Computer Vision Based Head Tracking from Re-configurable 2D Markers for AR *

Yong Liu, Moritz Störring, Thomas B. Moeslund, Claus B. Madsen, Erik Granum
Computer Vision and Media Technology Laboratory, Aalborg University
Niels Jernes Vej 14, DK-9220 Aalborg, Denmark
{yliu,mst,tbm,cbm,eg}@cvmt.dk

Abstract

This paper presents a computer vision based head tracking system for augmented reality. A camera is attached to a head mounted display and used to track markers in the user's field of view. These markers are movable on a table and used as interface with virtual objects. Furthermore, they are used as landmarks to track the user's viewpoint and viewing direction by a homography based camera pose estimation algorithm. By integrating this computer vision based tracker with a commercially available InterSense tracker, we take the advantages of the small jitter of the former one without losing the tracking speed of the latter one. For static and slow head motion the system has less than 0.3mm RMS of position jitter and 0.165 degrees RMS of orientation jitter.

1. Introduction

One aim of Augmented Reality (AR) is to give the user the impression that the virtual objects are a part of the real world, i.e., that they are not moving with respect to real objects but rather are 'glued' to them. A crucial part of an AR system is the tracking of the user's viewpoint and viewing direction with respect to the real scene. To enable a realistic overlay of virtual objects on a real scene the tracking has to be accurate and fast with a rather low latency [1]. Different tracking technologies are used in AR-systems. Among the most popular ones are magnetic, acoustic, inertia, and optical systems and hybrid versions of these [5]. Inertia tracking provides high update rates and low latencies [2]. It performs best in the non-static case, i.e., whenever the tracked head is moving. In the static case inertia tracking has the disadvantage of fading. Magnetic, acoustic, and optical systems are more appropriate for the static case or for slow movements. Magnetic systems are sensitive to metal in its working environment. Acoustic and optical system need a free line of sight and are sensitive to temperature changes and illumination changes, respectively. Furthermore when the user is more or less static, i.e., examining a virtual object, the magnetic and acoustic systems are jittery. This disturbs the user's experience because the virtual overlay jitters with respect to the real world. Therefore some papers have suggested fusion of several technologies, e.g [2].

This paper introduces a tracking system which integrates a commercial InterSense IS-600 tracker with computer vision based tracking. The IS-600 is an acoustic and inetial cube integrated tracker. The computer vision based tracking uses a head mounted camera (HMC) and a homography-based pose estimation algorithm. In our AR system, we use a number of 2D color coded movable markers that have two functions: 1) they are used as interfaces to virtual objects, i.e. the user can move a marker and through this move a virtual object that is attached to this marker. 2) they serve as landmarks for the proposed head pose tracking method. The computer vision based pose estimation is used when the user is close to static. When the user is moving, the data from the IS 600 tracker is used instead. In this way, both the advantages of the IS tracking and computer vision based tracking are combined to speed up tracking and decrease jitter. A full description of the work can be found in [3].

2. Homography based Pose Estimation

A homography based camera pose estimation algorithm is used for the head pose estimation [4]. Since all the 2D markers are located on the same 2D plane, there is a homography between these markers and their images. The rotation $\mathbf{R}$ and translation $\mathbf{t}$ is related to the homography between the 2D images and 3D plane as

$$\overline{\mathbf{H}} = \pm \left(\mathbf{r}_1 \quad \mathbf{r}_3 \quad \mathbf{t} \right) , \tag{1}$$

[1] Supported by the ARTHUR project under the European Commissions IST program (IST-2000-28559)

where $\mathbf{H}$ is the homography matrix, $\mathbf{r}_1$ and $\mathbf{r}_3$ are the 1-th and the 3-th column vectors.

For more than four 2D-3D correspondences, an initial linear estimation for the homography can be obtained. Then a Levenberg-Marquadt optimization method can be used for the final rotation and translation estimation. In this method, four coplanar points are minimum for pose estimation. If there are more than four markers, one of them maybe moved at a time. The CV tracking system detects the moved markers by back projection and uses the non-moved markers for pose computation. Then the CV tracker registers the new position of the moved marker and go to the next tracking step. In this way, the configuration of the 2D markers allows to be changeable. Therefore, each individual movable marker can be used to interact with the virtual objects. Meanwhile, the whole marker set can be exploited for head tracking.

3. Experimental Results

Two sets of experiments were carried out: 1) the computer vision based tracking and the IS-600 were compared for the static case. 2) the error accumulation of the computer vision base tracking due to the movable markers was evaluated.

Fixed 2D markers

In this experiment, we keep both the 2D markers and the HMC static, then the reported poses of the HMC show how big the jitter is. A pure InterSense tracking (STD1, STD3) and a computer vision based tracking (STD2) are shown in the Fig 1.

Movable 2D markers

This experiment shows the error accumulation for the movable 2D markers by the pure computer vision based tracking. We use 6 markers for the experiment. We move one of the markers at a time. After moving all the markers, the position estimation error is around 5 cm. The modified moved marker is going to fit the HMC pose and the HMC pose is optimized to fit the marker reprojections. In other words, the HMC and the markers are adjusted to fit each other, but the HMC pose and marker positions are running away from the true ones. In practice, we set a threshold for determining if the accumulation is big enough for re-estimating all the markers by the InterSense tracker. Once the HMC pose runs away from InterSense tracker exceeding the threshold, the HMC pose will be reset by the Intersense Tracker and the marker positions will be re-estimated for the following tracking.

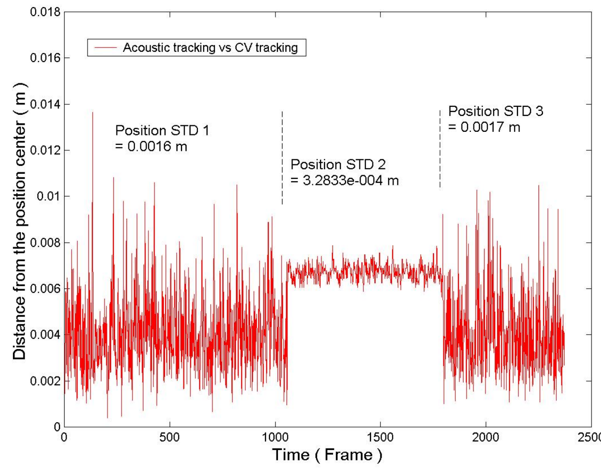

Figure 1. Position jitter of InterSense tracking and computer vision based tracking

4. Conclusions

A head pose tracking system for AR was introduced using computer vision based tracking in combination with an InterSense tracking system. A set of movable 2D markers is used for the homography based pose estimation. The position jitter is about 5 times smaller than that of the IS-600, and the orientation jitter is about 2.4 times smaller.

By integrating with the InterSense tracking system, we take the advantages of the small jitter of the computer vision based tracking without losing the tracking speed of the InterSense system.

References

[1] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355 – 385, Aug. 1997.

[2] E. Foxlin and L. Naimark. Vis-tracker: A werable vision-inertial self-tracker. In *Proceedings of the IEEE Virtual Reality*, pages 199 – 206, 2003.

[3] Y. Liu, M. Störring, T. B. Moeslund, C. B. Madsen, and E. Granum. Computer vision based head tracking from re-configurable 2d markers for ar. *The 12th Danish conferenceon pattern recognition and image analysis* , 2003. http://www.cvmt.dk/∼ yliu/publications.htm

[4] P. Sturm. Algorithms for plane-based pose estimation. *International Conference on Computer Vision and Pattern Recognition*, pages 706–711, 2000.

[5] G. Welch and E. Foxlin. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 6(22):24 – 38, 2002.