HAL

archives-ouvertes.fr

Mobile Augmented Reality: Applications and Spe-cific Technical Issues

Nehla Ghouaiel, Jean-Marc Cieutat, Jean-Pierre Jessel

▶ To cite this version:

HAL Id: hal-01006627

https://hal.archives-ouvertes.fr/hal-01006627

Submitted on 16 Jun 2014

Mobile Augmented Reality: Applications and Specific Technical Issues

Nehla Ghouaiel[1], Jean-Marc Cieutat[2], and Jean-Pierre Jessel[3]

Abstract. Although human's sedentary nature over time, his wish to travel the world remains as strong as ever. This paper discusses how imagery and Augmented Reality (AR) techniques can be of great help not only when discovering a new urban environment but also when observing the evolution of the natural environment. The study is applied on Smartphone which is currently our most familiar device. Smart phone is utilized in our daily lives because it is low weight, ease of communications, and other valuable applications. In this chapter, we discuss technical issues of augmented reality especially with building recognition. Our building recognition method is based on an efficient hybrid approach, which combines the potentials of Speeded Up Robust Features (SURF) features points and lines. Our method relies on Approximate Nearest Neighbors Search approach (ANNS). Although ANNS approaches are high speed, they are less accurate than linear algorithms. To assure an optimal trade-off between speed and accuracy, the proposed method performs a filtering step on the top of the ANNS. Finally, our method calls Hausdorff measure [15] with line models.

Keywords: Mobile Augmented Reality, Building Recognition, Machine Vision

1 Introduction

Augmented reality was first used in 1992 by T. Caudell, and D. Mizell to name the overlaying of computerized information on the real world. Subsequently, the expression was used by P. Milgram, and F. Kishino in their seminal paper "Taxonomy of Mixed Reality Visual Displays" [13]. In this paper, they describe a continuum between the real world and the virtual world (nicknamed mixed reality) where augmented reality evolves close to the real world whereas augmented virtuality evolves close to the virtual world (see figure.1).

[1] Nehla Ghouaiel, F-64210 Bidart, France, `n.ghouaiel@estia.fr`
[2] Jean-Marc Cieutat, F-64210 Bidart, France, `j.cieutat@estia.fr`
[3] Jean-Pierre Jessel, F-31062 Toulouse, France, `jean-pierre.jessel@irit.fr`

In 1997 Ronald Azuma developed a complementary definition which he completed in 2001 [14] and which, along with Milgram & Kishino's approach, gave two commonly admitted definitions of augmented reality. According to Azuma, an augmented reality definition is one which complements the real world with (computer generated) virtual objects so they seem to coexist in the same space as the real world. He defined the features of an augmented reality system in both cases according to the following three properties:

1. Combining real and virtual: In the 3D real world 3D entities must also be integrated.
2. Real time interactivity: This namely excludes films even if the previous condition is respected.
3. 3D repositioning: This enables virtual entities to be made to visually coincide with reality.
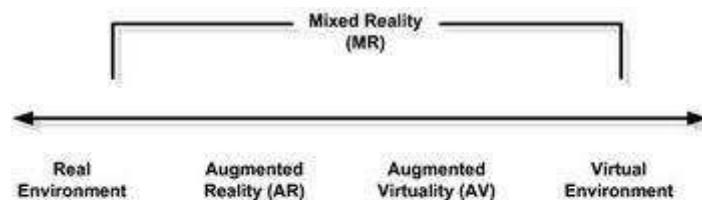
Fig. 1. Virtuality Continum

Displaying augmentations can be done with direct or indirect vision by inducing an additional mental load. In direct vision case, the display uses metaphors such as mirrors; smartphones open like windows onto the environment, vision through glasses or windows, etc.

This chapter is divided to three main parts. The first part presents our proper definition of augmented reality, then, we define the general technical types of augmented reality systems and mobile augmented reality systems. The second part summarizes our contribution on mobile augmented reality. It includes a sensor-based graphic application for urban navigation and a virtual human based augmented reality application. The third part details our hybrid method for building recognition. It combines the potentials of Speeded Up Robust Features (SURF) points and features lines. Our method relies on Approximate Nearest Neighbors Search (ANNS) approach, described by Muja et al. [11]. ANN's methods are known for their speed but they are less accurate than linear algorithms. To assure an optimal trade-off between speed and accuracy, the proposed method performs a filtering step on the top of the ANNS. Finally, our method calls Hausdorff measure with line models [15].

2 Augmented Reality

2.1 Our Definition of Augmented Reality

In [6] ,[17], we proposed a general definition of augmented reality as being the combination of physical and digital spaces in semantically linked contexts especially for the objects of associations lie in the real world. On the contrary, augmented virtuality can be defined as the combination of physical with digital spaces in semantically linked contexts, but where the task's objects lie in the world of computing, states that the systems considered aim to make interaction more realistic.

All the definitions proposed in literature leave little room for multimodality. However, augmented reality nowadays not only exceeded the stage of repositioning virtual indices in a video flow but also proposes sound and even tactile augmentations. To take into account the multimodal aspect of real world, we also propose a new definition of augmented reality: **Augmented reality is the superposition of sensory data (digital or analog) to the real world, so that pursuing a definite goal, it seems to coexist with the real world.** Our definition of augmented reality includes previous definitions to be more general and efficient.

2.2 Mobile Augmented Reality

Technology advances in mobile computing have promoted the development of augmented reality applications. Actually, handheld computers are becoming smaller and lighter. Nowadays they are more accessible and cheaper thanks to highly competitive industries. Therfore, mobile augmented reality aims a wider audience than ever before, as the users own mobile devices and already know how to handle them. Hollerer et al. depict basic components and infrastructure required for mobile augmented reality systems [4]:

- Mobile Computing Platform
- Displays for Mobile AR
- Tracking and Registration
- Environmental Modeling
- Wearable Input and Interaction Technologies
- Wireless Communication and Data Storage Technologies

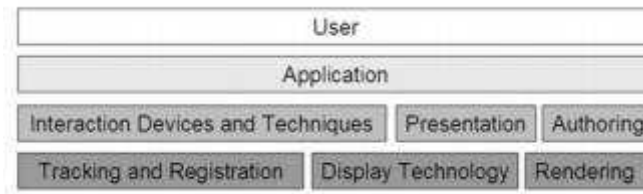## 2.3 Technical Constitution of an Augmented Reality System

Fig. 2. Buildings Blocks for Augmented Reality [5]

Bimber et al. defined general building blocks representing fundamental components of augmented reality [5] :

Base Level: This is the most critical part of an augmented reality system. In the fact, tracking and registration problem are the most fundamental problems in AR research. Much research effort is spent to improve performance, precision, and robustness of tracking systems. In effect, precise alignment between the projected image and the features on the display surface is highly dependent on tracking.

Besides tracking, display technology is another basic building block for augmented reality. Head-mounted displays are the first display technology for AR applications. Today, it is possible to substitute them by Smartphone or tablet screens. The third basic element for augmented reality is real-time rendering. Since AR mainly concentrates on superimposing the real environment with graphical elements, rendering methods should operate in real time.

Second Level: This intermediate level is situated on the top of base level, as can be seen from the figure below. It includes: interaction devices and techniques, presentation, and authoring. Ideas and early implementations of presentation techniques, authoring tools, and interaction devices/techniques for AR applications are just emerging. Some of them are derived from the existing counterparts in related areas such as virtual reality.

Application level: This level represents the interface to the user. Effectively, it is the user-oriented software part of an augmented reality system. At present, it is possible to totally implement an augmented reality application by the use of dedicated Software Development Kit (SDK).

User level: This last layer is finally the user of the application. User studies have to be carried out to provide measures of how effective augmented reality system is.
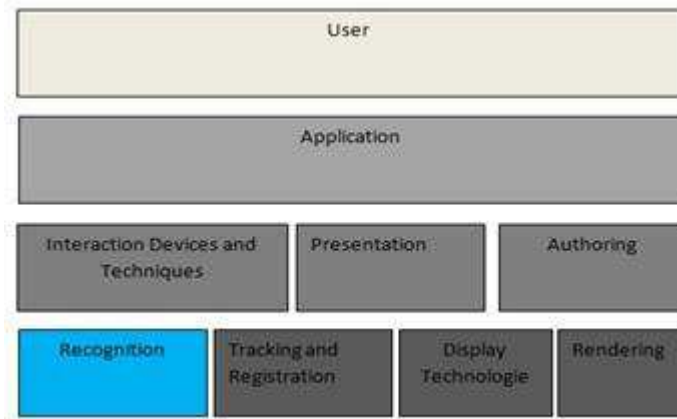
Fig. 3. Modified Technological Constitution of an AR System

Bimber et al. [5] forgot to mention recognition in the base level of AR. Therefore, as highlighted by figure.3, we propose a modified version of building blocks of AR. Since augmentation processes treat each object differently, recognition needs to be achieved. Thus, it is primordial for an augmented reality system to identify the object in front of the camera. Hence, in the last section of this article, we deal with technical aspects of object recognition.

3 Exemples of Mobile Augmented Reality Applications

Campus information system is one of the first mobile augmented reality systems. It was proposed in 1997 by Feiner et al. [10] in their paper entitled "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment". Campus information system aims to assist users in exploring the campus space. As the user moves around the campus, his see-through head mounted display overlays notes on campus buildings, as shown in the figure below. With the emmergence of mobile devices, augmented reality systems turn from heavyweight to lightweight. In fact, thanks to embodied sensors, computing platform and camera, Smartphones or tablets could be used by themselves in a mobile augmeted reality system. In this section, we present examples of mobile augmented reality applications we developped for either indoor or outdoor environments.

3.1 Urban Environment

3.1.1 Augmented Reality Browser.

Tourists visiting an urban environment for the first time may face a lot of problems. They may, for example, not initially have a precise destination [2]. On the other

hand, in any urban environment there are Points of Interest (POIs), where visitors may easily miss if these POIs are not well known or difficult to locate. This type of POI may be described as hidden. In [2], D. McGookin shows how visitors can pass by statues without actually seeing them. In this case, the first question facing tourists confronted with unfamiliar urban environments is: What is worth visiting in the city? The most appropriate answer to this question in such situations should at least contain all the POIs (the most interesting places to visit in urban environments in this case) with highest priority ranking. Priority ranking POIs are those situated close to the visitor's position as well as those considered to be the city's symbols (for example the Eiffel Tower in Paris). To distinguish common land navigation point by point (in which the destination is determined) from navigation in which the destination is not known in advance, we have chosen to call the latter multipoint navigation.

One of the aims of augmented reality is to enhance perception or the visibility of the physical world. The Smartphone's screen acts as a window onto the real world whose video flow can be augmented. We use the geo-referenced data of objects to inform users about their location as shown in figure 4, for example, where the location information of different POIs located close by can be seen. Our system calculates the user's position based on GPS data. It then filters the database so as to only display POIs close to the user. Filtering calculates the distance between the user and the referenced objects using the Haversine formula [16]. With regard to the display, annotations are added to the real scene, which are visible on the smartphone's screen as illustrated in figure 2. For this purpose, we use the Vision See through (VST) technique [3], widely used in augmented reality applications. Just like the documented reality functionality relating to augmented reality, our video flow can be enriched with information identifying what can be seen with the camera. The layout of annotations informs users about the spatial location of POIs with regard to their geographical position. For example, the annotation in the top left means that the POI in question is in front of the user on the left.
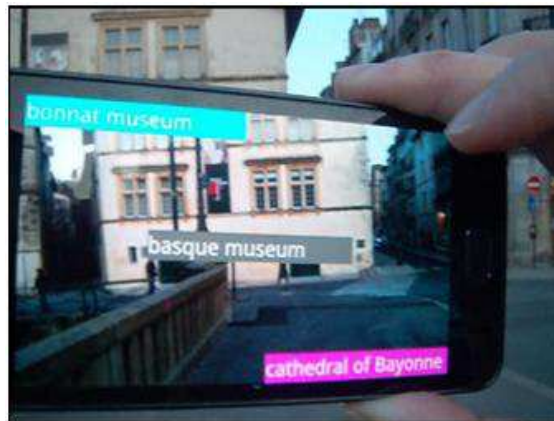

Fig. 4.   Visual Interface of Our Augmented Reality Browser

3.1.2 Virtual Human Guide

Virtual humans represent a natural communication method. Indeed, based on a multimodal interaction mode, a virtual human guide can join gestures to speech, which remember human beings' communication. In this section, we suggest the use of virtual human guides in order to augment touring cultural visits. Educationally rich visits and visitor engagement is also one of the most important factors in the tourism industry. AR has huge potential to actively involve tourists in learning about the visited environment and exploring various museum settings and artifacts like never before.

The church of Sainte Eugenie, named after Napoleon III's wife, Empress Eugenie de Montijo, is a neo-Gothic church of gray stone that dominates the old harbor of Biarritz. To showcase the notable architecture of Sainte Eugenie's church, we integrated a virtual guide in the real scene. Figure 5 below shows the virtual human in a didactic situation.

Fig. 5. Virtual Human in Real Scene

In another application, we overlay digital texture on top of buildings facades. The digital texture holds a virtual human animation. The virtual human highlights the history and the singularities of a particular building. Thus, he attempts to encourage the visitor to enter the POI and explore it.

Fig. 6. Virtual Human Animation in AR Scene

4   Recognition

4.1   Features Points

The use of QR codes [12] generates visual pollution. They are also difficult to deploy in outdoor environment. Therefore, in this section we describe alternative solution to QR code which is features points. Features points are interesting points in an image. Obviously, they are rich in terms of local information contents and stable under local and global perturbations in the image domain such as illumination, brightness, and affine transformations.

Harris corner detector [7] is a well-known feature points' detector, which was proposed in 1988. Harris corner detector uses the eigenvalues of the second moment matrix to determine corner points. However, this detector suffers from scale variance.

SIFT detector introduced by Lowe [8] in 2004, is a scale-invariant detector. The relative descriptor, computes a histogram of local oriented gradients around the interest point and stores the bins in a 128 - dimensional vector (8 orientation bins for each of 4 * 4 location bins).

SURF detector and descriptor, is derived from SIFT. It was coined in 2006, by Bay et al. [9], as a novel scale and rotation-invariant detector and descriptor. It shares with SIFT the same concept of local features descriptors based on the neighborhood of the interest point. Nevertheless, SURF differs in how the interest points are selected and described. SURF detector is based on the Hessian matrix because of its good performance in computation time and accuracy. It relies on the determinant of Hessian matrix for selecting the location and the scale of a feature point. Given a point x = (x, y) in an image I, the Hessian matrix H(x, σ) in x at scale σ is defined as follows:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \qquad (1)$$

Where $L_{xx}(X, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\partial)$ with the image I in point x, and similarly for $L_{xy}(X, \sigma)$ and $L_{yy}(X, \sigma)$. The extraction of SURF descriptor is performed in two steps. The first step consists of finding the orientation to a circular region around the interest point. Then, a square region aligned to the selected orientation is constructed, and therefore the SURF descriptor is extracted from it. Thanks to the use of integral images, SURF detector is faster than others point features detectors. An integral image can be rapidly computed from an input image and used to speed up the computation of the SURF descriptors for that image. The value of the integral image I(x) in a point (x, y) is the sum of all the pixel values of the input image I between the point and the origin.

$$I_\Sigma(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i, j) \qquad (1)$$

The integral image enables fast computation of the intensities over any upright rectangular area of the image. This process is independent of the size of the image or of the area. The extraction of SURF descriptor consists of two steps. The first step fixes a reproducible orientation based on information from a circular region around the interest point. For that purpose, Haar-wavelet responses are computed in x and y direction, and this in a circular neighbourhood of radius 6s around the interest point, with s the scale at which the interest point was detected. The second step constructs a square region aligned to the selected orientation, and extracts the SURF descriptor from it.

4.2 Matching Approach

The task of finding correspondences between two images of the same scene or object is part of many computer vision applications such as object recognition. Once visual features have been extracted from an image, they are matched against a set of features extracted from the other image. Feature descriptors contain a vector of real numbers. The simplest way to compare two features is to compute the Euclidean distance (or the squared Euclidean distance) between their associated descriptors. This computation is obviously slower if the dimension is higher, so descriptors with smaller vector (like the 64-dimensional SURF) are preferable over larger ones (like the 128-dimensional SURF). The distance between two descriptor vectors p and q is evaluated using Euclidean metric:

$$dist(p, q) = \sqrt{\sum_{i=1}^{64} (p_i - q_i)^2} \qquad (2)$$

Linear search for nearest neighbor is costly for real-time applications. Hence, many methods are interested on approximate nearest neighbor search. Our approximate nearest neighbor search is based on the Fast Library for Approximate Nearest Neighbors (FLANN) library proposed by Muja et al [11]. FLANN contains a collec-

tion of algorithms for solving approximate nearest neighbors problem. These algorithms use, among many others, the hierarchical k-means tree or multiple randomized k d-trees. Their library automatically selects the optimal algorithm performing the best approximate nearest neighbor searches for a given dataset.

4.3   Recognition Algorithm

Approximate nearest neighbor search [11] is faster than linear search. However, it generates loss in accuracy. Obviously, it sometimes does not return optimal neighbors. This figure shows false matches generated by FLANN.

Fig. 7. False Matches with FLANN

To overcome this problem, we propose a filtering method on top of approximate nearest neighbor algorithm. Indeed, our recognition algorithm is split into two steps. The first step consists on the approximate nearest neighbor method followed by a filtering method. The second step is based on the Hausdorff distance [15] applied on line models. We note that our recognition method focuses on building recognition.

First, the test image is compared to image dataset using the approximate nearest neighbor method [11]. For each pair (test image, model image), the minimum distance between descriptors is computed. Next, the median of minimum distances is calculated. For pairs (test image, model image) that minimum distance is less than the computed median, we calculate the number of matches. At this stage, a match is considered positive if it fulfills this condition:

$$d < median \qquad (3)$$

where

d : The relative descriptor distance

$median$: The previous computed median of minimum distances

Subsequently, we retain images giving a number of matches equal or higher than 4. This phase is called filtering.

In the second step of our recognition method, each selected model image is aligned according to the test image, using SURF correspondences. In fact, SURF correspondences are used to calculate the homography [20] relating the test and the model images. Given the 2D homography, points of the model image are transformed with respect to the test image. Next, lines segments are extracted from selected images. Line segments detection is achieved by Hough transformation [18]. Then, we carry out the clustering method proposed by Nieto et al. [19], in order to keep only orthogonal lines, which contribute to vanishing points computation. Hence, we obtain a line-based representation of building as shown in the following figure (figure 8).

Next, Hausdorff distance [15] is computed for each pair (test image, model image). The pair giving the smallest Hausdorff distance is considered to be the correct match.
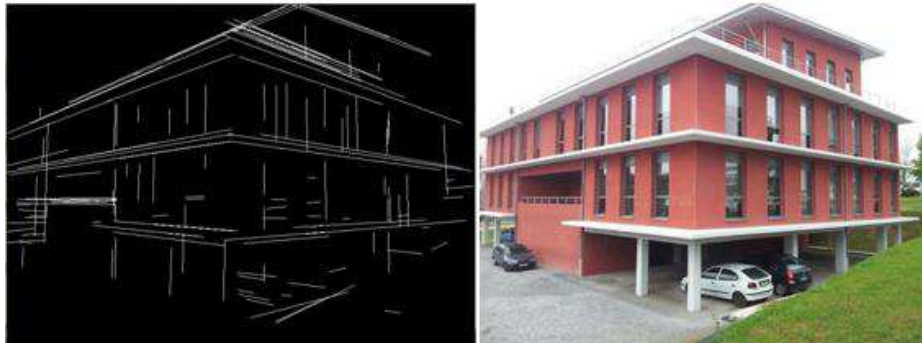

**Fig. 8.** Line model of Building

4.4   Tests and Results

We carried out an experimental study in order to measure the performance of the proposed recognition method. In fact, we compared a test image to a dataset containing fifty images of buildings.

Figure 9 shows the obtained values of minimum distances between matched descriptors. The median of minimum distances values of this dataset is 0.089661.
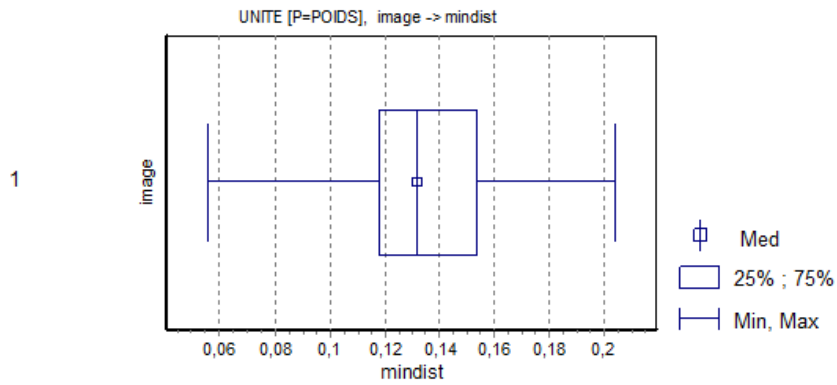
Fig. 9. Median diagram of minimum distances

Figure 10 shows the number of matches after and before filtering step, drawn respectively in blue and red. Only 22% of images participated to the second step of our algorithm. In this last step, an image is discarded if its alignment with the test image failed, otherwise, Hausdorff distance is computed. The obtained results gave that all the alignment processes failed expect the one performed with the correct match. The correct match returned a Hausdorff distance value equal to 10.098.
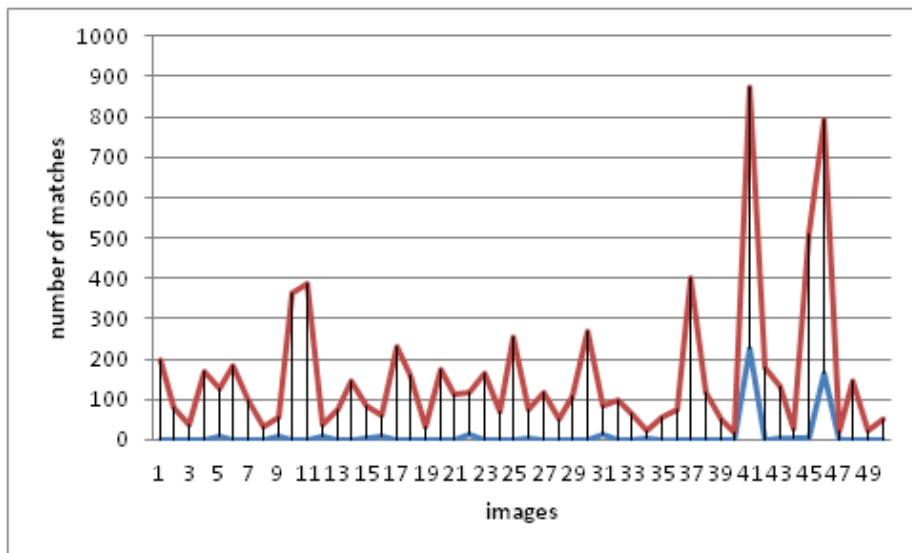
Fig. 10. Impact of Filtering Step

Fig. 11. Alignment Failure

Fig. 12. Successful Alignment. From left to right: test image, model image and rectified image

5   Conclusion

Augmented Reality (AR) describes the overlaying of computerized information on the real world. In this chapter, we gave our definition of augmented Reality: **AR is the superposition of sensory data (digital or analog) to the real world, so that pursuing a definite goal; it seems to coexist with the real world.** Our definition of augmented reality includes previous definitions but it is more general.

Object recognition is a primordial process in augmented reality. Thus, an augmented reality system should identify points of interest (e.g. Buildings, and artifacts) existing in the real scene, in order to apply correspondent augmentations to them. Hence, at the last section of this chapter, we depict our method for building recognition. Our proposition presents a hybrid method, which relies on both points and lines features. The obtained results view the strong performance of this method.

13

References

1. R. Azuma. A survey of augmented reality, Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pages. 355-385, August 1997.
2. D.Mcgookin, S.Brewster, and P.Priego. AudioBubbles : Employing Non-speech Audio to Support Tourist Wayfinding. In Haptic and Audio Interaction Design - HAID, pages. 41–50, Springer-Verlag, July 2009.
3. M.Kanbara, T.Okuma, H.Takemura, and N.Yokoya. A Stereoscopic Video See-through Augmented Reality System Based on Real-time Vision-based Registration. In Proceedings of the IEEE Virtual Reality Conference, pages. 255–262. IEEE Computer Society, March 2000.
4. Tobias H Hollerer and Steven K Feiner,Mobile Augmented Reality. Chapter Nine in book Telegeoinformatics: Location-Based Computing and Services: H. Karimi and A. Hammad (eds.), Taylor & Francis Books Ltd, Jan. 2004.
5. Oliver Bimber and Ramesh Raskar. Spatial Augmented Reality: Merging Real andVirtual Worlds. A K Peters, Ltd., 2005.
6. J.M.Cieutat, O.Hugues and N.Ghouaiel. Active Learning based on the use of Augmented Reality Outline of Possible Applications , International Journal of Computer Applications, 46(20), pages.31-36, May 2012.
7. C. Harris and M. Stephens. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, pages. 147-151, 1988.
8. D. Lowe. Distinctive image features from scale-invariant keypoints, cascade _ltering approach. IJCV, 60(2), pages.91 -110, January 2004.
9. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). Computer Vision Image Understanding, 110(3), pages.346-359, June 2008.
10. Steven Feiner , Blair Macintyre , Tobias Höllerer. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. 1997.
11. M.Muja, D.G. Lowe. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In VISAPP International Conference on Computer Vision Theory and Applications, pages. 331-340, 2009.
12. J. C.Interlando and C.Padilla. Decoding the (41,20,10) Quadratic Residue Code Beyond its Error-Correcting Capability, Applied Mathematical Sciences, Vol. 5, no. 46, pages.2261 − 2269, 2011.
13. P.Milgram. A taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information Systems, Vol E77-D, No.12, pages.1321-1329, December 1994.
14. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. IEEE Computer Graphics and Applications, pages.34-47, November 2001.
15. D. P Huttenlocher, G. A. Klanderman, W. J. Rucklidge. Comparing images using the Hausdorff distance. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(9), pp. 850-863(1993).
16. R.W. Sinnott. Virtues of the Haversine. Sky and Telescope, 68 (2), pages.159-159, August 1984.
17. Sébastien Bottecchia, Jean-Marc Cieutat, Christophe Merlo, Jean-Pierre Jessel. A new AR interaction paradigm for collaborative teleassistance system: the POA. Dans : International Journal on Interactive Design and Manufacturing, Springer, Vol. 3 N. 1, 2009.
18. I.P.V.C. Hough, a method and means for recognizing complex patterns, U.S. Patent No. 3069654, 1962.

19. M. Nieto and L. Salgado, Real-time robust estimation of vanishing points through nonlinear optimization, in IS&T/SPIE Int. Conf. on Real-Time Image and Video Processing, SPIE vol. 7724, 772402, 2010.

20. A.A. Agarwal, C. V. Jawahar, and P. J. Narayanan, A survey of planar homography estimation techniques, Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12, 2005