Markerless Image-based 3D Tracking for Real-time Augmented Reality Applications

R. Koch, K. Koeser, B. Streckel, J.-F. Evers-Senne
Institute of Computer Science and Applied Mathematics
Christian-Albrechts-University of Kiel, 24098 Kiel, Germany
email: rk@informatik.uni-kiel.de

Abstract

In this contribution we describe a visual marker-less real-time tracking system for Augmented Reality applications. The system uses a fisheye lens mounted on a firewire camera with 10 fps for visual tracking of 3D scene points without any prior scene knowledge. All visual-geometric data is acquired online during the tracking using a structure-from-motion approach. 2D Image features in the hemispherical fisheye image are tracked using a 2D feature point tracker. Tracking may be facilitated by orientation compensation with an inertial sensor. Based on the image tracks, 3D camera egomotion and 3D features are estimated online from the image sequence. The tracking is robust even in the presence of moving objects as the large field of view of the camera stabilizes the tracking.

1 INTRODUCTION

Augmented Reality (AR) systems aim at the superposition of additional scene data into the video stream of a real camera. One can distinguish between offline augmentation for special effects in video post production [3], and online augmentation, where a user typically carries a head mounted display. Additional information is either superimposed directly onto the video stream using video see-through devices or it is projected optically into the visual path of the users gaze direction [1, 2].

The technical and algorithmic demands for online AR are very challenging. The AR equipment must be carried by the user possibly for a long time, hence it should be lightweight and ergonomic and not hinder free movements. At the same time, computation of camera pose must be very fast and reliable, even in uncooperative environments with difficult lighting situation. This will require high computational demands on the system.

Recently, quite some research activities on online AR were undertaken. The work was inspired by the online tracking algorithms from robotics and computer vision. In robotics, the realtime SLAM approach (Simultaneous Localization And Mapping) has been used with non-visual sensors like odometry and ultrasound/laser sensors. These ideas were recently extended to visual tracking [4]. In computer vision, offline AR and visual reconstruction has been in the focus for some years. The dominant approach in this field is termed SfM (Structure from Motion), where simultaneous camera pose estimation, even from uncalibrated cameras, and 3D structure reconstruction is possible [8]. Both approaches have much in common and can be merged towards a versatile realtime AR system [6].

2 ONLINE AR SYSTEM DESIGN

In the following we will describe the components of an online AR system that allows robust 3D camera tracking in complex and uncooperative scenes where parts of the scene may move independently. It is based on the SfM approach from computer vision. The robustness is achieved in two ways:

1. A 190 Degree hemispherical fisheye lens is used that captures a very large field of view of the scene. If used in indoor environments, the hemispherical view will always see lots of static visual structures, even if the scene in front of the user may change dramatically. The system is therefore mainly designed for (but not restricted to) indoor use, because in outdoor scenes the sun light falling directly onto the CCD sensor will cause problems. These problems can be facilitated when CMOS sensors with logarithmic response and high dynamic range are used.

2. The 3D tracking is based on robust camera pose estimation using structure from motion algorithms [8] that are optimized for realtime performance. These algorithms can handle measurement outliers from the 2D tracking using robust statistics.

2.1 System

The goal of the AR system is a light-weight wearable solution that allows realtime augmentation via a HMD without obstructing the user motion. The computational load of such a system for simultaneous realtime tracking and augmentation is too high to be performed on currently available wearable computers. We have therefore designed the system with a lightweight wearable unit for the head-mounted display and the image acquisition, that is connected to a back-end PC via a wireless LAN access. In this contribution we are only concerned with the recording and tracking unit and do not handle augmentation.

The video camera system must be extremely small. We have chosen a 640x480 firewire camera with 12 mm microlens adapter and a microlens fisheye. The image quality of the fisheye lens degrades towards the boundary of the hemisphere, therefore the opening angle is reduced to 160 degree and a quadratic subimage with 400x400 pixel is processed, resulting in an angular resolution of 3 pixel/degree. The backend system is currently able to process 10 fps, thus a raw data rate of 1.6 MB/s is transferred through the WLAN channel.

In addition, the camera rotation is measured using a 3 DoF inertial sensor at 100 Hz rate. The rotation data is used to compensate fast head rotations and to predict image feature positions.

The backend system runs two separate threads (possibly on a 2-processor unit) that separate the 2D feature tracking from the 3D SfM pose and structure computation. The estimated 3D pose is handed back to the wearable unit and visual augmentation is superimposed onto the user view.

Figure 1 gives an overview on the system components.

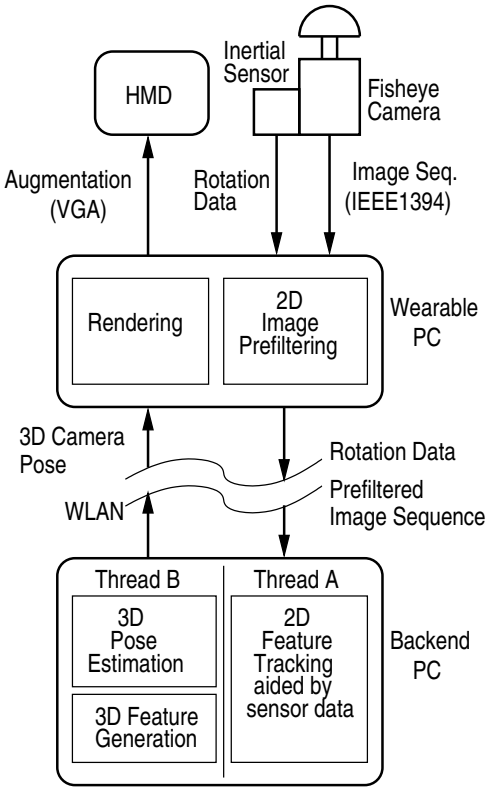

Figure 1: Overview of AR system.

2.2 Robust tracking from fisheye images

The 3D tracking system is divided into tracking initialization, 2D feature tracking and robust 3D pose estimation. The tracking is facilitated by the 3DoF inertial rotation sensor.

Initialization and 2D feature tracking: In an initial step, a set of salient 2D intensity corners are detected in the first image of the sequence. These 2D features are then tracked throughout the image sequence by local feature matching with the KLT operator [9]. If feature tracks are lost, new tracks are constantly reinitialized. The new tracks are merged with previous tracks in the 3D stage to avoid drift.

As we are handling spherical images from the fisheye lens, care must be taken to compensate the spherical distortions using local planar rectification. To further facilitate 2D matching, the 3D camera rotation velocity is measured by the inertial rotation sensor and the rotation is compensated in the images. Currently we compensate the rotation only, but a parallax compensation by backprojection of 3D features is planned.

3D feature tracking and pose estimation: From the given 2D feature tracks, a SfM approach [7] can be applied to estimate the metric camera pose and 3D feature positions simultaneously. Given a set of reliable 2D features, the Essential matrix between the views can be computed and the relative pose of the cameras can be extracted. Simultaneously, 3D feature points can be triangulated from the given 2D correspondences and the relative pose. The camera pose and the 3D feature positions are determined with a rotation relative to an initial camera position and up to an unknown overall scale. This scale must be inserted into the system from external data. The SfM is based on the assumption of a rigid scene where the estimated 3D features do not move between views. Therefore, care must be taken to handle moving objects and measurement outliers robustly.

Robustness of the estimation is introduced by robust statistical evaluation of feature matching and Essential matrix computation using RANdom SAmpling Consensus [7]. Thus, moving objects are treated as measurement outliers

that are discarded by the RANSAC. The tracking from fisheye cameras leads to an especially robust tracking for two reasons:

1. The wide field of view covers a very wide scene area and moving objects tend to be only in a small part of the scene. Therefore, most of the visible scene is static. Second, a camera mounted on a human head is subject to large and jerky rotations. This rotations are partially compensated by the rotation sensor, but still the head may rotate the camera quickly out of view. This will not happen easily with a fisheye camera with hemispherical view.

2. It can be shown that a wide field of view stabilizes the pose estimation [5]. For perspective cameras with small field of view, the motion towards the optical axis is always ill defined because the camera moves towards the focus of contraction (FOC). Only the motion perpendicular to the FOC can be estimated reliably. In a spherical image, there will always be an image position that is perpendicular to the FOC, hence the estimation of the camera motion is always reliable.

   A drawback of the spherical image is the low angular resolution of the image, hence the estimate for a fisheye lens camera will be less accurate than estimates from a sideways moving perspective camera with high angular resolution.

3 EXPERIMENTS

We have performed extensive experiments with the system. In the following section we will give some results on timing and on visual camera tracking.

Figure 2 shows camera tracking results of a sequence of 900 views. The camera was moved extensively through space and even rotated up to 180 degrees away from the initial pose. Still, tracking was possible since the wide field of view allowed that 3D features were visible for a long time.

To evaluate the timing, either 50 or 100 features were tracked on a 400x400 pixel image using a 3.0 GHz P4 with single and double processor PC. 2D and 3D tracking can be separated into 2 tasks that run either concurrently on the 1-CPU or parallel on the 2-CPU PC. The timing table in table 1 shows that 10 fps are indeed possible in this configuration for the 2-Processor PC.

No. feat.	2D	3D	1-CPU	2-CPU
50	30	68	98	68
100	40	105	145	105

Table 1: Timing of tracking per frame in ms.

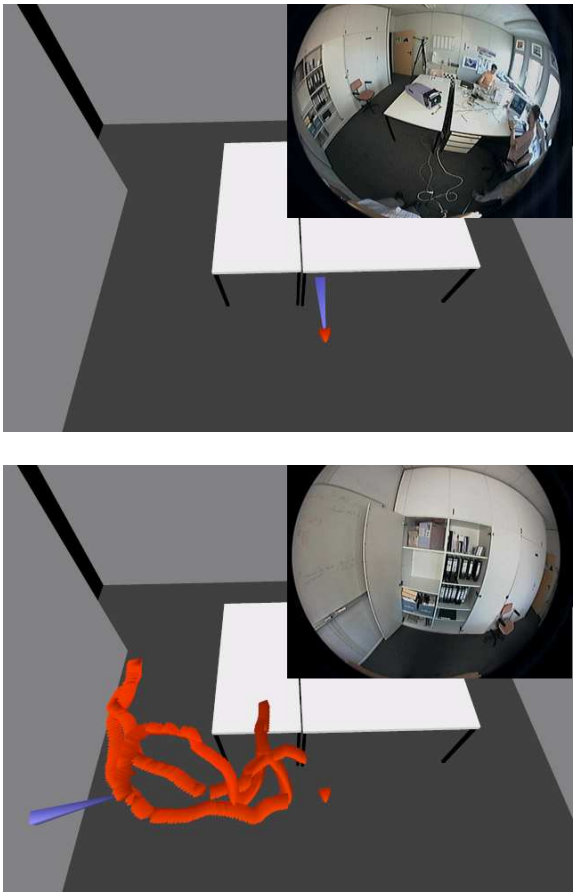

Figure 2: Visualization of 3D camera tracks (image 1 and image 650) with the original fisheye image in the upper right corner.

Figure 3 shows augmentation results, where the central section of the camera was mapped to a planar view and synthetic objects were placed on the real table. The objects remained in their allocated place without much jitter.

4 CONCLUSIONS

The presented approach shows that a robust markerless 3D tracking from a fisheye camera system is possible in realtime. The system presented is in an early stage and further fine-tuning is needed. The 3D processing is not yet optimized to speed and we foresee still some potential in this stage. Currently, there is no feedback from the 3D features into the 2D stage. A proper prediction from the full 6 DoF state of the system will enhance the current 3 DoF prediction. The covariances of the 3D features are currently evaluated numerically which is a costly operation. An analytical solution will further enhance speed. Furthermore, there is currently no solution on the computation of the ab-

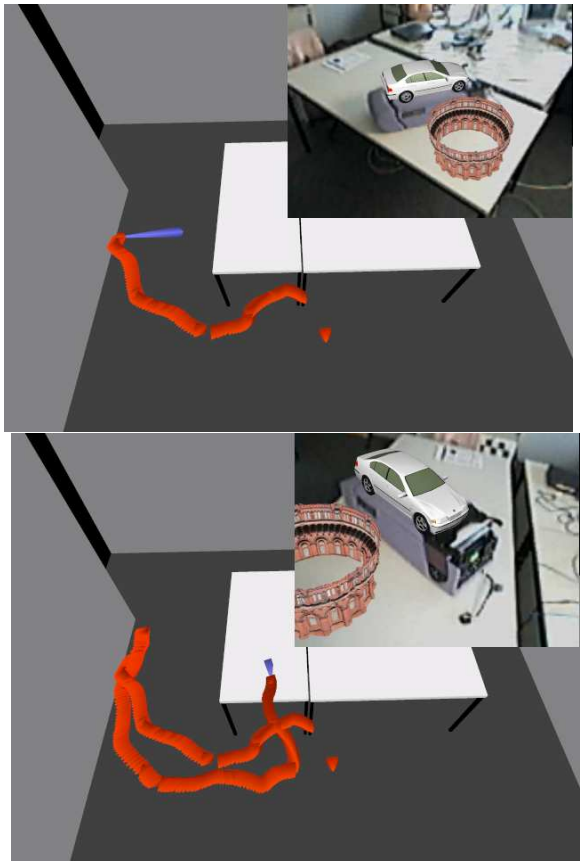

Figure 3: Visual augmentation of virtual objects in a real scene superimposed on central part of images 200 and 500.

solute scale of the reconstruction. Augmentation will need the transformation into the Euclidean world. Finally, we need better reinitialization strategies for the 2D tracks by recognizing objects and salient features from image data to minimize drift.

Acknowledgments

This work was funded partially by the German Ministry of Science project BMBF-ARTESAS and the European Commission project IST-2003-2013 MATRIS.

References

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

[2] R. Azuma, Baillot, Behringer, Feiner, Julier, and MacIntyre. Recent advances in augmented reality. In *IEEE Computer Graphics and Applications, Vol. 21 , No. 6*, pages 34–47, Nov. 2001.

[3] G. Bazzoni, E. Bianchi, O. Grau, A. Knox, R. Koch, F. Lavagetto, A. Parkinson, F. Pedersini, A. Sarti, G. Thomas, and S. Tubaro. The ORIGAMI Project – advanced tools and techniques for high-end mixing and interaction between real and virtual conten. In *IEEE Proceedings of 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, 2002.

[4] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings International Conference Computer Vision, Nice*, 2003.

[5] A. J. Davison, Y. G. Cid, and N. Kita. Real-time 3D SLAM with wide-angle vision. In *Proc. IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon*, July 2004.

[6] A. J. Davison, W. W. Mayol, and D. W. Murray. Real-time localisation and mapping with wearable active vision. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Tokyo*, 2003.

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge university press, 2000.

[8] M. Pollefeys, R. Koch, and L. J. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.

[9] J. Shi and C. Tomasi. Good features to track. In *Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, June 1994. IEEE.