Complementary Tracking and Two-Handed Interaction for Remote 3D Medical Consultation with a PDA

Greg Welch*
The University *of* North Carolina *at* Chapel Hill

Michael Noland†
Emergent Game Technologies

Gary Bishop‡
The University *of* North Carolina *at* Chapel Hill

Abstract

We report here on applying a complementary tracking and two-handed interaction paradigm to remote medical consultation using a mobile hand-held device. The basic idea is to track the relative pose between the mobile device and an integrated but detachable "patient surrogate" such as the device cover, enabling 6 DOF visualization of streaming vision-based 3D reconstructions of a remote patient. The two-handed device/target paradigm is complementary in that it enables 6 DOF user interaction with applications on the device, while making the tracking problem tractable.

Keywords: Tracking, Augmented Reality, Human-computer interaction, User interface, Two-handed.

Index Terms: I.3.1 [Computer Graphics]: Hardware Architecture—Input Devices I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1 Introduction

As part of a larger research project exploring the potential of 3D telepresence for remote medical consultation we are working on real-time systems to allow an expert medical advisor (doctor) to aid a remote medical advisee (technician) in the field [25, 24, 23]. The overall aim is to provide the advisor with a 3D sense of presence with the distant advisee in circumstances where time is critical, anxiety is high, and the advisee would welcome an expert consultation for diagnosis or treatment. Such 3D telepresence has been described as the "ultimate development in telemedicine" [16, 17].

An example of the circumstances where such consultation might take place is illustrated in Figure 1. The left image depicts an emergency medical technician in the field with a patient, using a portable multi-camera rig. The middle image depicts doctors at the hospital, viewing a dynamic 3D reconstruction of the patient. The right image depicts a doctor (a critical specialist) who is away from the hospital, viewing the dynamic 3D reconstruction of the patient using a wireless, tracked, hand-held personal digital assistant (PDA).

It is the right image in Figure 1 that is the subject of this paper. Our goal has been to develop or adapt tracking technology and user interface paradigms that would allow a remote doctor to use a PDA as a "magic lens" [5, 6, 10, 13] that simultaneously provides spatial awareness of the remote circumstances, e.g., via motion parallax, and viewpoint flexibility, e.g., to achieve a closer view or a different perspective to help address occlusions of the doctor's view.

2 Prototype System

As shown in Figure 2, our current prototype 3D medical consultation (3DMC) system consists of multiple components that would be associated either with the remote patient or the doctor: a portable camera unit (a), a portable compute cluster (b), and two consultant display device paradigms: a head-tracked stereo and autostereo station (c) and the tracked PDA (d). The latter is also shown in Figure 4. During operation the compute cluster (b) receives images from the portable camera unit (a) at 15 frames per second and uses a GPU-based view-dependent 3D reconstruction [26] to render proper images for the stereo/autostereo displays (c) or the tracked PDA (d). The tracked PDA component consists of a Toshiba e800 PDA; a PDA Server/Tracking Client; a Tracking Server; and a PDA Client. The PDA Server has a complete representation of the reconstructed data and receives the estimated pose of the PDA from the Tracking Server over the network. It then renders a view of the most recent reconstruction from the estimated viewpoint and compresses the view to send to the PDA. (The client/server components may be run on the same or separate machines.) The PDA Client receives compressed images and displays them, as well as relaying user input to the server, such as thumbwheel-controlled field-of-view.

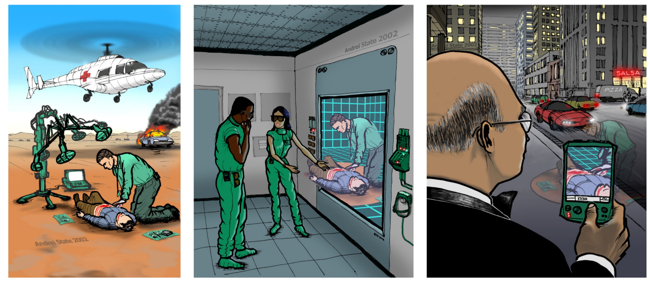

Figure 1: Conceptual sketch of 3D medical consultation. Left: technician in the field with camera rig. Middle: Surgeons in the hospital view dynamic 3D reconstructions. Right: A doctor away from the hospital uses a tracked PDA. (Sketch by Andrei State.)

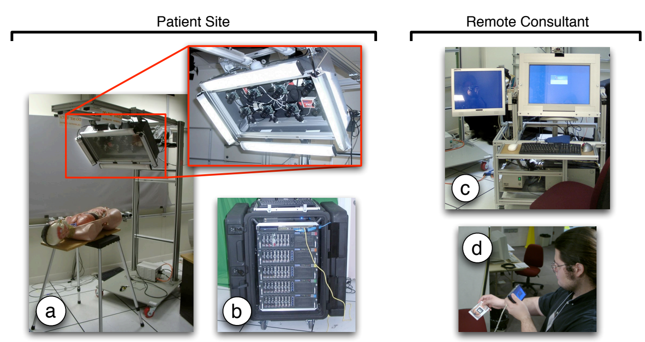

Figure 2: Our proof-of-concept 3DMC prototype, with patient site components on the left and remote consultant components on the right: (a) portable camera unit with eight Firewire cameras and high-frequency area lights; (b) compute cluster; (c) a transportable consultant viewing station with 2D and 3D (head tracked or autostereo) displays; (d) a tracked PDA mobile display.

*e-mail: welch@cs.unc.edu
†e-mail: joat@cs.unc.edu
‡e-mail: gb@cs.unc.edu

3 Exploring Tracked PDA Paradigms

Here we present a brief historical account of the progression of experiments and realizations surrounding our exploration of the tracked PDA paradigm.

3.1 Absolute One-Handed Paradigm

We initially envisioned having to solve the general problem of tracking the PDA with respect to the world. That difficult problem has, and continues to be, the subject of interesting work, generally targeted at outdoor Augmented Reality, where the visualization target is (or is embedded in) the world around the user [22, 2, 9, 15, 27]. The idea was to position the dynamic 3D reconstruction of the patient at some fixed place in front of the doctor, allowing them to physically walk around to view the patient from different perspectives, moving the PDA closer and farther away to effectively zoom in and out.

To explore this scenario we simply attached a HiBall-3000™ tracking system sensor to the Toshiba e800. Using the real-time pose information we streamed view-dependent imagery of a 3D reconstruction of a human patient model to the PDA. While the prototype functioned properly, our exploration brought to light several unforeseen issues. The biggest issue was related to using the "magic lens" to find the remote (virtual) reconstruction in the real world. We would gesture and tell the users where the patient was situated (virtually), and yet they would still spend quite a bit of time sweeping the PDA view around the tracked space, searching for the patient. If the user momentarily moved the PDA away, they would get "lost" and once again have to search the virtual space for the patient. A second issue was related to the ergonomics of the paradigm: the users could not effectively use the system while sitting down, without restricting their views.

3.2 Relative One-Handed Paradigm

To help address the problem of getting lost we decided to mark a place on a table that would be considered the patient, and then track relative to that place. The idea was to give the users a concrete notion of where the patient was. This helped, but still the user's lost track of the patient when they moved around to change viewpoints.

We then realized that the medical consultation scenario inherently involved visualization *centered about the patient*, which reduced the problem to one of tracking relative to a single nearby object, as opposed to the environment (the world). The natural proximity of this orbital paradigm reduces the tracking problem to a more tractable short-range problem.

3.3 Complementary Tracking & Two-Handed Interaction

Finally it occurred to us that we could use the PDA cover (for example) as a surrogate for the patient, and allow the doctor to hold the "patient" in one hand and to use the PDA as the "magic lens." This two-handed approach is complementary in that it provides a natural way to control the viewpoint, while also relaxing the tracking requirements. The tracking would be over a short range, it could rely on the human visual and proprioception systems to mitigate fixed errors/offsets, and the solution could be completely self contained, requiring no additional environmental infrastructure.

The notion of two-handed input has been around for some time, and has been shown to be useful for certain applications [3, 12]. Similar to our circumstances, Hinckley, et al. experimented with using a doll's head or rubber ball and various tools as "props" for neurosurgeons visualizing patient data [8]. While each object was tracked relative to the world, the interaction was handled with respect to the individual objects. Hinckley found that users could easily position their hands relative to one another quickly—a task we all do frequently. The scenario where a doctor is holding a patient in their hand could also be considered a *Worlds in Miniature* technique [14, 18]. Around the same time we began work in this area, Hachet et al. also recognized the complementary benefits and applied it to tangible map exploration [7]. Our application is similar but requires full 6 DOF (position and orientation) tracking to visualize the streaming 3D data from any perspective.

To explore the two-handed paradigm we initially attached one HiBall sensor to the PDA and a second to a PDA-sized metal plate (the surrogate) as shown in Figure 3. While we were still tracking each HiBall sensor relative to the system's ceiling (LED strips), we transformed the PDA-mounted sensor into the coordinate frame of the surrogate sensor before any further processing was done, to simulate relative tracking.

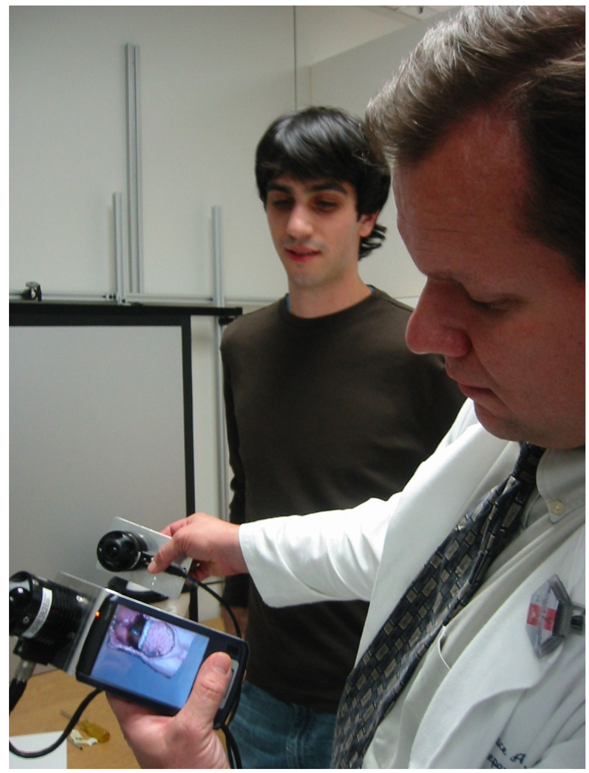

Figure 3: Max Smolens (left) and Dr. Bruce Cairns, M.D. (right) use a HiBall-3000™-based prototype to try the two-handed paradigm. Dr. Cairns holds a tracked PDA in his left hand. The PDA displays a 3D reconstruction of a medical training model. Dr. Cairns holds a tracked "patient surrogate" in his right hand, and his relative right/left-hand pose dynamically controls the rendering viewpoint.

Our initial demonstration to medical collaborator Dr. Bruce Cairns turned out to be very interesting and revealing. Initially he was absolutely certain that he would not want to use the system in the two-handed mode. He said that it would be too confusing, that he would tend to lose the surrogate (the PDA cover), and that instead he would probably simply place the surrogate on a table and walk around it using the PDA as a "magic lens." However after a few minutes of doing this with our prototype, he decided to pick up the surrogate. Then after a few minutes of two-handed manipulation he completely changed his mind and began extoll the virtues of the two-handed paradigm. He particularly mentioned two advantages: he didn't have to exert as much energy to look around, and he could see things in a way he could not in the real world. For example, he could "flip the patient over" to get another perspective, or to see around an occluding object (the remote technicians hands for example). The more he used the two-handed prototype the more enthusiastic he became about the mode of interaction.

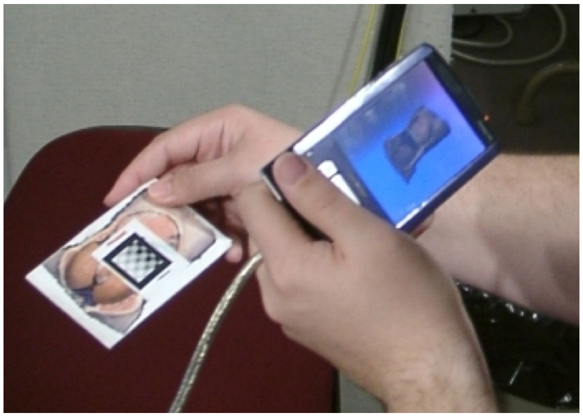

Figure 4: Current Implementation: Surrogate with fiducial marker (left) is tracked using a camera mounted on the PDA (middle).

3.4 Vision-Based Prototype

After using the HiBall tracking system to experiment with the two-handed paradigm, we developed a prototype using a vision-based tracking approach, with a camera on the PDA and a visual fiducial (marker) on the physical patient surrogate (Figure 4).

Our current implementation uses a gray-scale PointGrey DragonFly 640×480 camera mounted on the Toshiba e800 PDA. The camera is connected to the Tracker Server via a Firewire cable, which uses ArToolKit [1] to locate the fiducial and compute the relative pose of the surrogate attached to it. We currently obtain new pose estimates at about 8 to 11 frames a second, limited primarily by the network bandwidth.

The current prototype is not truly portable because of the camera link to a computer (laptop), so we plan on implementing the tracking on a PDA with a built in camera in the future. Wagner and Schmalstieg have ported and optimized ArToolKit for PDAs, and although their early results indicated that the primary bottleneck was frame capture, new PDAs and mobile phones appear to be better suited to video rate capture [20, 21, 7].

4 Discussion

Using a vision-based tracker such as ARToolKit has many advantages, not least of which is the system simplicity, where robust off-the-shelf hardware can be used. Cell phones are now commonly equipped with cameras and are also increasing in computational power and display size, presenting another possible platform for development [10]. However we are concerned about the need for very robust tracking under diverse lighting conditions, something that vision-based approaches in general can have difficulty in providing. For example, ArToolKit uses a fixed binary thresholding, which can work well in conjunction with a camera that self-adjusts brightness and/or shutter speed, but this has its limits. And while ARTag [4] offers significant advantages, the basic paradigm is still passive optical, and thus relies on appropriate lighting and a clear line of sight.

As a result, we are also investigating other tracking mechanisms using an active surrogate. Making the surrogate active (powered) offers opportunities to improve the optical signal-to-noise ratio and reduce computation. The simplest form of active tracking we are considering is to embed flashing LEDs in the surrogate [11], and observe these using a built in PDA camera. This would not require additional sampling hardware attached to the PDA, unlike other potential solutions, but it is still limited to operating at video rates (at best), and requires PDA-based image processing. Another alternative would be to use the same setup on the surrogate, but to equip the PDA with an analog position-sensing device such as a Lateral Effect Photodiode (LEPD). Using a dark-light-dark detection of the change in centroid produced by flashing LEDs, the pose of the surrogate could be computed in a fashion similar to that used in the HiBall. With narrow band filtering on the LEPD, this solution should be much more robust to changes in lighting. It might be possible to implement this using the built-in microphone (analog) input on the PDA for analog-digital conversion, or by adding a card-based AD converter.

While it should be possible to design a "grip" that will reduce the likelihood of users occluding optical fiducial patterns or LEDs, it is still possible a user (consultant) will do so while they are busy trying to help save someones life. As such, another solution we are interested in pursuing is short-range magnetic tracking, which would provide relative immunity to occlusion sensor views. Magnetic tracking has traditionally required bulky, energy hungry field generators and fairly large sensors, but there are now sensors available that fit within blood vessels.[1] Comparatively little work has been done to miniaturize the field generator, but our application requires a relatively small tracked volume and can tolerate reduced accuracy as long as tracking remains robust. Yet another alternative would be to use an acoustic approach such as in [19]. The PDA cover could have a small battery, a pseduo-random signal generator, and four small speakers embedded in it. The PDA could use the built-in microphone to simultaneously acquire the four CDMA (code division multiple access) signals, or separate microphones could be added similar to the LEPD case (above).

Of course making the surrogate active (powered) will lose some of the simplicity associated with a passive vision-based approach. For example, an active surrogate will require power and electronics, introducing issues such as recharging or replacing batteries at regular intervals. The surrogate also becomes harder to replace if lost, compared to a printed fiducial pattern glued to a PDA cover. These are tradeoffs we can better grapple with once we have a concrete working active prototype for comparison.

No matter what the approach to tracking, as mentioned earlier we believe that with a human in the loop controlling the viewpoint, small anomalies will be naturally and unconsciously compensated for: the user will just continue to move the PDA until the display presents the desired view.

5 Acknowledgments

We want to acknowledge collaborators Max Smolens and Herman Towles for various contributions to this work. We also want to acknowledge Prof. Henry Fuchs, Dr. Bruce Cairns, Prof. Ketan Mayer-Patel, and Prof. Diane Sonnenwald, our collaborators on the 3D medical consultation research project.

This research was supported by National Library of Medicine contract N01-LM-3-3514: "3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals."

References

[1] Artoolkit, September 2005.

[2] R. T. Azuma, B. R. Hoff, I. Neely, Howard E., R. Sarfaty, M. J. Daily, G. Bishop, V. Chi, G. Welch, U. Neumann, S. You, R. Nichols, and J. Cannon. Making augmented reality work outdoors requires hybrid tracking. In *First International Workshop on Augmented Reality*, pages 219–224, San Francisco, CA, USA, 1998.

[3] W. Buxton and B. Myers. A study in two-handed input. In *CHI '86: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–326, New York, NY, USA, 1986. ACM Press.

[4] M. Fiala. Artag, a fiducial marker system using digital techniques. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Vol-*

[1] Ascension Technology microBIRD.

ume 2, pages 590–596, Washington, DC, USA, 2005. IEEE Computer Society.

[5] G. W. Fitzmaurice. Situated information spaces ad spatially aware palmtop computers. *Communications of the ACM*, 36(7), July 1993.

[6] G. W. Fitzmaurice and W. Buxton. The chameleon: Spatially aware palmtop computers. In *ACM CHI94*, Boston, MA USA, 1994.

[7] M. Hachet, J. Pouderoux, P. Guitton, and J.-C. Gonzato. Tangimap: a tangible interface for visualization of large documents on handheld computers. In *GI '05: Proceedings of the 2005 conference on Graphics interface*, pages 9–15, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. Canadian Human-Computer Communications Society.

[8] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 452–458, New York, NY, USA, 1994. ACM Press.

[9] 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 & Graphics*, 23(6):779–785, 1999.

[10] M. Mohring, C. Lessig, and O. Bimber. Video see-through ar on consumer cell-phones. In *ISMAR '04: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, pages 252–253, Washington, DC, USA, 2004.

[11] L. Naimark and E. Foxlin. Encoded led system for optical trackers. In *ISMAR '05: Proceedings of the Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 150–153, Washington, DC, USA, 2005. IEEE Computer Society.

[12] R. Owen, G. Kurtenbach, G. Fitzmaurice, T. Baudel, and B. Buxton. When it gets more difficult, use both hands: exploring bimanual curve manipulation. In *GI '05: Proceedings of the 2005 conference on Graphics interface*, pages 17–24, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. Canadian Human-Computer Communications Society.

[13] W. Pasman and C. Woodward. Implementation of an augmented reality system on a pda. In *Proceedings of the 2003 International Symposium of Mixed and Augmented Reality (ISMAR 2003)*, Tokyo, Japan, November 2003.

[14] R. Pausch, T. Burnette, D. Brockway, and M. E. Weiblen. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 399–400, New York, NY, USA, 1995. ACM Press.

[15] G. Reitmayr and T. Drummond. Going out: Robust model-based tracking for outdoor augmented reality. *Proceedings of 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2006)*, October 2006.

[16] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93, Autumn 1993.

[17] J. Steuer. Defining virtual reality: dimensions determining telepresence. pages 33–56, 1995.

[18] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[19] N. Vallidis. *WHISPER: A Spread Spectrum Approach to Occlusion in Acoustic Tracking*. Ph.d., University of North Carolina at Chapel Hill, 2002.

[20] D. Wagner and D. Schmalstieg. First steps towards handheld augmented reality. In *ISWC '03: Proceedings of the 7th IEEE International Symposium on Wearable Computers International Symposium on Wearable Computers*, Washington, DC, USA, October 2003.

[21] D. Wagner and D. Schmalstieg. Handheld augmented reality displays. In G. Welch, editor, *Proceedings of 2nd Emerging Display Technologies Workshop (EDT 2006)*, pages 35–36, 2006.

[22] G. Welch. Hybrid self-tracker: An inertial/optical hybrid three-dimensional tracking system. Technical Report TR95-048, University of North Carolina at Chapel Hill, Department of Computer Science, 1995.

[23] G. Welch, H. Fuchs, B. Cairns, K. Mayer-Patel, D. H. Sonnenwald, R. Yang, A. State, H. Towles, A. Ilie, M. Noland, V. Noel, and H. Yang. Improving, expanding and extending 3d telepresence. In *Proceedings of the 2005 International Workshop on Advanced Information Processing for Ubiquitous Networks*, Christchurch, New Zealand, December 8 2005.

[24] G. Welch, D. Sonnenwald, K. Mayer-Patel, R. Yang, A. State, H. Towles, M. Bruce Cairns, and H. Fuchs. Remote 3D medical consultation. In *Proceedings of BROADNETS: 2nd IEEE/CreateNet International Conference on Broadband Networks*, pages 103–110, Boston, MA, USA, October 2005. Omnipress.

[25] G. Welch, R. Yang, M. Bruce Cairns, H. Towles, A. State, A. Ilie, S. Becker, D. Russo, J. Funaro, D. Sonnenwald, K. Mayer-Patel, B. D. Allen, H. Yang, M. Eugene Freid, A. van Dam, and H. Fuchs. 3d telepresence for off-line surgical training and on-line remote consultation. In S. Tachi, editor, *Proceedings of ICAT CREST Symposium on Telecommunication, Teleimmersion, and Telexistence*, pages 113–152, The University of Tokyo, Tokyo, Japan, December 2004. IOS Press (English) and Ohmsha (Japanese).

[26] R. Yang, M. Pollefeys, H. Yang, and G. Welch. A unified approach to real-time, multi-resolution, multi-baseline 2d view synthesis and 3d depth estimation using commodity graphics hardware. *International Journal of Image and Graphics (IJIG)*, 4(4):627–651, 2004.

[27] S. You, U. Neumann, and R. T. Azuma. Orientation tracking for outdoor augmented reality registration. *IEEE Computer Graphics and Applications*, 19(6):36–42, Nov/Dec 1999.