality and have not been successfully integrated with AR systems.
2. *Image Registration*: This process takes preoperative imagery and "warps" until it fits the tissues as encountered during the procedure. This is known as "Image to Patient Registration" and typically works by taking advantage of fiducial markers either naturally occurring or placed in the patient.
3. *Biomechanical models*. In this case, the behavior of an organ is simulated to determine a likely deformation. This deformation is then used as the basis for image warping or as a better "guess" of a starting point for photogrammetric systems described in (2), above.

We see in these papers and others that the medical community is open to the concept and use of AR in a variety of ways, and that the technology for integrating the motions of soft tissue with a virtual scene is approaching acceptable usability.

Based on the preceding then, for AR to work in open surgery training and simulation, it appears that the main issue to resolve is the simulation of the surgical site and the contextually correct tracking of the tools used in the procedure.

From the perspective of tool use, an AR open surgical simulator can be regarded as an augmented tangible user interface. "Tangible user interfaces couple physical representations (e.g., spatially manipulable physical objects) with digital representations (e.g., graphics and audio), yielding user interfaces that are computationally mediated but generally not identifiable as “computers” per se"[18].

In an open surgical simulation, user would manipulate the simulation using physical devices such as scalpels, drills, and other items. Within the context of training, these tools have information associated with them about how and when they should be used. Additionally, as the tools (and the user directly, as appropriate) interact with the surgical site, the state of the site is affected along with it's physical and informational structure.

To provide a structure that the user can interact with, there are several possibilities. Conventional mannequins such as those already used for some surgical procedure training (e.g. Simulab[19]) could be augmented to provide sufficient tracking that they could be incorporated in a simulation. Such an approach has the advantage of using currently available items and augmenting them so that they appear more lifelike and are capable of more sophisticated behaviors such as visually portraying bruising and bleeding. Disadvantages would include the focus on standardized models, and the additional complexity of performing an essentially destructive procedure (i.e cutting and removing items) in a way that can be undone so that another user can take advantage of the training system.

A second approach could be to use 3D printing to create a "destructible" surgical site that would be integrated with AR visualization. Printing of deformable materials is currently possible[20]. Additional steps would be to print in fiduciary marks into the material so that image registration can occur, and to be able to print models that consist of layers of material with different characteristics that mimic the behavior of human tissue. One significant advantage of such a system would be the use of patient-specific data, since the model is not under the constraints of mass production.

A third approach would be to build dynamic models that can react as if they were being cut and otherwise manipulated. A particularly interesting recent technological development has been the creation of the inFORM system by Follmer and Leithinger[21] this system provides for the real-time manipulation of a two-dimensional surface - essentially a scaled-up pin block. This system combines the dynamic surface with video projection to create a tangible, dynamic augmented surface. On could imagine adapting such a system so that it presents a number of layers to the user, each representing a physiological component of a test patient. As with the previous system, patient-specific data would be relatively straightforward to incorporate.

These approaches are not mutually exclusive. It is not difficult to imagine a mass-produced body that contains dynamic components that are able to move 3D-printed organs in lifelike ways, all overlayed by AR imagery that provides a realistic and effective environment for the student to not only learn but to explore possible alternatives for some of the most important medical cases as they come up in the future.

CONCLUSION

In this paper, we have attempted to show the history and potential applications of Augmented Reality as it has evolved under the influence of Moore's Law. From systems that were hard-pressed to maintain registration of a few components so that a simple view of a molecule or a printer tray could be presented to a user, modern AR systems have the computational capability to provide registered augmented imagery from unprepared imagery. By transfoming from a tracking-based technology to an image processing-based one, AR is now far more flexible. Scenes can now be "understood" by the imaging and rendering systems so that the object(s) of focus can be rendered in an effective context. Further, as the technology progresses further, the ability of AR systems to adapt to deformable objects will allow for the development of new classes of AR, where the ability to interact with a soft, deformable world in a contextually meaningful way is supported. This opens the door to - among other things, sophisticated hands-on applications such as open surgical training and OR support.

REFERENCES

1. Sutherland, Ivan E. "A head-mounted three dimensional display." Proceedings of the December 9-11, 1968, fall joint computer conference, part I. ACM, 1968.
2. Moore, Gordon E. “Cramming More Components onto Integrated Circuits,” Electronics, pp. 114–117, April 19, 1965.
3. Dourish, Paul. Where the action is: the foundations of embodied interaction. The MIT Press, 2004. 103-120
4. Feiner, Steven, Blair Macintyre, and Dorée Seligmann. "Knowledge-based augmented reality." Communications of the ACM 36.7 (1993): 53-62.
5. Wikipedia contributors. "Intel 4004." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 13 Nov. 2013. Web. 16 Nov. 2013.
6. Wikipedia contributors. "Intel 80486." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 12 Nov. 2013. Web. 16 Nov. 2013.
7. Azuma, Ronald T. "A survey of augmented reality." Presence 6.4 (1997): 355-385.
8. Cavallaro, Rick. "The FoxTrax hockey puck tracking system." Computer Graphics and Applications, IEEE 17.2 (1997): 6-12.
9. Wikipedia contributors. "Transistor count." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 14 Nov. 2013. Web. 16 Nov. 2013.
10. Recent advances in augmented reality. Azuma, Ronald, et al. "Computer Graphics and Applications, IEEE 21.6 (2001): 34-47.
11. Byron, Kelly, Linh Pham, and Joshua Situka. "Markerless Motion Tracking." Virtual Reality Applications Center, Iowa State University 9 (2008).
12. Perceptual issues in augmented reality revisited Kruijff, Ernst, J. E. Swan, and Steven Feiner. Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on. IEEE, 2010.
13. Kalkofen, Denis, Erick Mendez, and Dieter Schmalstieg. "Interactive focus and context visualization for augmented reality." Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2007.
14. Jacobs, Robert S., and Stanley N. Roscoe. "Simulator cockpit motion and the transfer of initial flight training." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 19. No. 2. Sage Publications, 1975.
15. Haque, Syed, and Shankar Srinivasan. "A meta-analysis of the training effectiveness of virtual reality surgical simulators." Information Technology in Biomedicine, IEEE Transactions on 10.1 (2006): 51-58.

16. Sielhorst, Tobias, Marco Feuerstein, and Nassir Navab. "Advanced medical displays: A literature review of augmented reality." Display Technology, Journal of 4.4 (2008): 451-467.

17. Baumhauer, Matthias, et al. "Navigation in endoscopic soft tissue surgery: perspectives and limitations." Journal of Endourology 22.4 (2008): 751-766.

18. Ullmer, Brygg, and Hiroshi Ishii. "Emerging frameworks for tangible user interfaces." IBM systems journal 39.3.4 (2000): 915-931.

19. "Simulab Corporation." : Medical Training Models and Surgical Simulators. N.p., n.d. Web. 17 Nov. 2013.

20. Bickel, Bernd, et al. "Design and fabrication of materials with desired deformation behavior." ACM Transactions on Graphics (TOG) 29.4 (2010): 63.

21. Follmer, Sean, et al. "inFORM: dynamic physical affordances and constraints through shape and object actuation." Proceedings of the 26th annual ACM symposium on User interface software and technology. ACM, 2013.

22. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 22 Aug. 2013. Web. 17 Nov. 2013.