MIXED REALITY WITH MOVEABLE 3D SEE-THROUGH DISPLAY AND VIDEO-BASED OBJECT TRACKING

Christos Conomis, Siegmund Pastoor, René de la Barré, Hans Röder, Volkert Tietje
Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institut,
Einsteinufer 37, 10587 Berlin, Germany
E-mail: {conomis, pastoor, barre, roeder, tietje}@hhi.fraunhofer.de

ABSTRACT

This paper presents a novel moveable mixed reality system with video-based tracking. The system consists of a high resolution video display which is mounted on the end effector of a passive kinematics chain with revolving joints. The monitor shows the live video stream with additional information superimposed in correct perspective onto the live scene. The user can freely move the system within a large work envelope and observe the target object from different viewpoints. Video and model-based object tracking is applied to accurately recover metric information approximately in real time.

Experimental results show that the system can be efficiently used for a number of applications including maintenance and repair, product design, and entertainment.

1. INTRODUCTION

Mixed Reality (MR) is a key technology that integrates dynamically generated 3D objects into a physical 3D space in real-time and can be used to navigate, manage and visualise information resources and to control tasks. The term "Mixed Reality" was introduced by Milgram [1] in order to enclose both Augmented Reality and Augmented Virtuality in the reality-virtuality continuum. Moreover MR systems should allow dynamic user interaction and correctly render interrelations between virtual and real objects in the mixed environment (such as occlusions, lights and shadows).

The potential benefits of MR systems have been widely recognized and during the past decade there have been substantial technological advances in this field. However concepts and adequate techniques for mobile and moveable systems are still a topic of ongoing research.

Being mobile expands the applicability and acceptance of MR systems to their full potential. On the other hand the performance of MR applications depends heavily on the registration methods and their properties which are used to fulfil the alignment task. In order to keep the real and the computer generated objects in fixed position, angle and size with respect to each other, it's crucial that changes of the user's viewpoint as well as changes of the real scene, such as movements of the real objects and changing occlusions, are fast and accurately detected and compensated.

In movable systems, the camera is usually mounted at the mobile platform and moved with the user. Hence small movements and rotations of the user produce large image motions which make the registration and tracking problem hard to solve. Registration mismatches and misalignments result in frame jitter and pose estimation inaccuracies. The last is of great importance especially for MR systems that aim at spatial visualisation with 3D displays since human vision perceives misalignment errors magnified at the stereoscopic distance.

The rest of the paper is organized as follows: In Section 2 we briefly discuss the related work in the field of moveable mixed reality systems. Section 3 presents our concept and gives a short overview of our system. Section 4 steps through the different stages of the mixed world generation process, and Section 5 presents some experimental results. Preliminary conclusions and directions for future work are given in Section 6.

2. RELATED WORK

Most of the movable MR systems proposed in related work use specialised see-through head-mounted displays (HMD). Although HMD-based MR systems require precise user-depended optical calibration which is not automated and prone to errors, they have great potential since they are wearable and leave the hands of the user free. Most of the reported HMD approaches rely on different sensors to ensure accurate registration (see [2] and [3]). Recently presented HMD systems that claim to have good results are found in [2], [3], [4], and [5].

Another class of MR systems is based on handheld devices such as Tablet PCs and PDAs. These approaches are closer to our system concept since they employ video see-through visualisation combined with video based registration. In contrast to our work these approaches process, due to the limited resources, still images [6], are implemented in client/server architecture and thus are not adequate for real time applications [6], or rely on a small

number of pre-stored patterns [7]. A large number of existing systems apply fiducial markers. However, markers limit the system's accuracy and require a "prepared" environment.

3. SYSTEM OVERVIEW

Figure 1 shows two variants of the moveable articulated video-MR system, the one (Fig. 1a) using video-based overlay and the other one optical superposition (Fig. 1b). The video-based system consists of four main parts:

1. a passive kinematics chain with five revolving joints which remains "at position" when no force is applied by the user,
2. a high resolution video display which is mounted on the end effector of the kinematics chain,
3. a video camera (or a stereo pair) fixed to the monitor, viewing the object from almost the same perspective as the user, and
4. a PC which is used for video-based object registration and rendering of the virtual scene.

a b

Figure 1: Moveable Articulated MR systems with video (a) and optical overlay (b).

The user may freely move the monitor within a large work envelope and observe the target object from different vantage points. In the video-overlay system (Fig. 1a) the monitor shows the live video stream with additional information superimposed in correct perspective. Video based object registration recovers the relative position and orientation between the target object and the camera in real time. The target object can be arbitrarily moved and repositioned during operation. No extra sensors are attached to the joints of the arm. The system enables natural real-time user interaction.

The concept of the moveable articulated video-MR system has several advantages over other existing approaches for mobile mixed reality. Opposed to other approaches, multiple users can view the display simultaneously, and the screen position and orientation can be "frozen" to study and discuss details of a certain view. This concept provides also an adequate platform for mounting state-of- the-art high-quality 3D displays such as the Variable Accommodation MR Display [8] which at the moment weighs several kilograms.

4. MIXED WORLD GENERATION

We aim at applications in real environments in which interaction occurs with certain pre-known technical objects of interest. In such situations we assume that a generic model of the target object consisting of significant features like points, edges, conics contours, texture and colour, is a-priori known. This model is used for object registration and visualization (rendering) purposes.

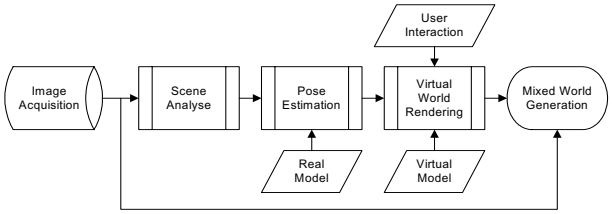

Figure 2: General flow of video and model-based mixed reality generation.

A general diagram of mixed world generation is given in Fig. 2. A camera continually captures the environment, and the acquired images are analysed. Adequate features are extracted from the images and compared with the model, in order to estimate the current target object's pose with respect to the camera coordinate system. The estimated pose as well as the user's interaction is then used to render perceptively correct augmentations of the virtual world. In video-based mixed reality the synthesized data are superimposed onto the live image.

4.1 Video and model-based object registration

Our video-based object tracking technique includes three essential steps: 1) object localisation in the image domain using appearance based features, 2) initial pose estimation of the camera viewpoint using geometry based primitives and 3) iterative refinement of pose estimation. For the following we assume that the camera's intrinsic matrix is known from off-line calibration.

Object localization is performed using colour histograms (CH). Apart from simplicity, this approach is very fast and thus suited for real-time applications. CH provide good results for extended viewpoints (camera

positions) since they are invariant against rotation, translation and scaling. In our work they are generated off-line under a few different light conditions and stored with the model. In the initialising phase candidate pixels are first extracted with CHs. The extracted pixels are then labelled into regions using the connected component algorithm and checked for their geometrical constancy using central moment analysis.

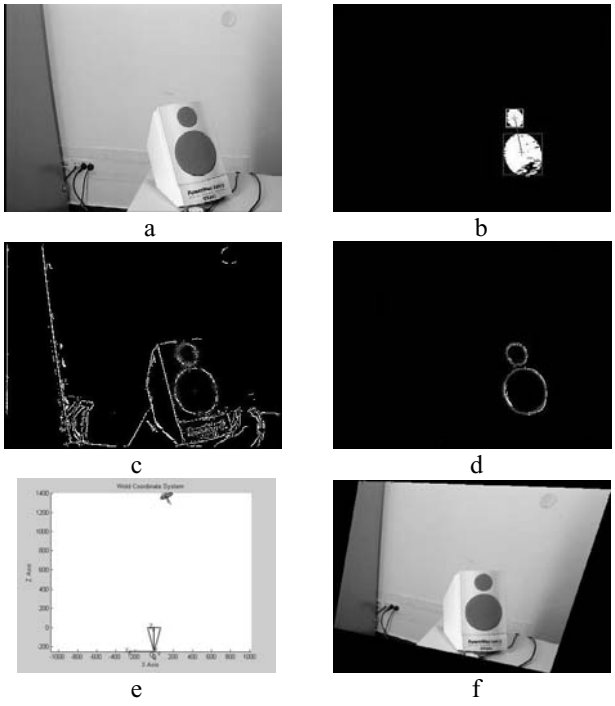

Figure 3: Overview of the initialisation process. a) original image, b) colour segmentation, c) edge and interest point features detection, d) fitted conics, e) 3D reconstruction from conics and f) rectified image from conics reconstruction results (shown for illustration only, no integral part of the initialisation process).

Once the target object is approximately localized we extract edges and fit the edges to basic geometric primitives such as lines and conics using a RANSAC based least squares method to ensure suppression of outliers. The initial pose estimate is obtained from a two-fold strategy, namely from homography decomposition and from conics rectification. Since homography decomposition is a well known technique (see [9]) we discuss here only the approach of pose estimation from conics. If the radius of a circle and the camera intrinsic matrix are given, we can recover the centre of the circle with respect to the camera (world) coordinate system and the normal vector of the circle's supporting plane up to a twofold ambiguity [10]. In case of two coplanar circles the whole rigid transformation can be recovered without ambiguity following a few steps: 1) resolve the ambiguity of the normal vector by choosing the common normal vector between the two conics; 2) compute the vector connecting the two centres of the circles; and 3) recover the third rotation axis as the cross product of the two vectors. The initialisation steps are shown in Figure 3.

To refine the pose estimate from the previous step, the estimation process is formulated as a minimisation problem of the error vector between the extracted edge segments and the projected model edges, which is solved iteratively. Since our initial pose estimate is close enough to the true one it takes only a few iterations to converge to the correct position without stacking in local minima.

4.2 Dynamic object generation and overlay

For dynamic object generation the projective transformations of a virtual camera and the model of the object are employed. The model is described by a triangle mesh as well as texture information of the model faces. The virtual camera is decomposed as

$$P_V = K_V \begin{bmatrix} R & | & t \end{bmatrix}$$

where K_v is the intrinsic camera matrix, R is the rotation matrix and t the translation vector between the model coordinate system and the camera coordinate system. R and t define the so called modelling transformation.

The virtual camera has the same viewing volume as the real camera. The modelling transformation is computed from the pose estimation algorithm and the virtual camera is updated every frame. We use the virtual camera and the user interaction to project and to render views of the model. The synthesized image is then superimposed onto the live video image.

5. RESULTS

The presented algorithms have been implemented on a standard PC Pentium IV 2.0 GHz equipped with a camera that provides colour images at full PAL resolution. As the user moves, the relative pose of the camera with respect to the object is computed and dynamically generated views of the object as well as scene information are superimposed at 17 frames per second onto the live video. The user can interactively choose different visualization modes while he/she is free to arbitrarily move the arm and reposition the target object.

Figure 4 shows example frames of a live scene of a loudspeaker taken from different viewpoints overlaid with the wire frame model as well as other information. A design application has also been implemented.

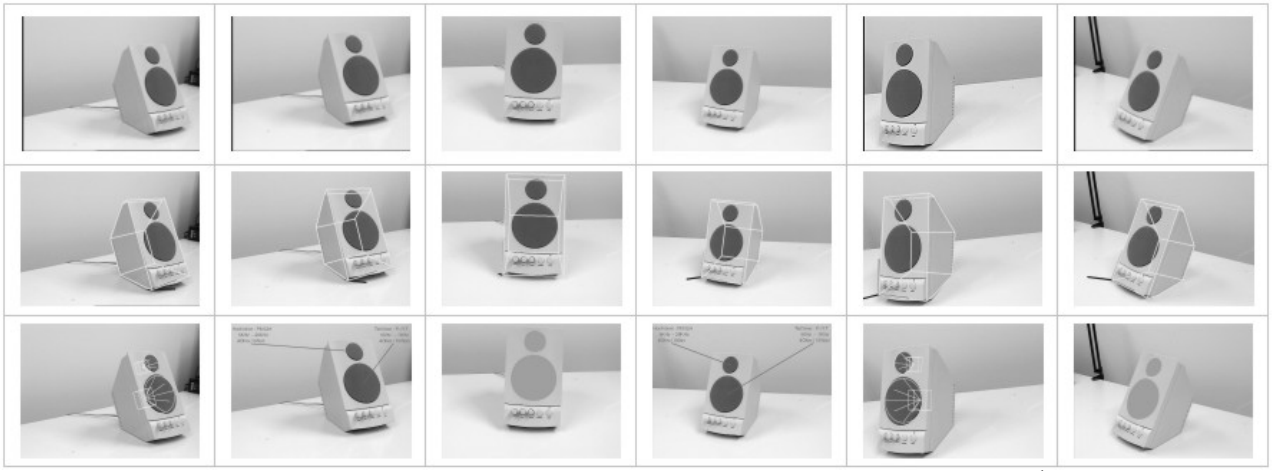

Figure 4: 1st row shows six original pictures out of a 800 picture sequence of a loudspeaker. 2nd row shows the same pictures superimposed with the wire frame model. 3d row shows various visualisation modus for the same pictures: 1st & 5th picture show structure visualisation of hidden parts, 2nd & 4th information annotation, and 3d & 6th change of exterior attributes (e.g. colour) in a design application.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a moveable articulated MR system with marker less model-based object tracking. Articulated mixed reality systems are a good compromise between desktop MR systems and handheld MR systems. They allow moveable interaction in a quite big working envelope such as mobile and handheld-systems while at the same time ensure stable, high quality visualisation comparable to a desktop MR system. In the future we intend to compare the existing video-overlay solution with the optical-overlay approach based on the recently constructed free-viewing VAMR (Variable Accommodation MR [8]) shown in Figure 1b.

7. ACKNOWLEDGEMENTS

This work is part of the mixed3D project which investigates autostereoscopic 3D displays with novel natural user interfaces. The mixed3D project is funded by the German Federal Ministry for Education and Research (BMBF) under grand 01 BD 250.
http://www.hhi.fhg.de/german/im/projects/mixed3d

8. REFERENCES

[1] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays, *"IEICE Transactions on Information Systems*", Vol E77-D, No.12 December 1994.

[2] R. Tenmoku, M. Kanbara and N. Yokoya, "A Wearable Augmented Reality System for Navigation - Using Positioning Infrastructures and a Pedometer", *IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 344-345, October 2003.

[3] A. J. Davison, W. W. Mayol, and D. W. Murray, "Real-Time Localisation and Mapping with Wearable Active Vision", *IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 18-27, October 2003.

[4] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods, "An Occlusion-Capable Optical See-through Head Mount Display for Supporting Co-located Collaboration", *IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 133-141, October 2003.

[5] G. Klein and T. Drummond, "Robust Visual Tracking for Non-Instrumented Augmented Reality", *IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 113-122, October 2003.

[6] S. Goose and G. Schneider, "Augmented Reality in the Palm of your Hand: A PDA-Based Framework Offering a Location-based, 3D and Speech-Driven User Interface", *Telecommunications and Mobile Computing, TCMC,* 2003.

[7] W. Pasman and C. Woodward, "Implementation of an Augmented Reality System on a PDA", *IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 276-277, October 2003.

[8] S. Pastoor and C. Conomis, "Mixed Reality Displays" in: O. Schreer, P. Kauff and T. Sikora (eds.), "3D Videocommunication - Algorithms, concepts and real-rime systems in human centred communication", *Wiley & Sons Ltd.*, Chichester, 2005 (in print).

[9] G. Simon, A. W. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene", *Proc. IEEE and ACM Int. Symp. on Augmented Reality ISAR 2000*, pp. 120-128, 2000.

[10] S. Rad, K. C. Smith A, B. Benhabib and I Tchoukanov, "3D location estimation of circular features for machine vision", *IEEE Trans. Robot Automation*, vol. 8, pp.624-640, Oct 1992.