CMNE/CILAMCE 2007
Porto, 13-15 Junho, 2007

AUGMENTED REALITY TO AID CONSTRUCTION MANAGEMENT

Asla Sá[1], Manuel Loaiza Fernández[1], Alberto Raposo[1], Álvaro Maia da Costa[2] and Marcelo Gattass[1]*

1: Tecgraf - Computer Graphics Technology Group
Computer Science Department
Pontifical Catholic University of Rio de Janeiro
R. M. S. Vicente, 225 – Gávea, Rio de Janeiro, RJ - Brazil
e-mail: {asla, manuel, abraposo, mgattass}@tecgraf.puc-rio.br

2: CENPES – Petrobras Research Centre
Ilha do Fundão – Rio de Janeiro, RJ - Brazil
e-mail: amcta@petrobras.com.br

Key words: Augmented Reality, Computer Vision, CAD Models, Construction.

Summary. *Nowadays, before being constructed, most buildings have geometric models defined by their CAD systems. Modern CAD systems are moving from drawing programs to geometric modelers yielding interesting new possibilities. Great benefit can be derived if an Augmented Reality system can yield online images of the construction augmented with the rendering of the CAD models still not constructed. With these possibilities a computer aided construction inspection system can be built providing significant gains of productivity by mainly avoiding construction errors. This paper focuses on Computer Vision challenges and algorithms as part of a bigger effort to make Computer Aided Civil Engineering a more integrated process spanning from design to production.*

1. INTRODUCTION

Nowadays, before being constructed, most buildings and other constructions have geometric models defined by their CAD systems. Modern CAD systems are moving from drawing programs to geometric modelers yielding interesting new possibilities. However, currently the 3D geometric models of the building and the terrain are normally used only for design and marketing purposes. They normally are not used as a daily tool for the construction management.

The engineering geometric models are starting to be used in Virtual Reality (VR) applications for diverse purposes, such as design review, ergonomic studies [1] and safety training [2]. However, VR applications do not provide the user with information about the surrounding real space, being totally independent of the place the user is located.

Great benefit can be obtained with the use of Augmented Reality (AR), which combines the real surrounding environment with virtual objects superimposed over it. The potential of AR in the field of construction management can be characterized by two scales: outdoor (large scale) and indoor (detailed, small scale).

As an example of the outdoor scale, if cameras are installed in the construction site, an AR system can yield online images of the construction augmented with the rendering of the CAD models still not constructed. Furthermore, 3D Photo or 3D scans can be used to generate geometric models of the current constructed parts and these models can be compared with the design models. With these possibilities a computer-aided construction inspection system can be built providing significant gains of productivity by mainly avoiding construction errors.

Considering the indoor scale, engineers, using mobile equipment (e.g., laptops or PDAs), may check internal installations, such as electrical conduits, hydraulic pipes, etc.

This paper discusses computer systems designed to aid construction management. These systems are part of a bigger effort to make Computer Aided Civil Engineering a more integrated process spanning from design to production. This paper focuses mainly in the AR and Computer Vision challenges and the computer algorithms to meet them.

The following section presents an overview of AR technologies. Then, section 3 discusses the main challenges related to the use of AR and Computer Vision in construction applications, and some results are presented. Finally, some conclusions and suggestions of future work are yielded in Section 4.

2. AUGMENTED REALITY

The utilization of Computer Graphics and Geometric Modeling to visualize real data brings fundamental changes in the methodology of work in several different areas, from Medicine to Civil Engineering. The possibility to iteratively visualize and manipulate virtual models permits a better comprehension and analysis of large amounts of data.

The research in this subject is broad and its spectrum includes VR, Augmented Virtuality and AR [3]. In Figure 1 the VR spectrum is illustrated, the extreme points are pure Reality and pure Virtual Reality, in between there is the Mixed Reality where virtual and real elements are mixed up. Near the Reality extreme there is Augmented Reality that is based on the real world with virtual elements inserted on it. Approaching the other extreme the basis changes to a virtual ambient with real elements inserted on it (for example, videoconference inserted in a virtual environment).

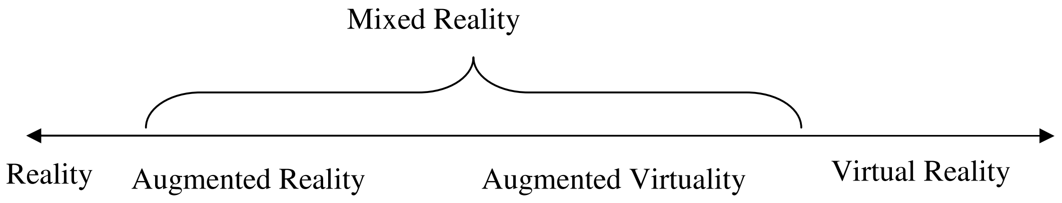

Figure 1 – Virtual Reality Spectrum.

The possibility to compose virtual and real information permits to give information to the user that cannot be obtained by human senses. Possible applications of AR includes, as an example, the inspection of a mechanical component which information is virtually shown composed with the object image in the real world. Also in medicine AR is useful helping in surgery, for example indicating where to make the surgical incision [4]. In this work we are interested in potential applications of AR for Civil Engineering.

In an AR system, the composite of real and virtual images can be made using head mounted displays or video [5]. A challenge is to position the virtual and the real objects in the 3D scene in order to produce a coherent visualization of the mixed reality; this is a hard problem to be solved (called registration problem) [6].

To solve the challenging registration problem, some well-defined steps have to be studied. The recovery of the user initial position relative to the real scene is known as camera (or user position) calibration. The subsequent tracking of its movements to update the camera position in the real world is the tracking phase. The virtual object to be projected onto the scene has to be modeled, and many techniques to recover geometry of real objects can be used, from CAD systems to 3D photography techniques. The final step is the visualization of the real world composite together with the virtual object; if the alignment and registration has been well solved in the previous steps there is the occlusion problem left to be solved in the visualization phase. The above steps will be detailed in the following sections.

A characteristic that influences on the project development decision is related to the real ambient where the augmented reality system intends to be used. Indoors and outdoors environments have fundamental differences related to the possibility of controlling illumination. The possibility to positioning fiducial markers to help the calibration and tracking phases is also a characteristic of the real ambient that influences on the AR system.

The challenge of Engineering is to work with the existing elements in the construction site. Different from medical or museum applications that have a controlled environment, construction sites require robust equipments. Moreover, the placement of markers (fiducials) may also be more complicated given movement of the construction equipment and the changing nature of the scene. With these special conditions the solutions for the use of AR in constructions sites requires innovative solutions.

3. COMPUTER VISION CHALLENGES IN THE CONSTRUCTION AREA

The inputs of AR applications are images taken from cameras positioned by the user. The processing of such images to obtain useful information to the AR application uses several techniques originally developed in the context of Computer Vision. The recovery of camera position, known as camera calibration problem, as well as the feature tracking in a video sequence are common problems in Robotics, and are also the main input processing tasks of an AR application. The use of fiducial markers inserted onto the real scene can be of great help to solve these problems, especially if real time is requested. Once the camera position in real world is known, and the interesting scene features were tracked, the virtual object to be inserted onto the scene has to be coherently positioned related to the scene, occlusion can possibly occur and has to be treated at this step in order to produce convincing output images. In the engineering context the virtual objects consists of architectural elements or engineering plants. In the following subsections we describe the above problems in more detail.

3.1. Camera Calibration

One of the important elements of the visualization process is the virtual camera. The simplest camera model is the one that has no optical system (no lens): it corresponds to the rudimental model of the pinhole camera where light enters through a small hole in a box, projecting itself onto a plane. The produced image is simply a perspective projection of the three-dimensional scene. This simple camera model is the basic model used in computer graphics to define the virtual camera for producing synthetic images. For this reason, applications that involve capture and synthesis of images commonly use this model. The pinhole camera has seven degrees of freedom representing the parameters:

- Position (3 degrees of freedom);
- Orientation (3 degrees of freedom);
- Focal distance (1 degree of freedom).

These parameters determine a projective transformation T which associates to each point **p** in the image plane. In other words, given a point **p**, the virtual camera transformation generates the point **T**(**p**) = **p**’ in the image. We pose an inverse problem related in a natural way with the virtual camera:

1. Given point **p** and **p**’, determine the camera transformation **T** such that **T**(**p**) = **p**’.

It is clear that the knowledge of just one 3D point and its 2D projection is not sufficient to determine the seven camera parameters. A useful application of (1) is: given an image acquired by some input device (video camera or photographic camera), determine the camera parameters (position, orientation, focal length, etc.). This problem is known as camera calibration (Figure 2(a)).

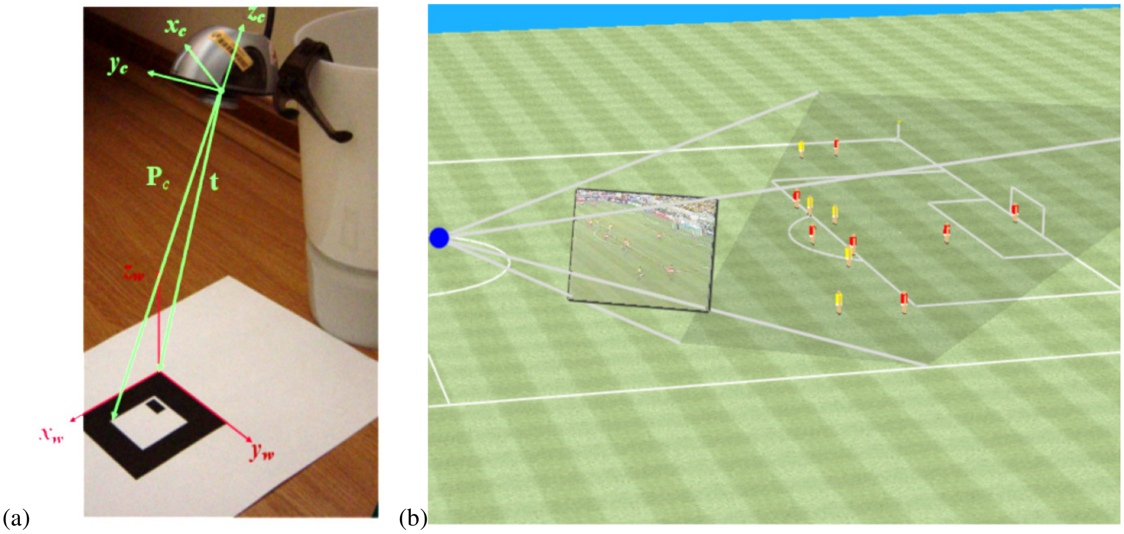

Figure 2 – Illustration of the camera calibration parameters (a) and a virtual camera calibrated according to a real camera in a soccer image (b).

Camera calibration has a large number of applications. Among them we could mention:

- In AR applications, we need to synchronize the parameters of the virtual camera with those of the real camera in order to obtain a correct alignment of the elements in the image composition.
- TV broadcast of sports often uses 3D reconstruction of a given play to help spectators better understand what has happened. In order to do this, it is necessary to know the camera parameters (Figure 2(b)). More information about this type of application can be obtained in [7].

Problems involving the estimation of parameters are naturally posed as a continuous optimization problem, where the objective is to determine the set of parameters that minimize the adjustment error. In camera calibration problem, the parameters to be estimated are the camera intrinsic (optical characteristics of its optical system) and extrinsic (position and orientation) parameters. A good explanation of the calibration problem is found in [8].

Suppose now that we are given the image of a scene and that we need to add to this image some additional synthetic elements. For this, it is necessary to generate a synthetic camera with the same parameter values of the camera that captured the scene. If these parameters are not known, they must be estimated, based on information from the image.

Suppose that there exist *n* points in the image whose spatial coordinates $\mathbf{m}_i = (x_i, y_i, z_i)$ ($i = 1, \ldots n$) are known. By associating these coordinates to the corresponding coordinates (u_i, v_i) on the image, we obtain 2*n* equations whose variables are the camera parameters. Therefore, the computation of the parameters consists in solving this system of equations.

However, in general it is not possible (neither desired) to solve this system exactly. Typically, we use a large number of points in order to calibrate the camera because using a small number of samples leads, in general, to considerable estimation errors. Thus, the system of equations generated by associating points in space to corresponding points in the image has more equations than variables and, in general, does not have a solution. That is, for any set of parameters, there will be discrepancies between the observed image (u_i, v_i) of a point (x_i, y_i, z_i) and the image (u_i', v_i') obtained with these parameters. The set of parameters to be adopted must, then, be chosen in such a way that these discrepancies are as small as possible.

Although general optimization methods can be used to solve the above problem, specific iterative methods have been devised for this purpose. The most popular is Tsai's algorithm [9], which also takes into consideration the presence of radial deformations caused by an optical system. Other important algorithms are presented in [10, 11].

Real cameras are not well approximated by the pinhole model. In real applications lens geometric distortion should be considered in the model. The most common distortion corrected is radial distortion and lens illumination fall off.

3.2. Marker vs. Markerless Tracking

In this subsection we describe the feature tracking in computer vision. There are two main distinct approaches concerning feature tracking in video sequences:

- Tracking based on markers
- Markerless tracking

A marker is a predefined object present in the scene that can be automatically detected by image processing. Since markers started to be used to help in solving computer vision tasks, several distinct types of markers have been proposed to facilitate the tracking task. The main characteristics of a good marker are: 1) it should be easy to detect in the scene; 2) it should be easily distinguishable from other markers present in the scene; 3) it should be robust to detect in case of partial occlusion. Another consideration is about the tracking technology – that influences on the marker design – among them we can cite mechanical, magnetic, acoustic, inertial and optical devices.

We will focus on application that uses optical devices and, consequently, when describing tracking based on markers we also focus our attention on optical markers.

When using optical devices, an approach used to obtain information useful to calibrate the camera position is to extract special patterns from the scene that are known in real world. Here the use of markers splits the approaches into two classes: one uses synthetic markers completely defined by user, meaning that its real dimensions are also pre-defined by user [12] and the other look for geometric characteristics naturally present on the scene that can be detected and tracked in subsequent images [13].

Tracking Based on Fiducial Markers

A fiducial marker can be defined as a pattern inserted into the scene by the user that satisfies as well as possible the 3 characteristics of being a good marker, that is, it is easily detectable and distinguishable from other markers, and is robustly detected in case of partial occlusion.

There are two main groups of fiducial markers: based on points and planar. Markers based on points have a predefined number of points to be detected that are used as input to solve the correspondence between real world and the image plane (Figure 3). A planar marker permits the detection of more points that are useful to guarantee the robustness in the pattern detection as well as in the camera calibration process.

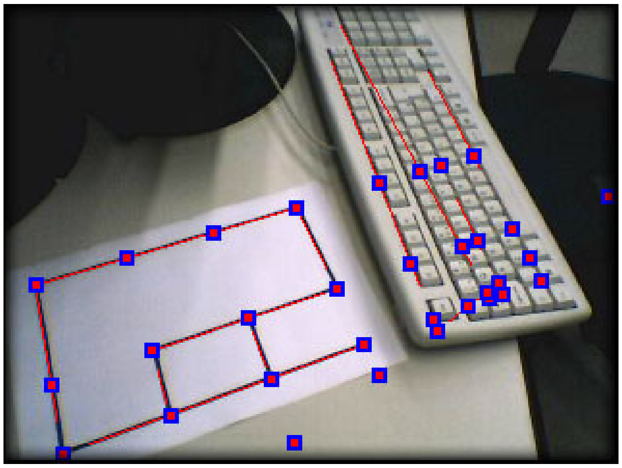

Figure 3 – Example of markers based on lines.

The use of planar fiducial markers is facilitated by the use of the software ARToolkit [14]. This software is able to detect planar markers defined by the user and inserted into the scene, based on the resultant pattern detection it is possible to calibrate the camera position relative to the pattern position. In addition, if the real marker dimensions are known, the scale ambiguity can be solved, and a virtual camera can be positioned consistently relative to the real coordinates allowing the insertion of virtual objects into the scene. An AR application to Manufacturing Planning can be found in [15]; other applications are in [16].

In the construction scenario, tracking based on markers may be suited to internal environments, for instance to check installation details. However, in the scale of buildings, it is not useful, since it is not possible to include optical markers around a construction. Therefore, markerless tracking is an essential technology in this context.

Markerless feature extraction

The second approach that we mention is to look for features naturally inserted into the scene of interest, that are good candidates to be used as markers. These features can be tracked in the image sequence and will be used to infer the camera position relative to the scene. These scene features can be object silhouettes based on basic line segments [17, 18, 19, 20] as well as vanishing points [10, 21].

The main reasoning is to take advantage of well defined fixed elements into the scene that are easy to be tracked. This approach is hardly based on image processing for computer vision purposes like line detection and edge detection algorithms, as well as on image segmentation techniques (Figure 4). Camera photometric calibration is also of great use.

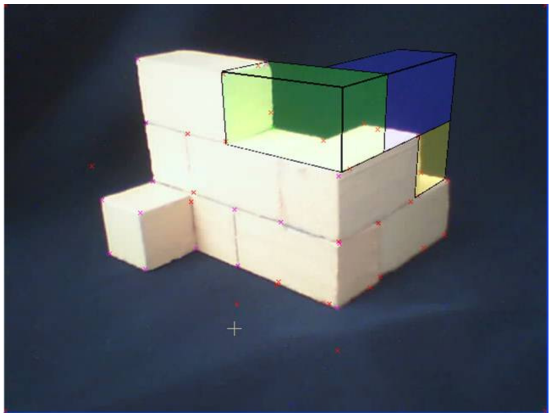

Figure 4 – AR image based on markerless feature extraction (in this case, the extracted features are the blocks corners).

The tracking based on features of the image is adequate for scenarios like a building. However, in a construction site, the situation can be even more challenging, since the scene is very dynamic. The image from where the features are extracted is very “volatile”, since floors or walls may appear in a daily basis, changing the scene objects. The solution goes in the direction of adaptable tracking solutions, like the one presented in [22].

3.3. Recovering scene geometry

The most common way to obtain a geometric model of an architectural building is by using CAD systems. The construction of CAD models of existing buildings is usually a laborious task.

An alternative way to reconstruct a 3D model of a real building is by adopting a computer vision approach, that is, to collect images of a real scene and automatically infer the scene geometry from the given images. If the scene to be reconstructed consists of an existing architectural building, then characteristics such as parallelism and orthogonality naturally present in most of the buildings can be explored to help the model reconstruction task. Application fields such as architectural restoration and town planning as well as studies of sound and electromagnetic wave propagation can greatly benefit from the solutions to 3D building from photographs.

The two main approaches used to recover 3D models of existing buildings are conceptually based on distinct reasoning. In the *Geometry-Based* approach the user uses a 3D modeling tool to recover the model. The alternative approach comes from Computer Vision techniques based on stereo vision and epipolar geometry, it is the *Image-Based* approach where images of the desired building are automatically (or semi-automatically) analyzed in order to find the building blocks. Once the model has been obtained, the original photographs can be used to re-project textures onto the model producing a result that is visually very similar to reality [23, 24].

A good survey on the subject is given at [25]. To fix the idea, we recall that, when geometry information is to be extracted from images, there is the need to calibrate the camera position relative to the scene from references present in the image. In this case, the techniques described in section 3.1 and 3.2 are useful.

3.4. Visualization

A difficulty encountered in the use of CAD models in real-time visualization, as is the case of AR applications is the fact that the engineering models are not constructed for this kind of visualization. In some cases, the models are visually simplified representations, serving only as schematic representations of the characteristics to be analyzed. In other cases the models are too large and complex to be visualized in real time. Currently, it is still necessary a chain of adaptation steps (usually manual) to adequately convert CAD models to real-time models [26].

Occlusion and image registration are typical problems in any AR application. Image registration is the alignment between the real and virtual objects, essential to provide realism to AR images. The occlusion deals with the problem of real objects in front of virtual ones. If occlusion is not considered, the AR scene understanding may be severely compromised.

The occlusion of virtual objects by real ones can be treated using “ghost” virtual objects, similar to the real ones. These “ghost” objects are not rendered, but used in the depth buffer in order to provide the occlusion of the visualized virtual objects behind the real ones. In Figure 5, there are 8 real objects represented by “ghost” objects (a, b), which are visualized in wireframe on top of the real ones (c). Figure 5 (d) shows the composition of real and additional virtual objects.

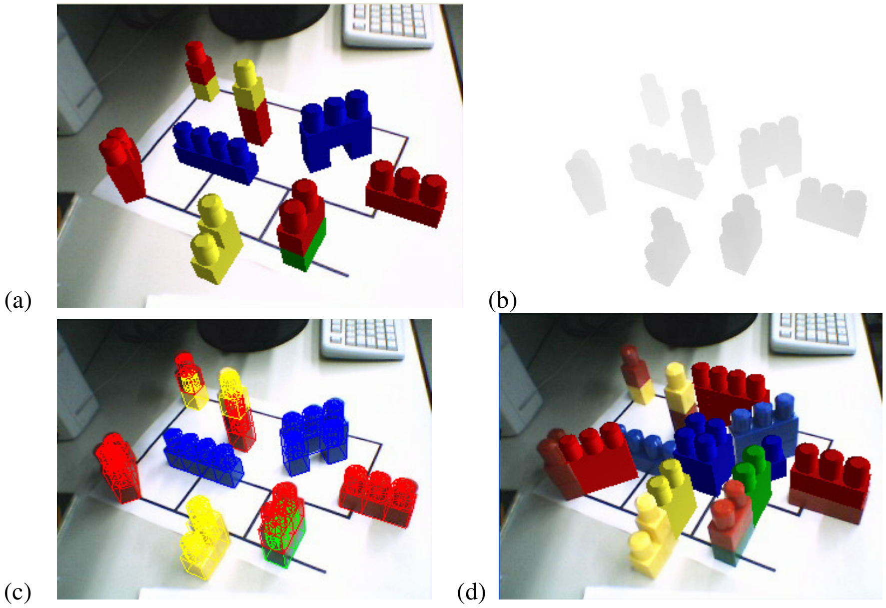

Figure 5 - “Ghost” virtual objects drawn in the color buffer (a) and in the depth buffer (b). The same objects are drawn in wireframe on top of the image of the real objects (c). Image of the real objects composed with additional partially occluded virtual objects (d).

4. CONCLUSION

The idea of using VR and AR in Civil Engineering and Architecture is not new. Since the beginning, these areas have been considered one of the potential applications for VR and AR. However, until recently, VR has been used just used for marketing purposes.

Currently, some projects are making use of AR and VR technologies for construction. Some visible results are been achieved in cultural heritage applications: Archeoguide [27] offers personalized augmented reality tours of an archaeological site; a prototype of Archeoguide was installed for test at *Greece’s Olympia archaeological site*. In Archeoguide the user carries a mobile unit (that can be a laptop, a pen tablet or a palmtop) capable of informing the user's approximate position within the site to a server unit. Based on the approximate initial position, an optical markerless tracking is used to compare the user's view to a set of reference images from the site and infer the 2D transform necessary to re-project augmented reality information from the user's point of view.

Another application related to cultural heritage is the Parthenon project [28]. In this project the geometric model is obtained using 3D scanners, while texture is recovered by taking high quality photographs of the site that are processed to produce high quality textures that are independent of the original illumination condition. The Parthenon project aims a high quality model recovery and an impressive resultant movie is available in Internet, where the Parthenon site is shown completed with many sculptures that are physically distributed in several museums around the world.

Nevertheless, these projects are developed in controlled and steady environments. An ongoing construction is even more challenging in terms of using AR and VR, due to its highly dynamic nature. This paper highlighted some of these challenges, indicating that they can be solved with modern Computer Vision techniques adapted to the specific characteristics of a construction environment.

ACKNOWLEDGEMENTS

This work was developed in Tecgraf/PUC-Rio that is a Laboratory mainly funded by PETROBRAS.

REFERENCES

[1] G. Monacelli, F. Sessa and A. Milite, "An Integrated Approach to Evaluate Engineering Simulations and Ergonomics Aspects of a New Vehicle in a Virtual Environment: Physical and Virtual Correlation Methods", *FISITA 2004 30th World Automotive Congress, Reference F2004F406*, Barcelona, Spain(2004), pp. 23-27.

[2] M. Haller, R. Holm, J. Volkert and R.A. Wagner, "VR based safety training in a petroleum refinery". *In The 20th Annual Conf. of the European Association of CG*, Eurographics, Milano, Italy(1999).

[3] P. Milgram and F.A.. Kishino, "Taxonomy of Mixed Reality Visual Displays", *Institute of Electronics, Information, and Communication Engineers Trans. Information and Systems (IECE special issue on networked reality)*, Vol. **E77-D**, pp. 1321-1329, 1994.

[4] W. R. Sherman and A. B. Craig, Understanding Virtual Reality: Interface, Application, and Design, San Francisco, *CA: Morgan Kaufmann Publishers* , 2003.

[5] J. Rolland, R. Holloway and H. Fuchs, "A Comparison of Optical and Video See-Through Head-Mounted Displays", *SPIE Proceedings: Telemanipulator and Telepresence Technologies*, Vol. **2351**, pp. 293-307, 1994.

[6] R. T. Azuma, "A Survey of Augmented Reality", *Presence: Teleoperators and Virtual Environments*, pp. 355-385, 1997.

[7] Juiz Virtual, http://www.visgraf.impa.br/juizvirtual/

[8] L. Velho and P. C. Carvalho, Mathematical Optimization in Graphics and Vision, *Siggraph Course Notes* (2003)

[9] R. Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. *3*, pp.

323-344, 1987.

[10] R. Cipolla, T. Drummond and D. Robertson. “Camera calibration from vanishing points in images of architectural scenes”. *In Proc. 11th British Machine Vision Conference*, 1999.

[11] Z. Zhang,“ A flexible new technique for camera calibration”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. **22**, pp. 1330-1334, 2000.

[12] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski, “ARQuake: An Outdoor/Indoor Augmented Reality First Person Application”. *In 4th International Symposium on Wearable Computers*, Atlanta(2000), pp 139-146.

[13] A. R. Dick, P. H. S. Torr and R. Cipolla. “Modelling and Interpretation of Architecture from Several Images”. *International Journal of Computer Vision*, Vol. **60**, pp. 111-134, (2004).

[14] H. Kato and M. Billinghurst, “Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System”. *In Proceedings of the 2nd International Workshop on Augmented Reality*, San Francisco, USA(1999), pp 85-94.

[15] F. Doil, W. Schreiber, T. Alt and C. Patron, “Augmented. Reality for Manufacturing Planning”. *In Proceedings of the. Workshop on Virtual Environments*. ACM Press. Zurich, Switzerland(2003). Pp. 71 - 76.

[16] J. M. Salles, A. J. Capó, J. Carreras, R. Galli and M. Gamito, “A4D: Augmented Reality 4D System for Architecture and Building Construction”, *Conference on Construction Applications of Virtual Reality*, Virginia, USA(2003).

[17] C. Baillard, C. Schmid, A. Zisserman and A.W. Fitzgibbon, “Automatic line matching and 3d reconstruction of buildings from multiple views”. *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, IAPRS (1999). Vol.**32**, pp. 69-80.

[18] A.R. Dick, P. Torr and R. Cipolla, “Automatic 3d modelling of architecture”. *In Proc. 11th British Machine Vision Conference (BMVC)*, Bristol(2000), pp. 372-381.

[19] A. Bartoli, and P. Sturm, “Structure From Motion Using Lines: Representation, Triangulation and Bundle Adjustment”, *Computer Vision and Image Understanding*.(2005) Vol. **100**, pp. 416-441.

[20] F. Szenberg, “Acompanhamento de Cenas com Calibração Automática de Câmeras”, *Tese apresentada ao Departamento de Informática da PUC-Rio como parte dos requisitos para a obtenção do título de Doutor em Ciências em Informática*, 2001.

[21] L. Grammatikopoulos, G. Karras and E. Petsa, “Camera calibration combining images with two vanishing points”. *Int. Arch. of Photogrammetry, Remote Sensing and Spatial Information Sciences*, (2004), pp. 99-104.

[22] G. Bleser, H. Wuest and D. Stricker, Online camera pose estimation inpartially known and dynamic scenes, *IEEE/ACM International Symposiumon Mixed and Augmented*

Reality (ISMAR), pp. 56-65 USA, CA, Santa Barbara, (2006).

[23] R. Cipolla, D. Robertson and E. Boyer, Photobuilder – 3D Models of Architectural Scenes from Uncalibrated Images, *IEEE International Conference on Multimedia Computing and Systems,* Firenze, vol. I, pp 25-31, 1999.

[24] E. P Debevec, Modeling and Rendering Architecture from Photographs. *PhD Thesis, University of California* (1996).

[25] G. M. Farinella and G. Mattiolo, Solutions to 3D Building Reconstruction from Photographs, *Eurographics Italian Chapter Conference,* 2006.

[26] J. Berta, Integrating VR and CAD, *IEEE Computer Graphics and Applications*, Vol. **19**, pp. 14–19, 1999.

[27] V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, D. Stricker, T. Gleue, P. Dähne, L. Almeida: Archeoguide: An Augmented Reality Guide for Archaeological Sites. *IEEE Computer Graphics and Applications 22(5): 52-60* (2002)

[28] J. Stumpfel, C. Tchou, N. Yun, P. Martinez, T. Hawkins, A. Jones, B. Emerson and P. Debevec, Digital Reunification of the Parthenon and its Sculptures, *4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, Brighton, UK, (2003).