

Model-based Tracking with Stereovision for AR

Hesam Najafi

Gudrun Klinker

Computer Science Department, Technische Universität München

Boltzmannstraße 3, Garching bei München, Germany

{najafi|klinker}@in.tum.de

Abstract

This demo shows a robust model-based tracker using stereovision. The combined use of a 3D model with stereoscopic analysis allows accurate pose estimation in the presence of partial occlusions by non rigid objects like the hands of the user. Furthermore, using a second camera improves the stability of tracking and also simplifies the algorithm.

Keywords

Visual Tracking, Marker-less Tracking, Model-based Tracking, Stereovision

1. Introduction

In this Demo we show a robust model-based tracker using stereovision that can handle partial occlusions. While the offline 2D-3D registration problem can be considered as essentially solved, robust real time tracking is still an open issue.

Model-based visual tracking methods rely on a 3D model of the target object(s) and try to compute a 3D pose that correctly re-projects the features (e.g. points, edges, line segments) of a given 3D model into the 2D image. After an initialization of the system where initial matches between 2D features and the corresponding 3D model features are established, 2D tracking algorithms like KLT [4] are used to track the features from frame to frame [3][7][1]. The resulting 2D-3D matches of every frame are used for pose estimation using standard methods e.g. Tsai's algorithm [6].

There are mainly three major problems arising when tracking 2D features. First, the matched features may drift or even be completely wrong (*outliers*) when pixel-based correlation techniques establish frame-to-frame correspondences. Secondly, since the camera can move, the initial features may not be visible all the time, so new features need to be found and tracked properly at run time. And thirdly, due occlusions e.g. caused by user's hands, the target object may be only partially visible.

To handle these problems new visible features need to be added and matched online. Several approaches to monocular tracking [3][7] try to detect and discard outlier matches by using robust estimation techniques like RANSAC, which tend to be quite time consuming.

We show a tracker that incorporates stereoscopic vision with the 3D model for this purpose. Stereoscopic analysis provides the important epipolar constraint [2]. By applying this constraint to the stereo image pair, outliers (false correspondences from monocular tracking) can be easily detected and rejected, avoiding the more time consuming robust estimation techniques. Using a second camera also improves the accuracy of the tracking. Furthermore, it improves greatly the stability and simplifies the algorithm.

The requirement of an existing 3D model is, in practice, not an issue since such models already exist in many applications or can be created using either automated techniques or commercially available products. In this demo we used the 3D modeling program *Canoma* based on single images taken from the target objects.

2. System Overview

Two cameras are mounted side by side as a stereo camera system. We use the method proposed by Zhang [8] to determine the intrinsic matrices of both cameras and the relative transformation between them.

For the initialization we let the user select several feature points interactively in each image, which are then locally refined to satisfy the epipolar constraint, and the corresponding model points. Using these 2D-3D correspondences, the model is registered with the images. From then on, the system tracks the optical flow of salient features using the KLT tracker [4], rejects outliers based on the epipolar constraint, and updates the pose of the camera in every frame (Figure 1). A feedback loop supplies new salient feature points in each frame to make the tracking more stable under various conditions, e.g. occlusions by the user's hands. The algorithms are described in more detail in the following two sections.

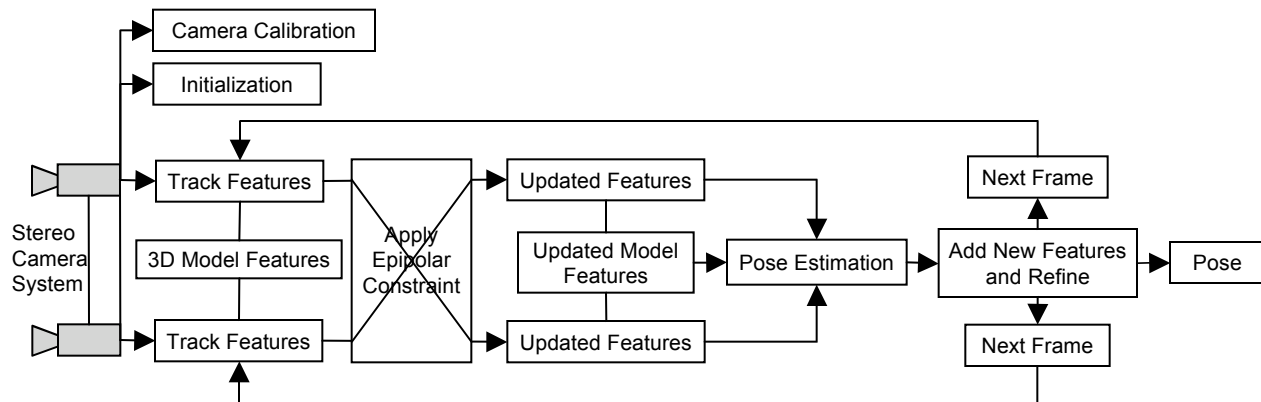


Figure 1. Overview of the model-based stereo tracking system.

3. Stereo 3D Tracking

First we detect strong corners in both images using the Shi-Tomasi algorithm [5]. The feature points are then tracked in each camera image independently using a pyramidal implementation of the Lucas-Kanade algorithm [4]. However the matched points may drift or even be wrong. Therefore, the epipolar constraint is applied to reject outlier matches. The epipolar constraint states that if a point p in the first image and a point q in the second image correspond to the same 3D point in the real world, they must satisfy the following equation $q^T F p = 0$, where F is the *fundamental matrix* that is the algebraic representation of the epipolar geometry between two images [2]. This equation means that point q must pass through the epipolar line defined as Fp in the second image and vice versa.

By applying this constraint to the stereo image pair, outliers can be easily rejected and the pose of the camera is estimated using Tsai's algorithm [6].

3.1. Adding New Features

To ensure that there will be a sufficient number of feature correspondences for pose estimation in the consecutive frames, new stable features are added based on the following two criteria:

Texture: The feature point has rich texture information. For this purpose strong corners are selected from the image based on [5], then back-projected to the 3D model in order to get their 3D coordinates, considering only the interest points that are on the object surface.

Visibility: The feature must be visible in both cameras. A feature point is visible in both images if the respective 3D model points are close enough to each other.

Adding new features by every frame alleviates the tracker drifting problem and improves the accuracy and stability of the tracker.

4. Summary and Conclusion

This demo shows an efficient and robust stereovision tracker for AR applications that can handle partial occlusions. Using a second camera improves the accuracy and the stability as well as simplifies the algorithm.

5. References

- [1] R. Behringer, J. Park, V. Sundareswaran. *Model-Based Visual Tracking for Outdoor AR Applications*, ISMAR 2002.
- [2] R. Hartley and Z. Zisserman., *Multiple View Geometry in Computer Vision*, Cambridge New York, Cambridge University Press 2000.
- [3] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar, N. Navab. *Marker-less Tracking for AR: A learning-Based Approach*, ISMAR 2002, Darmstadt, Germany
- [4] B.D. Lucas, T. Kanade. *An iterative image registration technique with an application to stereo vision*. In Proc. Int. Joint Conference on Artificial Intelligence, pages 674-679, 1981.
- [5] J. Shi, C. Tomasi. *Good Features to track*. In Proceedings of the IEEE Conference on Computer Vision and pattern Recognition, pages 593-600, Seattle, WA, June 1994.
- [6] R.Y. Tsai. *A versatile camera calibration technique for high-accuracy 3D machine vision metrology using of the shelf TV cameras*. IEEE Journal of Robotics and Automation, RA-3(4):323-344, 1987.
- [7] L. Vacchetti, V. Lepetit, P. Fua. *Fusing Online and Offline Information for Stable 3D Tracking in Real-Time*.
- [8] Zhengyou Zhang. *A Flexible new technique for camera calibration*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.