ArcheoGuide – a mobile augmented reality system for archeological sites - A solution to the tracking problematic -

Didier Stricker*, John Karigiannis†, Vassilios Vlahakis, Patrick Dähne**, Nikos Ioannidis†
*Fraunhofer Institut für Graphische Datenverarbeitung.
**Zentrum für Graphische Datenverarbeitung
†Intracom S.A
E-mail: stricker@igd.fhg.de; nioa@intracom.gr;

Zusammenfassung:
Das ArcheoGuide-System stellt ein mobiles, multimediales Informationssystem dar, das dem Besucher archäologischer Stätten neue Wege der Wissensvermittlung eröffnet. Durch ein mobiles Endgerät werden ihm multimediale Informationen (Bild, Text und Ton) ortsabhängig bereitgestellt. Zusätzlich werden mit Hilfe von Augmented-Reality-Technologien virtuelle Monumente in eine Datenbrille lagerichtig in der Umgebung eingeblendet. Der Besucher blickt so beispielsweise auf eine Ruine, während der Computer ihm die virtuell rekonstruierten Gebäude in ihrer ursprünglichen Pracht präsentiert. Das Projekt ARCHEOGUIDE (Augmented Reality-based Cultural Heritage On-site GUIDE) ist ein europäisches Projekt, das im IST Rahmen gefördert wurde (IST-1999-11306).

Abstract
ARCHEOGUIDE (Augmented Reality-based Cultural Heritage On-site GUIDE) is the acronym of a project, funded by the EU IST framework (IST-1999-11306), and pursued by a consortium of European organizations. The system allows the user/visitor to experience a Virtual Reality world featuring computer-generated 3D reconstructions of ruined sites without isolating him from the "real world". The key feature of the system is the position and orientation tracking technique used in determining the user's viewpoint. This is an image registration technique based on phase correlation. This paper presents the technique in detail and gives practical examples of the application of the technique in the ARCHEOGUIDE system. In particular, it describes how calibrated reference images may be matched to live video by image transformation to the Fourier domain and phase calculation for translation, rotation, and scaling operations.

Keywords: Augmented Reality, Position Tracking.
Project URL / Projekt URL: http://Archeoguide.intranet.gr.

1. Introduction / Einleitung

ARCHEOGUIDE is a multi-user Augmented Reality system for visitors of cultural heritage sites; such sites are very sensitive and any kind of interference must be kept to an absolute minimum. The system provides 2D and 3D navigation assistance to the user with intuitive and familiar user interfaces (from Internet Explorer windows) and also automatically launches audiovisual presentations about the site depending on the user's position and orientation and declared interests. Augmented Reality reconstruction of the most important temples of the site is achieved from selected viewpoints using novel and sophisticated tracking techniques that will be discussed in this paper.

The visitors, in the beginning of their session, provide a user profile indicating their interests and background, and they optionally choose a tour from a set of pre-defined tours. The system guides them through the site, acting as a personal intelligent agent. Meeting the functional requirement of displaying augmented reality reconstruction of the temples and other monuments of the site, depends on the position – orientation tracking component of our system. Correct object registration and occlusion handling of course requires having an adequate position and orientation component and a detailed model of the site's static environment. An accurate model of the site (digital elevation model, orthophotos etc.) was obtained using accurate photogrammetry and site maps.

The rest of this paper is organized as follows: in section 2, we give a brief description of the overall architecture of the system. The extremely complicated problem of position tracking in an outdoors environment as well as the specific technique that was adapted in our system is analyzed in section 3. Subsequently, section 4 outlines the problems of multi-modal user-computer interaction. We summarize our conclusions in section 5.

2. Overview Of System's Architecture

A more detailed description of our system's architecture is contained in [1], in this paper we will provide only an outline of the architecture for the purpose of completeness. Major requirement that our design is expected to fulfill is the support of multiple concurrent users without any serious sacrifices on the response time of the system. The architecture is a *client/server* architecture, where the clients are the wearable computers of the Mobile Units (MUs) equipped with a wireless network connectivity card. A wireless network with a sufficient number of Access Points (AP) provides connectivity to the server who is responsible for updating the contents of the MU's database whenever the user is moving to an area about which there is no content available. Graphically, we depict system architecture in figure 1.

The hardware components of the system include a Site Information Server, the Mobile Units, the Wireless Local Area Network (WLAN) and a Differential Correction GPS reference station, while the software components address the storage, tracking, interaction, distribution and rendering needs of the overall system.

The Mobile Unit software has been written in C++ (for speed) whereas the server components were written in Java (for maximum portability.)

3. Position Tracking

3.1 General Description

In order to integrate the virtual objects into the real environment, i.e. augment the user's view; we need to determine user's exact position and direction of view. There is a large number of tracking technologies today that offer position and orientation tracking with high

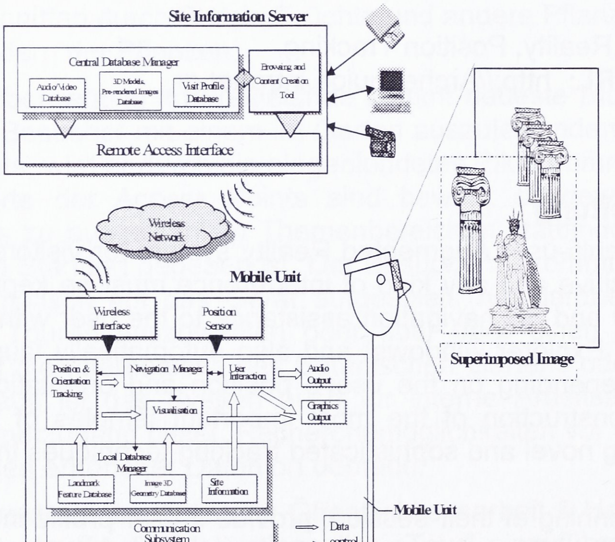

Figure 1 Example of caption.

precision and low latency [2], [3], [4], [9]. However, none of the available systems is suitable for outdoors usage with sufficient precision as required by ARCHEOGUIDE. Integrating different tracking technologies into one hybrid system seems to be the most promising approach [4].

A first rough positioning of the user is given by using the GPS and compass system. For exact tracking we use image-based techniques. The only additional hardware that we need in order to perform the vision-based tracking is a tiny off-the-shelf camera attached to the user's HMD. The system can determine the user's viewpoint based solely on the video image. This brings us to the concept of "*image registration*", something we used in ARCHEOGUIDE and which will be discuss next.

3.2 Describing Image Registration

This method for determining the user's viewpoint assumes that a sufficient number of calibrated reference images are stored in the database, indexed according to the spatial coordinates from which they were taken (successive images overlap to a good degree). The areas from which these reference images are taken comprise the so-called "*selected areas*" or simply "*augmentation view-points*". These areas must be carefully selected in each installation site of the system so as to give a good overall sense of the site (should have open view of most of the site's important areas and cover them well). The method then performs a comparison between the images that the user sees via the camera attached to the HMD, and a number of reference images in the database whose indexes are close to the coordinates provided by the GPS and compass devices. The matching is performed by considering the image as a whole (*global method*) instead of identifying landmarks in each image (*local method*). When the two images being compared overlap in approximately 30-40% or more of the area, the method can correctly compute the *warping* transformation of one image to the other. This (invertible) transformation is then used to provide accurate head-pose estimation since the coordinates of the database image are known. The entire process is shown in figure 2. In fact, in the case when

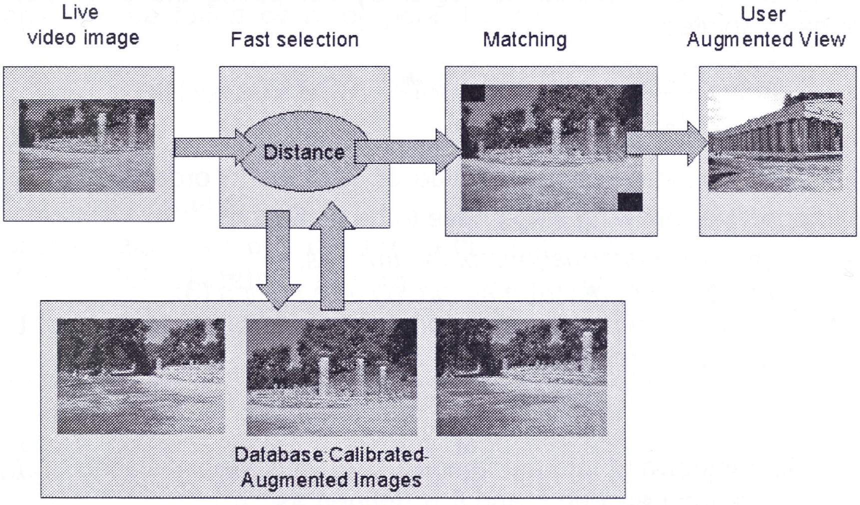

Figure 2 Example of caption.

there are pre-rendered images of the 3D reconstructions of the monuments rendered from the same point of view of the stored site images, the same warping transformation can be used to transform (fast and accurately) the pre-rendered image and then display it on the user's HMD thus achieving 3D visualization as well. Keeping in mind the general description of *Image Registration* process we move one step further in order to see different approaches and analyze the one we employed.

3.2.1 Image Registration Approaches

The approaches for image registration can be classified in three main categories [6], based on the kind of data which is being processed:

1. *Pixel-based*
2. *Feature-based and*
3. *Phase-correlation –based methods.*

The first one compares the raw image pixel data between the two images [7]. Methods of the second class extract image features, like corners, edges or contours [5]. The algorithm, which assigns the extracted features of the two images, depends very much on the class of features which are extracted. The third class of registration algorithms is the phase-correlation one. The input data is transformed in frequency space, where the correlation occurs. This algorithm is well known for its robustness to illumination differences between user and reference image and local features of the images.

3.2.2 Registration with Phase-Correlation

In this sub-section a description of the phase-correlation image registration technique [8] is given which will show how to recover the 2D translation, rotation and the scale factor.

3.2.2.1 Translation

Let f_1 and f_2 be two images differing only in a 2D translation $t(t_x, t_y)$. The images are related as follows:

$$f_2(x,y) = f_1(x - t_x, y - t_y) \qquad (1)$$

The Fourier function F_1 and F_2 of the two images are given by the Fourier shift-theorem:

$$F_2(\xi,\eta) = e^{-j2\pi(\xi t_x + \eta t_y)} F_1(\xi,\eta) \qquad (2)$$

F_2 and F_1 are two arrays of complex numbers. By computing the cross-power-spectrum, the phase of the images is isolated:

$$\frac{F_1(\xi,\eta)F_2^*(\xi,\eta)}{\left|F_1(\xi,\eta)F_2^*(\xi,\eta)\right|} = e^{j2\pi(\xi t_x + h t_y)} \qquad (3)$$

Where: $F_2^*(\xi,\eta)$ is the conjugate-complex value of $F_2(\xi,\eta)$. In order to estimate the translation between the two images the following steps have to be achieved:

1. *Compute the Fourier transformations of the images.*
2. *Compute the cross power spectrum according to equation* (3).
3. *Compute the inverse Fourier transformation of the cross-power-spectrum.* This is a real function which contains an impulse at the coordinates (t_x, t_y).

3.2.2.2 Rotation

If the image $f_1(x,y)$ is transformed into the image $f_2(x,y)$ with the translation $t(t_x, t_y)$ and the rotation with angle φ, then the relation between f_1 and f_2 is defined as:

$$f_2(x,y) = f_1(x\cos\phi_0 + y\sin\phi_0 - t_x, -x\sin\phi_0 + y\cos\phi_0 - t_y) \qquad (4)$$

According to the shift theorem of the Fourier transformation, we obtain:

$$F_2(\xi,\eta) = e^{-j2\pi(\xi t_x + \eta t_y)} F_1(\xi\cos\phi_0 + \eta\sin\phi_0, -\xi\sin\phi_0 + \eta\cos\phi_0) \qquad (5)$$

A rotation in the spatial domain generates a similar rotation in the frequency domain. The magnitude spectra M_1 and M_2 of F_1 and F_2 are related as follows:

$$M_2(\xi,\eta) = M_1(\xi\cos\phi_0 + \eta\sin\phi_0, -\xi\sin\phi_0 + \eta\cos\phi_0) \qquad (6)$$

An adequate way to represent the spectral function is to use a polar coordinate system. A point $P(\xi,\eta)$ in the magnitude spectra is represented by a point $P(r,\varphi)$. Both magnitude spectra in polar coordinates are then defined as:

$$M_2(r,\phi) = M_1(r,\phi - \phi_0) \qquad (7)$$

A rotation is represented as a translation of value φ_0 in the polar-transformed magnitude images. This translation can be easily found by the phase-correlation technique, and thus the rotation angle.

3.2.2.3 Scale

The scaling parameter between two images can be found in a similar way. Let $f_2(x,y)$ the scaled image of the image $f_1(x,y)$ with the factors (α,b), so that:

$$f_2(x,y) = f_1(ax,by) \qquad (8)$$

Then, the Fourier spectra of both images are related as follows:

$$F_2(\xi,\eta) = \frac{1}{|ab|} F_1(\frac{\xi}{a},\frac{\eta}{b}) \qquad (9)$$

If the horizontal and vertical axis of the frequency domain are scaled in a logarithmic way, the scaling parameters can be found as a translation in the frequency domain. This can be written as:

$$F_2(\log\xi,\log\eta) = \frac{1}{|ab|} F_1(\log\xi - \log a, \log\eta - \log b) \quad (10)$$

By applying the phase-correlation technique, the translation *(loga, logb)* can be found and thus the scaling factor *(a,b).*

3.2.2.4 Rotation and Scale

In most of the cases, the horizontal and the vertical scale factors are equal. A rotated and scaled copy of one image can be found by a log-polar transformation of the magnitude images (see equation 7)

$$M_2(\log r,\phi) = M_1(\log r - \log a, \phi - \phi_0) \qquad (11)$$

3.3 Implementation and Results

The Fourier transformation can be computed efficiently with the method of the "Fast Fourier Transformation" (FFT) [10]. Thereby, the image must be square and with dimension 2^n. In our implementation, the left and right borders are cut and the image is scaled down to the next 2^n dimension.

Figure 3 shows an example of registration of the two images (a) and (b). The Fourier transformation of the images is computed and represented by the images (c) and (d) (Power Spectrum). The cross power spectrum is then deducted and transformed back in the space domain. As represented in the 3D representation (e), the peak function is well defined. It gives without ambiguity the component (t_x, t_y) of the image translation t. Finally the two images are added by bilinear interpolation as re-sampling method.

The first results from this approach to tracking run at speeds of 5 frames per second on a low-end laptop PC, and several optimizations are being developed to increase this speed to 10 – 15 frames per second on a high-end laptop. The next two snapshots (figure 4) show the results of the tracking on the site. Images of a video sequence are registered sequentially. The chosen resolution of the camera is 320x240 pixels. The tracking and rendering works at 10Hz on a Toshiba laptop with GeForce graphic card.

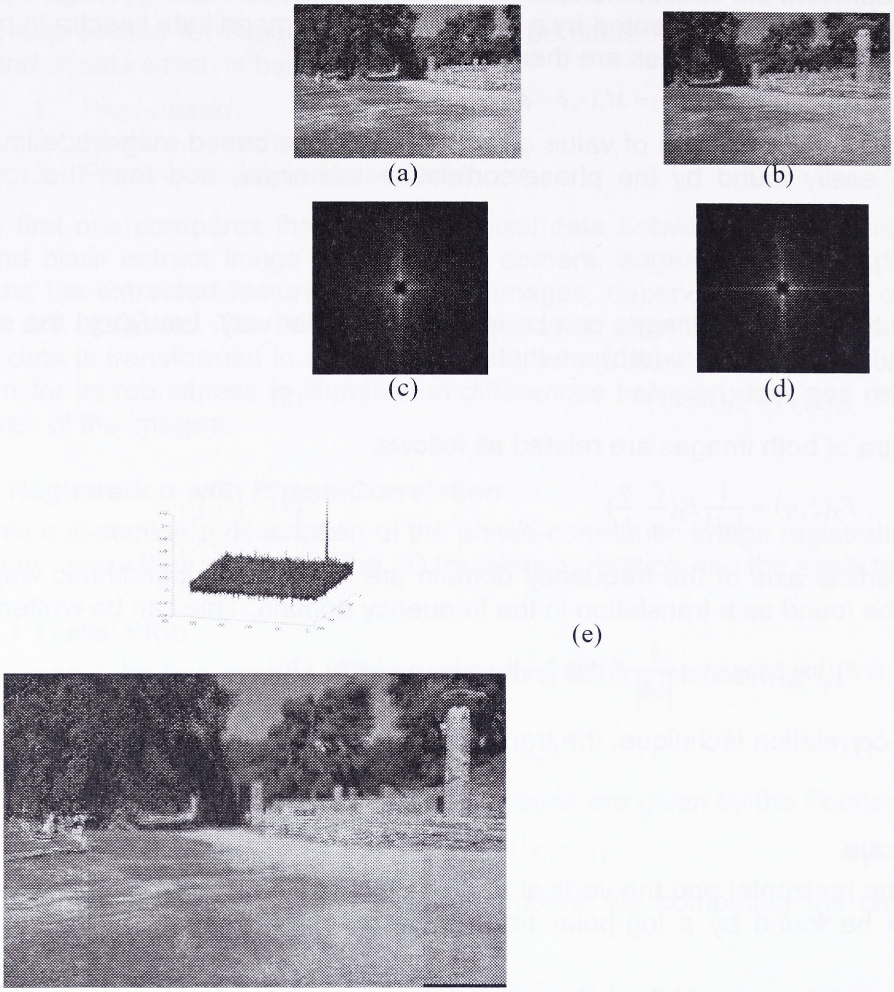

Figure 3 Example of Registration.

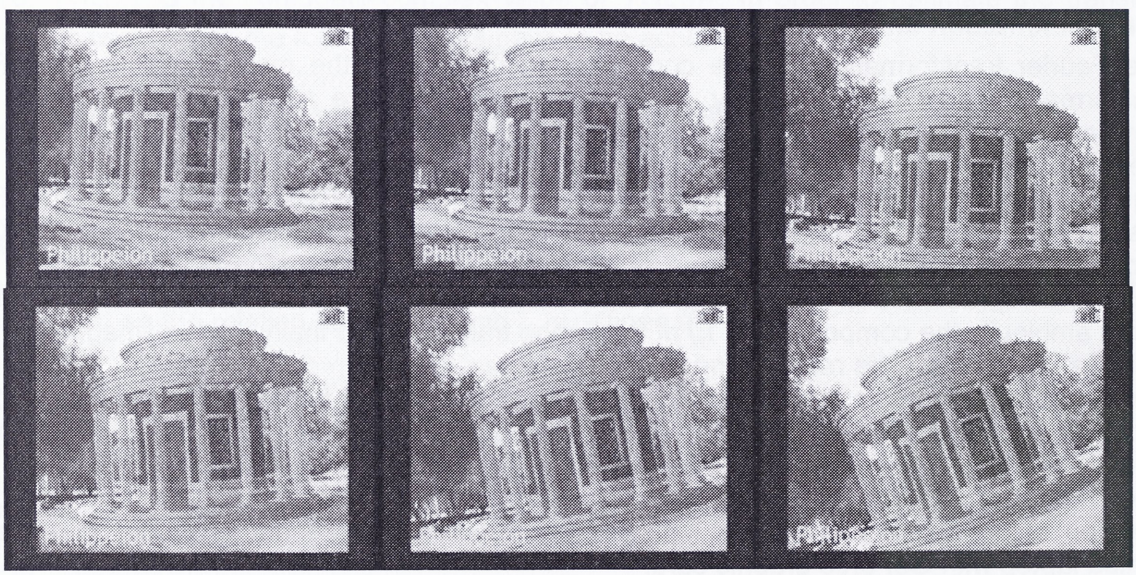

Figure 4 Examples of Real Time Motion Estimation.

So overall, the approach seems to hold the best promise for accurate head-pose estimation in wide outdoors areas. Its combination with image-based rendering techniques is natural and offers the ability to provide photo-realistic Augmented Reality from a number of selected areas in the site that is independent of the scene's complexity. The tracking method is far more robust than landmark recognition based methods.

Whereas the latter methods suffer the danger of loosing tracking due to occlusions of the landmarks from the user's point of view (due to other objects such as humans, animals etc. obstructing the user's field of view), our method is far more robust because it is a global method requiring any overlap in the images of only 30-40%. It also spares the site from interventions of placing artificial landmarks (the number of which increases drastically with the amount of area in which tracking is to be provided.) Further, it is very robust to illumination changes between reference and user images. Its only downside currently seems to be its requirements on storage space, as we must store (and index according to point of view) a number of images for each selected area in which accurate tracking (for augmentation) will be required. The GPS information will help the system in case of calibration and recalibration. In addition to this the GPS and compass will aid the system in predicting the user's path and which monument he will approach next, in order to prepare the data to use next. The method by itself does not require an accurate 3D model database as it works seamlessly with image based rendering techniques. It can be used along restricted paths using "wallpapers" (continuous image textures captured from the real scene.)

The next pictures show actual results from augmenting the area of ancient Olympia (the temple shown is the great temple of Zeus.) The photorealistic nature of the augmented pictures could be achieved in real-time only with pre-rendered views of the high-quality models of the temples that we had available.

4. User Interaction

The interaction of the visitor with the system requires advanced multi-modal interaction techniques since we intend to minimize use of common equipment such as mouse or keyboard. For the first prototype of system however we are developing 3D GUIs based on game-pads or pen-table interactions, on touch screens that change their content as the user moves from area to area. In that way, we treat the user as the mouse cursor on an active image map that is the catopsis of the site itself.

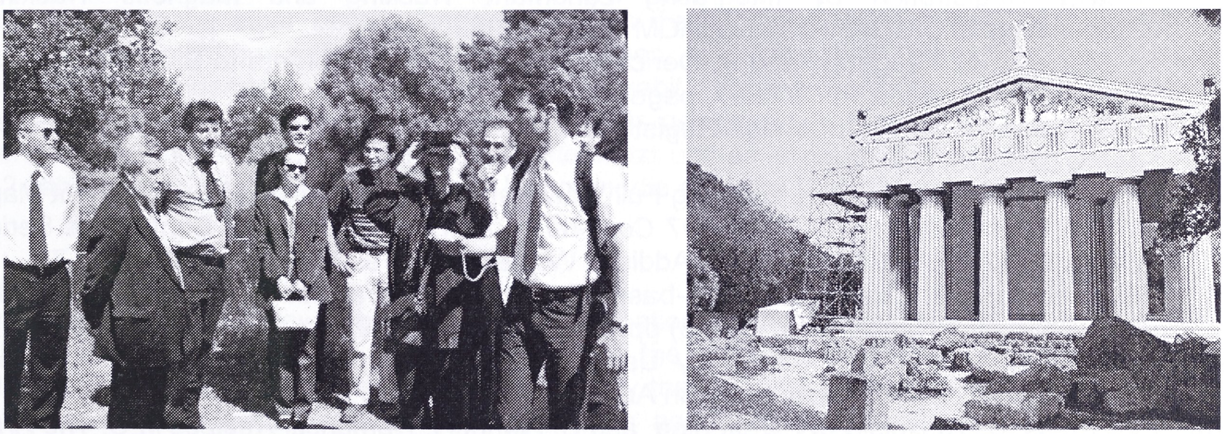

Figure 5 User Tests and Augmented Picture of Ancient Olympia.

5. Conclusions

ARCHEOGUIDE is a research project pursued by a consortium of European organizations and it is funded through the European Union IST framework (IST-1999-11306). The results of the image-based registration techniques described in this paper hold the best promise for the application of Augmented Reality in *sensitive* outdoors areas such as cultural heritage sites where minimizing the intrusion to the site by the system is crucial. The results of the overall approach in user trials performed in Olympia were very encouraging for the realization of a commercial version.

6. Acknowledgents

We would like to acknowledge the EU who selected ARCHEOGUIDE to be financially supported within the IST program under the contract number IST-1999-11306.

7. ARCHEOGUIDE Consortium

The ARCHEOGUIDE consortium is made up of the following partners:
Intracom S.A. (Greece), Fraunhofer Institute of Computer Graphics (IGD) (Germany), the Computer Graphics Center (ZGDV) (Germany), the Centro de Computacao Graphica (CCG) (Portugal), A&C 2000 Srl (Italy), Post Reality S.A. (Greece) and the Hellenic Ministry of Culture (Greece).

8. References

[1] D. Stricker, P. Daehne, F. Seibert, I. T. Christou, R. Carlucci, N. Ioannidis, "Design & Development Issues in ARCHEOGUIDE: An Augmented Reality based Cultural Heritage On-site Guide", Proceedings of the Intl. Conf. on Augmented & Virtual Environments and 3D Imaging" (EuroImage 2001), Myconos, Greece, 2001.

[2] R. Azuma, "A Survey of Augmented Reality", Proceedings of SIGGRAPH 1995, Course Notes #9 (Developing Advanced Virtual Reality Applications), ACM Press, 1995.

[3] R. Holloway, "Registration Errors in Augmented Reality Systems", Ph.D. dissertation, University of North Carolina at Chapel Hill, 1995.

[4] A. State, G. Hirota, D. T. Chen, W. F. Garrett, M. A. Livingston, "Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking", Proceedings of SIGGRAPH 1996, ACM Press, 1996.

[5] I. Zoghiami, O. Faugeras, and R. Deriche. Using Geometric Corners to Build a 2D Mosaic from a Set of Images. In *CVPR97*, pages 420-425, 1997.

[6] L.G. Brown. A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4) pages 325-376, Dec. 1992

[7] R. Szeliski and H-Y. Shum. Creating Full View Panoramic Mosaics and Environment Maps. In T. Whitted, editor. *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 251-258. ACM SIGGRAPH, Addison Wesley, Aug. 1997. ISBN 0-89791-896-7

[8] B. Reddy and B. Chatterji. An FFT-based Technique for Translation, Rotation, and Scale-invariant Image Registration. *IP,* 5(8) pp. 1266-1271, Aug. 1996.

[9] F. Seibert, "Augmented Reality by Using Uncalibrated Optical Tracking", Computers & Graphics 23 (1999), Special Issue on Augmented Reality, pp 801-804.

[10] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK, 2 edition, 1992.