Position Paper
We Don't Need No Stinkin' Calibration

Jim Vallino
Department of Computer Science
Rochester Institute of Technology
716-475-2991 (voice)
716-475-7100 (fax)
jrv@cs.rit.edu
http://www.cs.rit.edu/~jrv

Early work in augmented reality grew out of virtual reality research domains. These initial augmented reality systems applied the same methods and technology to solve what at first appeared to be a similar problem: correctly render a scene of virtual objects as the user changes viewpoint in the world. This is indeed similar for augmented and virtual reality systems. In virtual reality systems our sense of presence is primarily controlled by the degree to which the visual stimulus presented corresponds to our kinesthetic senses. Augmented reality systems have an additional performance constraint, that being, the correct registration between the user's view of the real scene and the virtual objects augmenting it. It is a more difficult task to maintain a compelling sense of presence when there are discrepancies between these two visual stimuli.

It is no mystery to researchers working in the area that the core problem in augmented reality is this accurate registration of the virtual computer generated images with the user's view of the real scene. This is highlighted in most of the papers in the literature [1-5]. A taxonomy for registration methods used by augmented reality systems described in the literature is given in Table 1. It is based on whether the system uses position sensing to monitor user location in the workspace and whether it requires calibration of the workspace and/or a video camera viewing the scene.

The accuracy of registration is measured by both static and dynamic error components. Static registration error provides a lower bound on accuracy. An augmented reality system will never perform better than its static accuracy. One contribution to static error is the method used for measurement of the user's viewpoint. In classic systems (assuming that the field is mature enough that some systems would be considered classics) viewpoint is measured by position sensing with magnetic sensors, such as the Polhemus sensor. Non-linearities in the response of these sensors, due to the presence of large metal objects in the

Table 1 - Taxonomy for approaches to registration in augmented reality

	Position Sensing	
Camera and/or Scene Calibration	Yes	No
Yes	Janin, et al [6] Feiner, et al [7] Rastogi, et al [8] Tuceryan, et al [2] State, et al [3]	Grimson, et al [9] Mellor [10] Hoff, et al [11] Neumann and Cho [12]
No	State, et al [13]	Uenohara and Kanade [14] Kutulakos and Vallino [15] Vallino [16]

workspace, introduce a static registration error [4, 17]. Systems using a video camera to view the scene require information about the camera's parameters, such as focal length, that are obtained via careful calibration procedures. Any errors in calibration will be reflected in static errors also. The primary contributor to dynamic registration error is latency in the system [4]. This latency comes not only from the time required to perform basic computational steps but also latencies in position measurement and the time to obtain the next frame of video in systems that incorporate cameras.

To achieve correct registration requires determination of the relationships between multiple coordinate systems [2]: world, camera, virtual object. Using magnetic position sensors and metric calibration of cameras these coordinate systems are referenced to a common world coordinate system defined as a standard Euclidean coordinate system. This has the advantage that the reference system is easy to conceptualize and you can take out a tape measure and measure locations in the world. The downside is that you require this metric information to compute the necessary relationships between coordinate systems. A desire to simplify the registration process by eliminating the need for this metric information along with all position measurement and camera calibration motivated much of my thesis work [16] in augmented reality at the University of Rochester. Augmented reality was a natural application of recent work in computer vision research that extracted structure and motion from scenes using uncalibrated cameras [18-20].

The technique of augmenting reality using affine representations relates all the coordinate systems to a common non-Euclidean affine coordinate system. The definition of this coordinate system is obtained at runtime from the projections of four non-coplanar feature points that are tracked through the video sequence.

One of the absolute beauties of this approach is the simplicity of the mathematics. An augmented reality system ultimately needs to compute the projection matrix that the computer graphics camera uses to render the virtual objects. Using affine representations the projection matrix is created directly from the projection of feature points in a video image. Even after working steadily in this environment for several years I am still amazed that it works as well as it does. At this point, one must ask the question: "Can this metric-free method scale up and be viable for the long term such that it warrants continued work?" My answer would be a qualified yes.

This method definitely has its disadvantages. One can argue that the simplicity of computing the projection matrix is offset by the additional requirement of tracking feature points in real time. (Note that any augmented reality technique based on computer vision methods [3, 10-12] has this additional requirement). Tracking of arbitrary targets is still an open problem in computer vision research. If, however, you are willing to engineer your problem with regard to the particular features being tracked, the technology is available for accurate real-time tracking of feature points. In my own work I engineered my experiments to minimize the tracking problem by using color segmentation to track features. Moving to a more natural setting will require additional work on the feature tracking subsystem.

This non-Euclidean method defines coordinates as the linear combination of projections of feature points. Since this coordinate system is only computed at runtime apriori placement of virtual objects can not be performed. Applications operating in an unknown environment will, in general, be required to do runtime placement of virtual objects and may be particularly well suited to this method. The University of Rochester is investigating one such application. It is a military or crime interdiction video surveillance and monitoring scenario, displayed in Figure 1. Here the system generates an augmented view of a scene in a command and control center based on ground and aerial surveillance and detection of common feature points. One can also envision another scenario where aerial surveillance is used to augment a ground-fighter's view of a battle scene.

Approximating the true perspective operation of a video camera with an affine camera [19] introduces errors particularly when the distance from camera to object diminishes. This will cause the system to produce a static registration error. To mitigate this error the use of projective representations [21] should be researched. Affine representations have their place and have been found to provide more accurate reprojection results for large object to camera distances [22]. An adaptive technique that automatically switches to the most appropriate representation, as the camera moves or its lens is zoomed, may yield the best results.

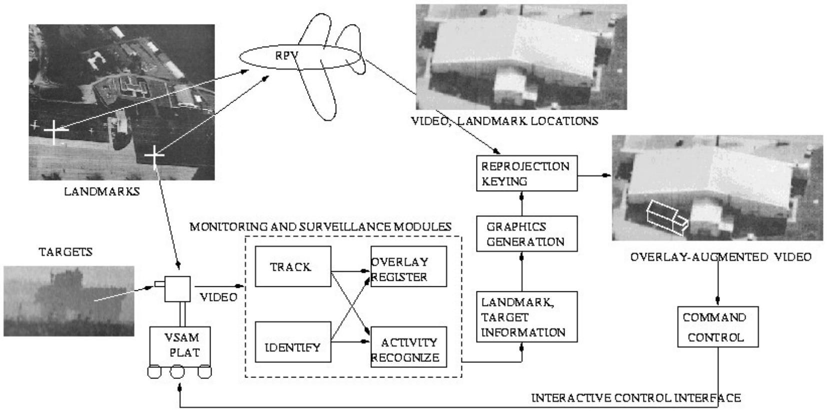

Figure 1 - Video Surveillance and Monitoring Application

It should come as no surprise that an augmented reality system based on affine representations is not immune to latency problems generating dynamic errors. In my system, I measured system latencies on the order of 70 to 90 msec. or 2 to 3 frames of video. Without metric data or position sensing many of the approaches [23] for minimizing errors due to latency can not be directly applied. In my testing, I found that simple filtering and forward prediction applied directly to the projection of the feature points in the image yielded marked improvement in registration. The results of this are shown in Figure 2. The mean Euclidean pixel error between a physical point moving in the scene and it's reprojected virtual point is shown on the Y axis. The X axis is the number of frames of forward prediction applied to the feature point locations assuming that feature motion is modeled by constant velocity in the image. The several plots are for different filtering methods applied to the feature points for noise reduction. Across the board improvements were seen for 2 and 3 frame forward prediction of feature point projections. These results show that methods are available for improving latency in a system using non-metric representations. On the graphics side there are limitations with this method due to the nature of the projection matrices computed. In general, the affine projection matrix will not be orthonormal. For many standard computer graphic techniques, such as lighting computations, an orthonormal system is a requirement. If more photorealistic rendering of the virtual objects is required then additional research must be undertaken to determine the correct method to execute these computer graphic algorithms in a common affine coordinate system.

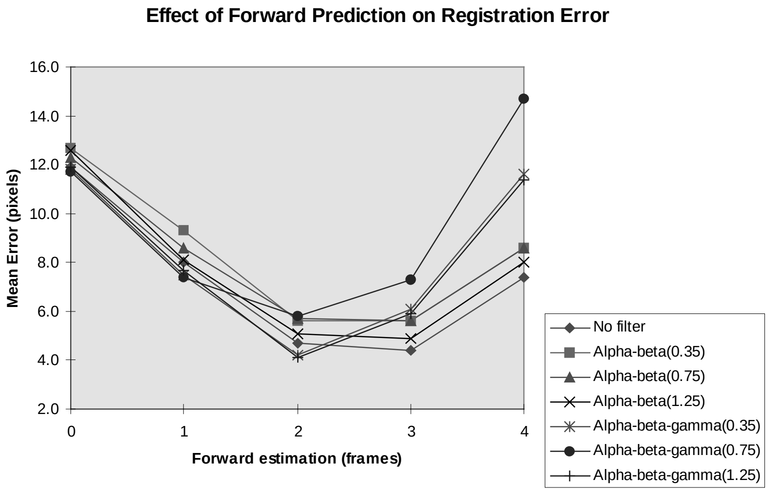

Figure 2 - Dynamic Registration Error Reduction

Affine representations are one end of the spectrum of computer vision based methods for implementing augmented reality systems. Despite its disadvantages it is an attractive method that has potential for certain applications. Applications that will work in an unknown environment where apriori calibration and measurement will not be possible seem particularly well suited to the method. Potential avenues of research activity exist to improve the performance of the method. So far there has been little work [3] that tries to integrate the different methods for performing registration to optimize the overall result. More research activity in that area may yield synergies that are currently unknown.

Bibliography

[1] M. Bajura and U. Neumann, “Dynamic Registration Correction in Video-Based Augmented Reality Systems,” *IEEE Computer Graphics and Applications*, vol. 15, pp. 52-60, 1995.

[2] M. Tuceryan, D. S. Greer, R. T. Whitaker, D. E. Breen, C. Crampton, E. Rose, and K. H. Ahlers, “Calibration Requirements and Procedures for a Monitor-Based Augmented Reality System,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, pp. 255-273, 1995.

[3] A. State, G. Hirota, D. T. Chen, W. F. Garrett, and M. A. Livingston, “Superior augmented reality registration by integrating landmark tracking

and magnetic tracking," presented at Proceedings of the ACM SIGGRAPH Conference on Computer Graphics, 1996.

[4] R. L. Holloway, "Registration Error Analysis for Augmented Reality," *Presence*, vol. 6, pp. 413-432, 1997.

[5] R. Azuma and G. Bishop, "Improving static and dynamic registration in an optical see-through hmd," presented at Proceedings SIGGRAPH '94, 1994.

[6] A. L. Janin, D. W. Mizell, and T. P. Caudell, "Calibration of head-mounted displays for augmented reality applications," presented at Proceedings IEEE Virtual Reality Annual International Symposium '93, Seattle, WA, 1993.

[7] S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-Based Augmented Reality," *Communications of the ACM*, vol. 36, pp. 53-62, 1993.

[8] A. Rastogi, P. Milgram, and J. J. Grodski, "Augmented Telerobotic Control: a visual interface for unstructured environments," : http://vered.rose.utoronto.ca/people /anu_dir/papers/atc/atcDND.html, 1995.

[9] W. E. L. Grimson, G. J. Ettinger, S. J. White, P. L. Gleason, T. Lozano-Perez, W. M. W. III, and R. Kikinis, "Evaluating and Validating an Automated Registration System for Enhanced Reality Visualization in Surgery," presented at Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95, Nice, France, 1995.

[10] J. P. Mellor,"*Enhanced Reality Visualization in a Surgical Environment*," Masters Thesis, AI Lab, Massachusetts Institute of Technology, Cambridge, MA, 1995.

[11] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer Vision-Based Registration Techniques for Augmented Reality," presented at Proceedings SPIE Vol. 2904: Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, Boston, MA, 1996.

[12] U. Neumann and Y. Cho, "A Self-Tracking Augmented Reality System," presented at Proceedings of ACM Symposium on Virtual Reality Software and Technology, 1996.

[13] A. State, D. T. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura, and H. Fuchs, "Case Study: Observing a Volume Rendered Fetus within a Pregnant Patient," presented at Proceedings of the 1994 IEEE Visualization Conference, 1994.

[14] M. Uenohara and T. Kanade, "Vision-Based Object Registration for Real-Time Image Overlay," in *Computer Vision, Virtual Reality and Robotics in Medicine: CVRMed '95*, *Lecture Notes in Computer Science*, N. Ayache, Ed. Berlin: Springer-Verlag, 1995, pp. 14-22.

[15] K. N. Kutulakos and J. R. Vallino, "Calibration-Free Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, pp. 1-20, 1998.

[16] J. R. Vallino,"*Interactive Augmented Reality*," PhD Thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1998.

[17] B. D. Adelstein, E. R. Johnston, and S. R. Ellis, “A Testbed for Characterizing Dynamic Response of Virtual Environment Spatial Sensors,” presented at Proceedings of 5th Annual Symposium on User Interface Software and Technology, Monterey, CA, 1992.

[18] J. J. Koenderink and A. J. van Doorn, “Affine Structure from Motion,” *Journal of the Optical Society of America A*, vol. 8, pp. 377-385, 1991.

[19] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*. Cambridge, MA: The MIT Press, 1992.

[20] D. Weinshall and C. Tomasi, “Linear and Incremental Acquisition of Invariant Shape Models from Image Sequences,” presented at Proceedings 4th IEEE International Conference on Computer Vision, 1993.

[21] O. D. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?,” presented at Proceedings of Second Euopean Conference on Computer Vision, 1992.

[22] C. Wiles and M. Brady, “On the Appropriateness of Camera Models,” presented at Proceedings of the Fourth European Conference on Computer Vision, Cambridge, UK, 1996.

[23] K. Zikan, W. D. Curtis, H. A. Sowizral, and A. L. Janin, “A note on dynamics of human head motions and on predictive filtering of head-set orientations,” presented at Proceedings of SPIE Vol. 2351: Telemanipulator and Telepresence Technologies, 1994.