Review: Markerless Tracking using Planar Structures in the Scene

Wei Zhang
Department of Computer Science
University of California, Santa Barbara
wei@cs.ucsb.edu
http://www.cs.ucsb.edu/~wei

Abstract

In this paper, the author describe a vision-based position tracker for augmented reality that operates in unstructured environment. This simplifies the general camera-tracking problem by only requires one or more planar surface visable in the scene, which is a common special case. In addition, the tracked plane represents a natrul reference frame so that the alignment of real and virtual coordinate systems is simplified as well. The author also provide an implementation of a practical and reliable tracker to show its validity. Their method grasps the key of markerless tracking: detect and track something in the image as an alternative of markers. Therefore finding a planar surface may be a pretty straightforward solution. However, their method is also limited by this point. Futhermore, their implementation is not completely automatic, it requires the user to choose 4 points of a rectangular on the plan manually during the initialization.

1. Introduction

1.1. The Problem

Estimating the pose and movement of a camera (virtual or real) in which some augmentation takes place is one of the most important parts of an augmented reality (AR) system. Legacy vision-based trackers are based on tracking of markers. The use of markers increases robustness and reduces computational requirements. However, their use can be very complicated, as they require certain maintenance. Therefore, directly using the scene features for tracking is desirable.

1.2. Some Previous Work

Now the problem turns to be which feature to use and how. The most common approach is model-based tracking, which exploit some structures in the scene that their measurements are already known. This method provide precise result, but is greatly limited by requiring the manual intervention to construct the model.

Another technology that provides general, accurate registration is known as structureand-motion estimation, or move-matching. It simultaneously estimate camera motion, and the 3D structure of the imaged scene. The major limitation is that it needs a lot computation and a bunch of adjustments, so that cannot be used in real-time application.

1.3. This Paper

In this paper, the author try to provide a markerless solution with only slight limitation on the type of scene. Some previous works had already done something similar such as using sets of parallel or orthogonal lines in the scene to determine the 3D reference frame. So their idea of using planar surface doesn't contribute too much. But they propose an efficient algorithm and a very clear framework to track the planar surface and do camera recovery fast and reliably.

In general, I think this paper is very well-written and easy to read. They first give a brief introduction to the problem and previous works, all with cons and pros. Then they briefly introduce the main idea of this work, and give a algorithm summary to show how it looks like. After that they describe every part of the algorithm in detail, the way that they explain is pretty easy to understand, but they do forget to add the equation numbers. Then they show us an implementation and some results as an example, and finally they discuss the advantage and limitation of this method.

2. Discussion of the Approach

2.1. Main Idea

The key idea of this paper is to track the planar surface in sequential images. By matching the interesting points between two images, the RANSAC algorithm gives out the planar homography between them. Therefore, if the planar homography between a planar surface in the world and in one of these images are determined, then problem is solved. This paper suggests to do this manually on the first image during initialization, by specifying a rectangular on a plane in the image.

2.2. Pros

There are several points that I like in this paper:

1. The choice of planar surface is a good idea, because such plane exists in most scenes that contain man-made objects.

2. They let the user to select a plane, and make it the XY plane in the world coordinate system. So all points on this plane have z equals zero, this makes future computation easier.

3. They give out a comprehensive explanation on how to compute the planar homography transformation matrices, as well on other steps, which makes their algorithm easy to repeat.

4. They argue that the best homography for the whole image(found by RANSAC) will correspond to the homography for the largest plane in the image. I'm not sure how robust it is, but if it's true then the hand-off precedure could be easier to implement.

5. The demos that they show is pretty impressive.

2.3. Cons

But there are more things that I have question with:

1. There have no equation numbers in the paper, that's a bad habit since this paper has more than 20 equations.

2. In *section 2.3*, they mention a little bit on some related works that detect parallel or orthogonal lines in the image to improve AR, and then skip them rapidly. I think these works are much more related to this paper, compare to what they have discussed in *section 2.1* and *section 2.2*. I think they should not intend to avoid the comparison with other competitive works.

3. In *section 3.6* they state that the '*hand-off*' procedure is easy, without giving enough details on how. And it becomes even more confusing when they admit in the conclusion that their system '*might fail*', '*when the tracked plane goes out of view*'. The problem here is that they doesn't explain how to choose the new plane and find three points on it. Is it done by RANSAC? Can RANSAC provide such robust result? Is there any experiment data to support this?

4. They rely on the RANSAC to compute the planar homography, which is the key part of their computation. But I can't find any experimental data analysis about this, especially about the accumulative drift. Besides, they doesn't really give a solution to solve this.

5. Finally, they didn't provide experimental analysis of their algorithm, thus I can't tell how good or bad this algorithm really is, nor compare to other related works.

3. Experiment

There are short demos, but no experiment or data analysis actually.

4. Conclusion

This paper propose a simple, fast and reliable solution to markerless camera tracking problem. Their major contributions are the mathematical framework for uncalibrated plane tracking and camera recovery.

However, the limitations of their solution are obvious:

- There must be more than one plane in the image, and must be detectable;
- The system might fail if the tracked plane move out of view;
- The system rely on RANSAC to compute the homography between images, but sometimes this is not reliable, for example, to track a big plane with many small rectangulars which all in the same size and pattern, such as a chessboard;
- How to do hand-off is still in question.

In order to make their argument stronger, the author need to:

- Give more details at a few points;
- Provide experimental data analysis;
- Show comparison with similar works;
- Finally and most important - add the equation number.