CONSTRUCTION AND MAINTENANCE OF AUGMENTED REALITY ENVIRONMENTS USING A MIXTURE OF AUTONOMOUS AND MANUAL SURVEYING TECHNIQUES

Joseph Newman, Friedrich Fraundorfer, Gerhard Schall, Dieter Schmalstieg
Institute for Computer Graphics and Vision, Graz University of Technology
Inffeldgasse 16a, A-8010 Graz, Austria
phone : +43 316 873-5011, fax : +43 316 873-5050
email: {newman, fraunfri, schall, schmalstieg}@icg.tu-graz.ac.at

Abstract: Augmented Reality (AR) provides a natural interface to the "calm" pervasive applications anticipated in large-scale Ubiquitous Computing environments. Such applications require detailed, coherent and up-to-date spatial models of the world. Most effort within the Ubicomp community has gone into maintaining the dynamic components of environmental state using sensors. Within the context of AR, these sensors have traditionally been precise, accurate, costly and short-range trackers. However, the modelling of seemingly static structures such as walls and doors is nevertheless important as the semantic significance of spatial events exploited by applications depends on these details. Rich environments, like office buildings, are subject to redevelopment and redeployment such that walls and other partitions are not permanent features that can be surveyed a single time. It is not sufficient for a model merely to be straightforward and cheap to build, it must also be maintainable at a reasonable cost. Each application may also demand a different model, especially as the perception of any given user of their environment depends on their point of view and the task in hand. Consequently, an AR application attempting to enhance or alter this perception must take these two factors into account. This paper presents an approach that combines manual and automated approaches to creating and maintaining an environment of AR and Ubicomp applications.

1. Introduction and Motivation

Augmented Reality (AR) has the potential to provide a natural interface to the "calm" pervasive technology anticipated in large-scale Ubiquitous Computing [9] environments. However, most AR applications have hitherto been constrained, by the working volumes of tracking technologies, to static spaces of a few cubic metres. Furthermore, an assumption has been made that sensors are deployed homogeneously and statically throughout the area of interest, resulting in a single one-off tedious off-line calibration.

An approach called Ubiquitous Tracking [10] attempts to automate the process of dynamically integrating arbitrary sensors in distributed sensor networks. Whilst focussing on the dynamic spatial relationships in a given environment. The semantic depth of events like "person A is in room W" depend not only on the concept of a *person* that can move, but also on the concept of a *room* that cannot. Nevertheless the *room* must nevertheless be measured, and meaning assigned to these measurements. Changes in building use, and even routine maintenance, mean that that new measurements may need to be made displacing or augmenting old ones. The integrity of the spatial model depends on the complete history of measurements, and care must be taken in the assumptions that are made.

Due to the expense and limited range of current commercial trackers designed for use by the Virtual and Augmented Reality communities, visual tracking of fiducial markers has become very popular. Attempts have been made at deploying markers over a wide area in order to extend tracking range. It is necessary to know the position and orientation of the markers as accurately as possible. Earlier surveying techniques involved the use of reflectorless Total stations to survey the positions of the markers. This manual approach is time-consuming, and presents a serious barrier to the introduction of AR to new environments. Therefore automatic methods are necessary to speed up the process. In this paper we propose the use of an autonomously navigating mobile robot to detect and localise the fiducial markers and build a model that can be used by existing AR systems. The mobile robot is equipped with a laser range finder to localise the robot as well as with a digital camera. The images taken from the camera are used to detect the fiducial markers. By fusing the 3D position of the markers with the laser based position of the robot the absolute pose of the fiducial markers can be calculated. The effectiveness of this new approach can be assessed by comparing the model obtained from the Total station with that obtained using the robot. Furthermore, a hybrid approach in which measurements from both the robot and the Total station is presented.

2. Related Work

In [7] techniques for rapidly and accurately surveying the locations of widely distributed markers with the theodolite-based measurement system Leica TPS 700 Total Station [8], whilst simultaneously building a model of the environment were described. To implement a wide area indoor tracking solution used a set of known markers that were distributed throughout the environment. Together with a geometric model of the building that includes the location of the well-known markers we can compute the user's location as soon as a marker is tracked by the optical tracking system.

The tracking library we used is called ARToolkit [6] but any vision-based fiducial tracker would be sufficient. ARToolkit markers (shown in Figure 1) are black and white patterns printed on paper and attached to marker templates. Each marker shows a different digital decoded pattern on it, so that unique marker identification is possible. These markers have very high contrast corners that are very useful features for calibration.

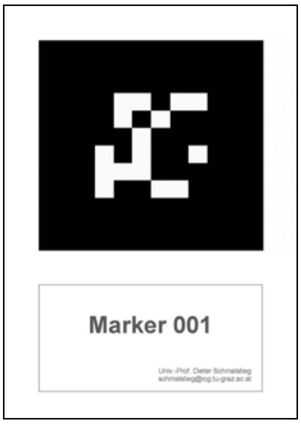


Figure 1: Fiducial Marker

3. Experiments

3.1. Autonomous measurement of fiducial markers

Our goal is to automatically create a map of deployed fiducial markers. A mobile robot capable of exploring unknown environments should detect fiducial markers and measure their position in a single coordinate system. Our PeopleBot (ActivMedia) robot is equipped with a laser range finder (LRF) (Sick LMS 200) and a 2MP digital color camera equipped with a wide-angle lense. The data from the LRF can be used to create a floor plan of the explored area. Laser readings are taken every 5cm and after a final registration of all the readings floor plan and robot positions (including orientation) are available with high accuracy. The map creation is done with the “ScanStudio” [1] software.

The detection of the fiducial markers is performed using the images acquired by the digital camera. After detection the 3D coordinates of the marker’s 4 corner points are computed using stereo reconstruction. The reconstructed marker points are transformed into the overall coordinate system determined using the LRF by fusing the camera coordinate system with that of the laser pose.

3.2. Marker detection

The images captured from the camera have a resolution of 1600x1200 pixels. Color information is discarded and all subsequent steps work exclusively on grayscale images. The lense has a field of view of 90°. The camera is calibrated; interior orientation as well as lense distortion is known. In a first step the images are resampled to compensate for the lens distortion, as the wide angle lense results in high radial distortion. To detect the markers in the images we extract Maximally Stable Extremal Regions (MSER) [2]. This local detector is threshold-based and is well suited to finding the deployed markers. However, this is a general approach and the detector also returns other stable image regions, which can be used as natural landmarks. The subsequent methods are applied to all detected landmarks. The classification of a detected landmark as a fiducial marker or a natural landmark is performed at the end of the mapping workflow. Figure 2(a) shows that three landmarks have been detected by the algorithm.

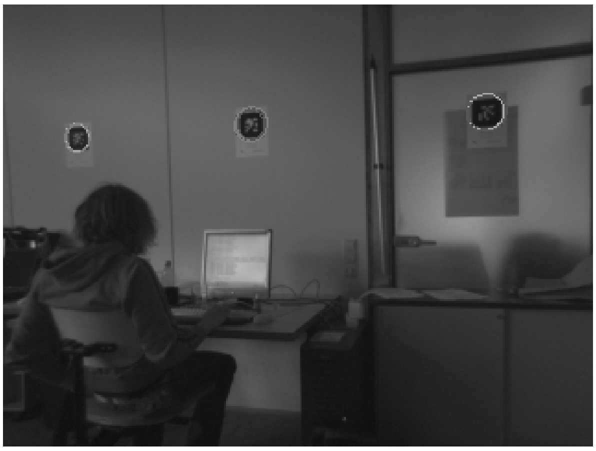
3.3. Marker reconstruction

3D reconstruction of the landmarks is done using a shape-from-motion approach (the markers are viewed from two or more different viewpoints which allows the calculation of the 3D position). Two nearby frames from the image sequence are selected and the essential matrix is calculated for the image pair. The estimation is done automatically. Harris corners are detected in both images and matched using normalized cross-correlation (see Figure 2(b)). The essential matrix is calculated on an inlier set obtained from RANSAC using the 5-point algorithm proposed by Nister [3].

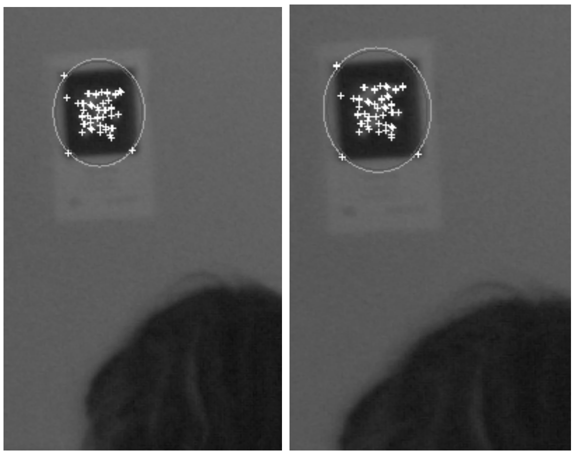
The next step is the reconstruction of the detected landmarks. As depicted in Figure 2(b) corresponding landmarks in both images are matched using SIFT descriptors [5] with a method proposed in [4]. The matching method returns a logical matching of the landmarks as well as accurate point matches within the landmarks. For every landmark we create a 3D reconstruction using the appropriate point matches. We assume that the landmarks are planar and do a robust plane fitting in 3D.

In the next step we now use our knowledge of the deployed fiducial markers to extract them from the set of all detected landmarks. This is done using basic image processing techniques, the markers in question consist of a black square surrounded by a white border. If we identify a landmark as fiducial marker we extract the 4 corner points in one image and project it onto the plane of the landmark in 3D to obtain the 3D coordinates of the corner points. We determine the scale factor of the metric reconstruction by our knowledge of the real size of the markers (153mm x153mm). The reconstruction is now in a canonical coordinate system.

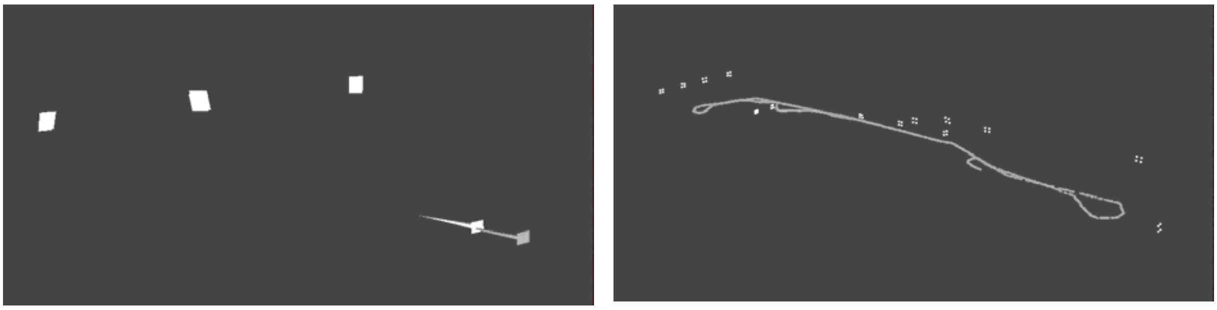
The final step is to transform the single reconstructions into the overall coordinate system established from the laser scanning. For that the 3D points of the markers are rotated and translated $p_{new} = Rp + T$. R is the rotation matrix describing the orientation of the robot and thus the camera and $T = [x, y, z]^T$ is the position of the robot for the frame of the image sequence used for the reconstruction. The final result can be seen in Figure 2(d), where the path the robot drove and the detected and reconstructed markers are drawn.



(a)



(b)



(c) (d)

Figure 2: (a) Image with 3 visible and detected markers. (b) Detected corresponding markers (c) Marker reconstruction in canonical coordinate system. (d) Map of the markers detected and reconstructed from the robot.

The accuracy of the results from the robot´s measurements is limited because of the use of COTS components (commercial off-the-shelf).

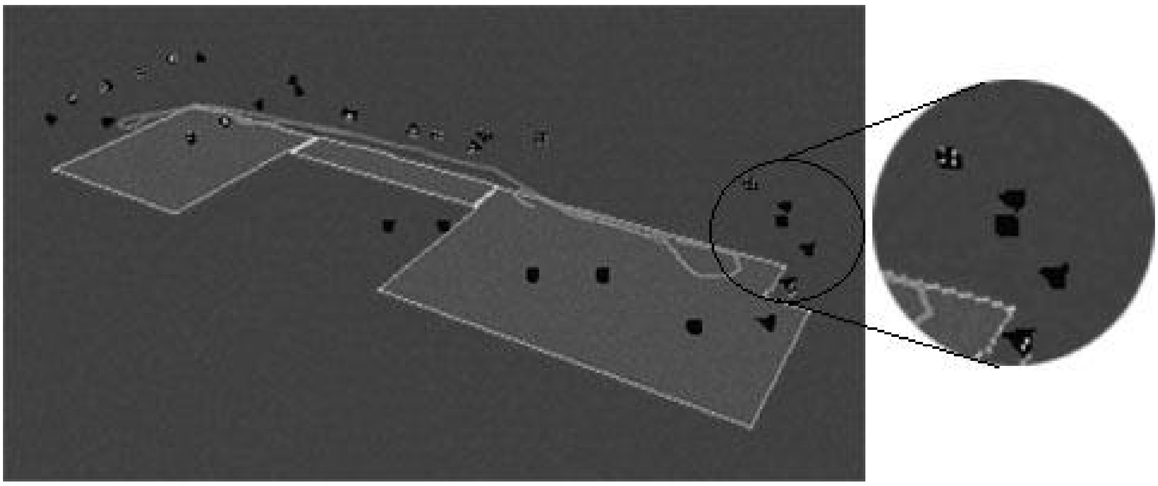


Figure 3: Superimposed map of the markers reconstructed from the robot and the map of markers measured with the Total Station

Figure 3 depicts a superimposed image of the map of markers reconstructed from the robot (shown in Figure 2 (d)) and the map obtained by the Total station including the floor plan and markers drawn as cones. In the zoomed-in area one can observe the superimposition of the markers.

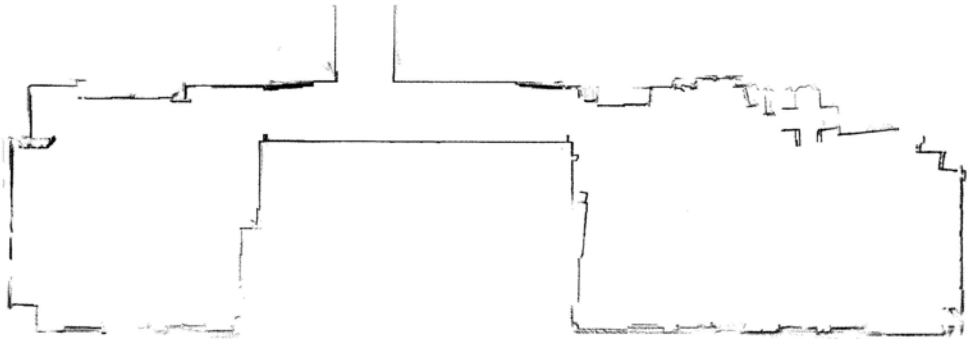


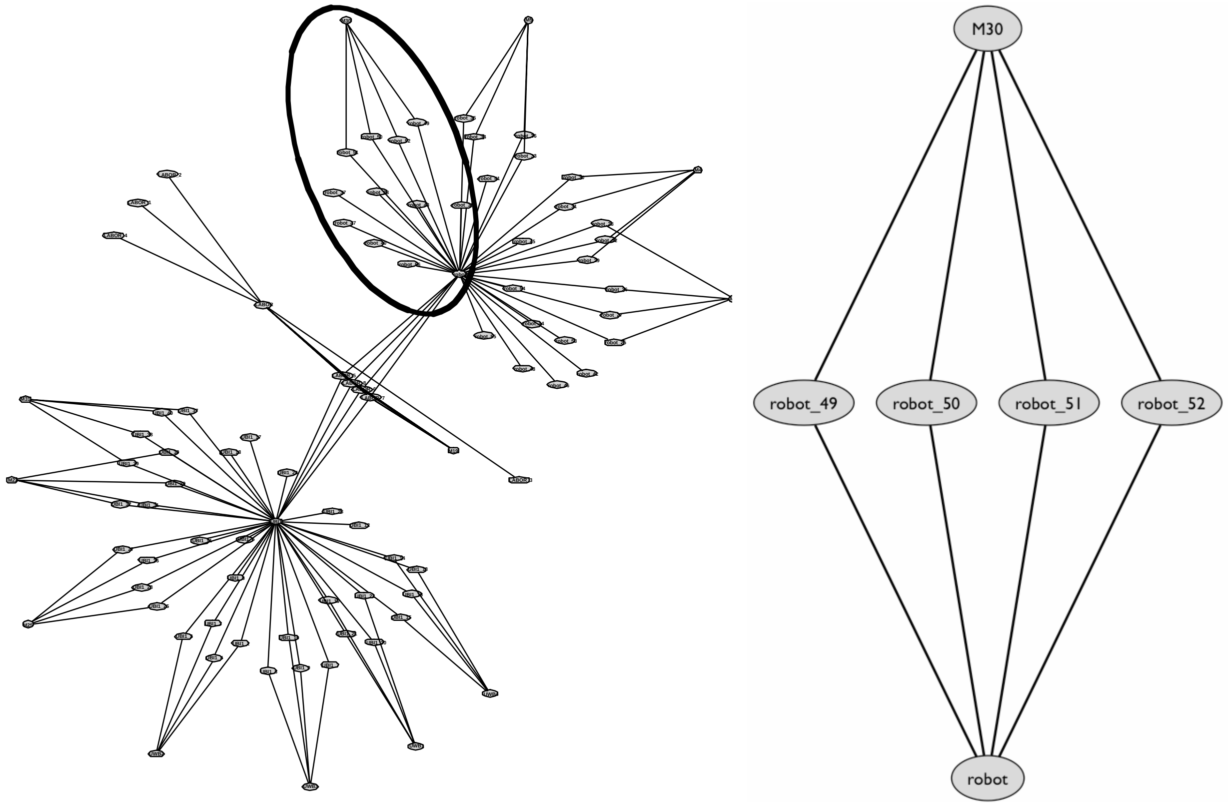
Figure 4: Two-dimensional floor plan measured by the robot

Figure 4 shows the floor plan of the part of the building that was measured with the laser range scanner mounted on the robot.

4. Surveying Strategy

Spatial relationships can, in general, be represented by a graph [11], in which objects are nodes, and spatial relationships between objects are directed edges. Each edge can represent either a measurement, or an estimate derived from other measurements. A complete spatial relationship (SR) graph would represent environmental state in its entirety and could be used to query relationships between two arbitrary objects. However, depending on the application in question the SR graph would be formed in different ways. As an illustration imagine two people, *A* and *B,* and two applications: a command-and-control view of a 2D map, and an AR application in which *A* and *B* are presented with an augmented view of one another. In the first case the viewer will want a globally optimised view of the locations of *A* and *B,* whilst the AR application will seek to minimise the registration error between *A* and *B* and their augmentations.

Figure 5(a) shows the spatial relationship graph of some of the measurements that were made in the course of surveying our lab in Graz. It has been pruned to reduce its size and complexity, and although still very large some structure is discernible with clusters of four nodes corresponding to the corners of the markers. The ellipse in this figure highlights a cluster that corresponds to the marker *M30* with points {*robot_49, robot_50, robot_51, robot_52*} being measured in the coordinate frame of the *robot*, that can be seen in Figure 5(b). A search throughout the entire graph in Figure 5(a) yields a set of similar subgraphs that corresponds to pairs of coordinate frames and common point features. A 4x4 homogeneous transformation matrix, and its inverse, is then calculated for each pair using a linear least squares solution. These transformations then paramaterise directed edges in a new graph seen in Figure 6. A choice can then be made as to which coordinate frame should form the root of model. A breadth-first-search starting at the root node in question, in the case of the graph in Figure 6 this root node is *LABOR*, and the transformation matrices parameterising each edge are concatenated by multiplication during the traversal. This yields yet another graph seen in Figure 7. The breadth-first-search does not result in an optimal solution, as there is no attempt to choose a well-conditioned set of common points for the least-squares solution. However, it can be used as the initial estimate in a non-linear optimisation, in which appropriate weightings are applied to each measurement.



(a) Pruned graph of spatial measurements (b) Subgraph of measurements

Figure 6: Graphs of spatial measurement

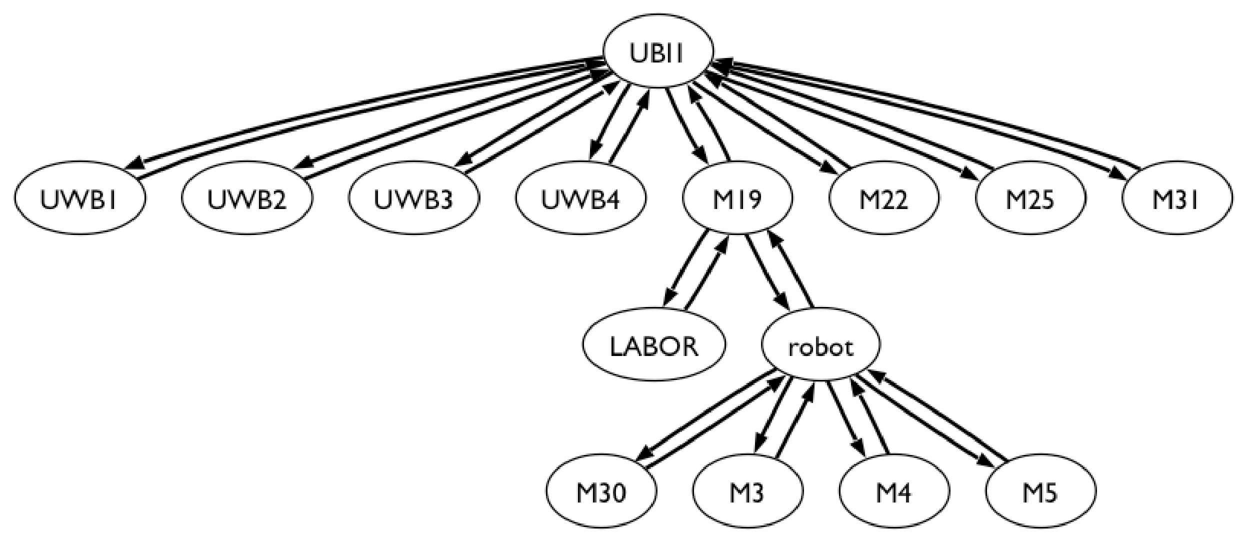


Figure 7: Graph of spatial relationships of coordinate systems

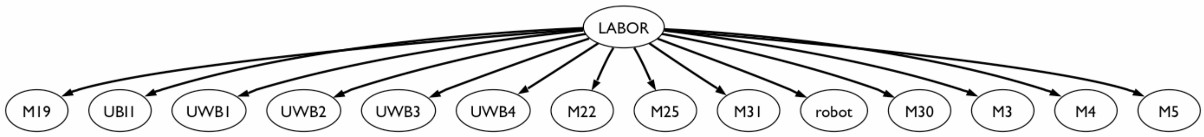


Figure 8: Graph of estimated spatial relationships with edges emanating from given root

This approach is applied to both measurements obtained by the robot and the Total station. This points to a composite strategy in which, the Total station can be used to survey important rooms with significant resources and equipment that needs to be surveyed and calibrated. These rooms are often cluttered and do not lend themselves to automatic surveying. However, these manually surveyed "islands" can be chained together using the robot that can easily navigate between rooms through corridors. Manually replicating this chaining process is extremely tedious as it requires multiple hops with point correspondences.

5. Future Work

The process of doing the non-linear optimisation described in the previous section is still being developed. Also, in order to be of practical use in real AR applications, the spatial relationship graphs will have to be mapped into a usable form (e.g. geometry) at interactive rates. This need not occur on a frame-by-frame basis on the client-side, but can take place as a background activity on a powerful server, or cluster of servers.

6. Conclusions

Ubicomp and AR applications require access to environmental state, namely spatial state. This state can be encapsulated in a model that must be kept up-to-date. Models cannot be created on a one-off basis, but must be capable of construction, augmentation and partial destruction in a way that is not just affordable, but also maintainable. A single model is not sufficient to cover the full spectrum of applications, and must be created on the fly according the unique needs of each application. We have demonstrated techniques whereby measurements from multiple sources can be easily combined to create bespoke models on a just-in-time basis.

7. References

[1] Gutmann J.-S. and Schlegel C, AMOS: Comparison of Scan Matching Approaches for Self-Localization in Indoor Environments, 1st Euromicro Workshop on Advanced Mobile Robots (Eurobot'96), 1996.

[2] Matas J., Chum O., Urban M., and Pajdla T., Robust wide baseline stereo from maximally stable extremal regions. In Proc. 13th British Machine Vision Conference, Cardiff, UK, pages 384–393, 2002.

[3] Nister D., An efficient solution to the five-point relative pose problem. In IEEE Trans. Pattern Anal. Mach. Intell. IEEE Computer Society, 2004.

[4] Fraundorfer F., Ober S., Bischof H., Natural, Salient Image Patches for Robot Localization Proc. 17th International Conference on Pattern Recognition, pp. 881 - 884, vol. IV, eds.: J. Kittler and M. Petrou and M. Nixon, publ.: IEEE CSP, 2004.

[5] Lowe D., Distinctive Image Features From Scale-invariant Keypoints. International Journal of Computer Vision, 60(2):91–110, November 2004.

[6] Billinghurst, M. and Kato, H., Collaborative mixed reality, in 'Proc. ISMR'99', Springer Verlag, Yokohama, Japan, pp. 261–284, 1999

[7] Schall G., Newman J., Schmalstieg D., Rapid and Accurate Deployment of Fiducial Markers for Augmented Reality, in Proc. 10th Computer Vision Winter Workshop (CVWW 2005). http://www.icg.tu-graz.ac.at/pub/pubobjects/schall05cvww [2005, June 13], 2005.

[8] Leica Geosystems. Advanced realtime tracking Gmbh. http://www.leicageosystems.com/corporate/en/products/totalstations/lgs5166.htm [2004, December 20], 2004.

[9] M. Weiser and J. S. Brown. Beyond Calculation: The Next Fifty Years of Computing. Copernicus, 1998.

[10] J. Newman, M. Wagner, M. Bauer, A. M. lliams, T. Pintaric, D. Beyer, D. Pustka, F. Strasser, D. Schmalstieg, and G. Klinker. Ubiquitous tracking for augmented reality. In Proceedings of ISMAR 2004, November 2004.

[11] B. Brumitt, J. Krumm, B. Meyers, and S. Shafer. Ubiquitous computing and the role of geometry. IEEE Personal Communications, pages 41–43, October 2000. 2