

3D Fiducials for Scalable AR Visual Tracking

John Steinbis*

William Hoff†

Tyrone L. Vincent‡

Center for Robotics Automation and Distributed Intelligence
Colorado School of Mines

ABSTRACT

A new vision and inertial pose estimation system was implemented for real-time handheld augmented reality (AR). A sparse set of 3D cone fiducials are utilized for scalable indoor/outdoor tracking, as opposed to traditional planar patterns. The cones are easy to segment and have a large working volume which makes them more suitable for many applications. The pose estimation system receives measurements from the camera and IMU at 30 Hz and 100 Hz respectively. With a dual-core workstation, all measurements can be processed in real-time to update the pose of virtual graphics within the AR display.

Keywords: Augmented Reality, Pose Estimation

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces

1 INTRODUCTION

A key challenge in augmented reality is accurately estimating the pose of the display in real-time with respect to the environment. For large scale outdoor systems (*e.g.*, a navigation aid for a city) it may be possible to use GPS and compass and/or tilt sensors [2, 1]. However, these types of sensors are not accurate enough for smaller scale applications or indoors, where the range to the objects of interest is small.

In this case, cameras and computer vision-based techniques for tracking features in the environment are attractive due to their accuracy and flexibility [5, 3, 6]. Ideally, we would like to track features that are already present in the scene. However, robust segmentation and tracking of features in unstructured environments is a very difficult task. To simplify the problem it is common to use specially designed fiducial targets, which have unique features that can easily and reliably be extracted from images.

One type of target is a planar pattern with high contrast markings, like a calibration grid. The markers used in the popular AR-ToolKit [4] are squares with black and white patterns that can be easily detected even under low lighting conditions. A single marker is sufficient for full 6 degree of freedom pose estimation.

Planar targets have been widely used in small scale, table-top type applications, where the objects are at arm's length [7, 8]. However, they may not be ideal for medium scale or outdoor applications. Scaling up the size of the workspace requires scaling up the size of the fiducial targets, so that they are visible at a distance. The target should also be visible from a range of viewpoints around the workspace.

An example of a medium scale application is shown in Figure 1. We developed a pose estimation system for tracking a C-130 airplane, in order to overlay virtual graphics (such as smoke, propellers spinning, etc) on the real airplane. The application was to

simulate dangerous scenarios during engine startup and train people in the procedures to deal with them, without risks to personnel and equipment.



Figure 1: C-130 with cone fiducial targets and virtual overlays.

In this paper, we describe a complete augmented reality system that uses ordinary traffic cones as 3D fiducial targets to track pose in real-time. The system operates as video see-through, with the user viewing virtual images overlaid on real world images taken by a camera. Although head-mounted operation is also possible, for testing and analysis our implementation uses a small hand-held display, with an attached camera and inertial measurement unit (IMU).

2 3D FIDUCIAL TARGETS

A planar target could be enlarged in order to track a large scale object (such as the C-130 airplane). So that the person can walk around the object, it is desirable that the targets be visible from any angle. One approach would be to place the target flat on the ground.

A planar target is best viewed from the direction normal to the surface of the target. In order to resolve the black and white shapes on the target, assume that the image of the shape on the camera plane needs to subtend a minimum angle of $\Delta\theta$ as measured at the camera center. Assuming the physical shape has minimum extent along some dimension of size L , then it can be shown that the viewing distance d at which the feature can be resolved is

$$d = \frac{L \cos \phi}{\Delta\theta}, \quad (1)$$

where ϕ is the viewing angle from the surface normal (assuming that $d \gg L$). The maximum distance occurs when the target is viewed from the direction normal to its surface. However, in cases where the viewer observes the target from a more oblique angle, we are limited to a shorter distance.

A 3D fiducial target may have an advantage over a planar target in these situations. Consider an ordinary traffic cone. The orange color is distinctive, so it can be easily segmented from the image. Consider a small surface patch ΔA on the cone. Assume that in order to segment the cone from the image, it needs to subtend a solid angle of $\Delta\Omega$ at the camera. Then the distance at which we can segment the cone is

$$d = \sqrt{\frac{\Delta A \cos \phi}{\Delta\Omega}}. \quad (2)$$

In this case, the maximum distance is along the direction perpendicular to the surface normal of the cone. This is more oblique to

*e-mail:jsteinbis@mines.edu

†e-mail:whoff@mines.edu

‡e-mail:tvincent@mines.edu

the ground, which is desirable in cases where the viewer is more distant to the target.

The “working volumes” of the planar target and the cone target are shown in Figure 2. As one can see, the shape of the volume for the cone target is more suitable for an application where the viewer is more distant from the target. Note that these are approximate working volumes, as more than one cone, or more than one image patch, are needed to fully determine pose.



Figure 2: “Working volumes” for planar target (left) and cone target (right).

Other 3D fiducials can be considered, but cones have several properties that make them ideal for pose estimation. Since they are symmetric around one axis, the image of the cone at a given elevation is the same from any direction, simplifying the segmentation problem. Indeed, the cone will either appear as a triangle with a rounded bottom or a circle. If it was desired to make the segmentation problem even easier, a sphere could be chosen, which always projects to a circle on the image plane. However, it is impossible to identify the orientation of the sphere, while for a cone viewed from the side, the two dimensional orientation in the image can also be measured, which is the projection of the cone’s vertical axis onto the image plane. Although this is insufficient to estimate the full pose of the camera from a single cone target, two cones are sufficient.

2.1 Cone Segmentation

The goal of the cone segmentation procedure is to automatically detect the 2D image coordinates of every cone’s peak, middle, and base points within the video stream when viewed from the side. Direct overhead view would require a different segmentation procedure not discussed here. As shown in Figure 3, the procedure may be decomposed into four distinct steps.

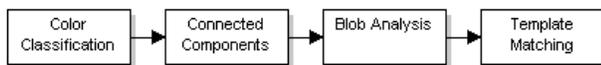


Figure 3: Cone Segmentation Procedure

The segmentation procedure relies heavily on the cone’s unique orange color, not usually found in most environments. Lighting conditions will however affect the color that appears in an image, but robust color classification can tolerate normal light fluctuations. The input to the color classification block is an RGB image and the output is a binary image. Pixels that are judged to be the color of an orange cone are assigned a one and zero otherwise.

The connected components step defines which pixels are connected to other pixels within a binary image. These blobs are either accepted or rejected by their geometric properties. Template matching is performed to refine the estimate of location, height, and orientation from the quick blob analysis. The templates are very simple and contain a single white triangle with black background. The search window is a rectangle that is 10% greater area than the bounding box of the blob.

2.2 Implementation

A real-time implementation using off-the-shelf components was developed, using an Extended Kalman Filter (EKF) to establish pose estimates from cones placed in a known triangular configuration. A gyro is also used, which can improve dynamic performance. The system runs at the frame rate of the camera, which is up to 30 Hz in good lighting conditions. Example overlays are shown in Figure 4.



Figure 4: Virtual 3D Models

In Figure 5 the pose estimates for this trajectory are plotted in one side and one overhead view. The estimated location and direction of the display is shown. Note that a large hemisphere around the targets is covered. In fact, 360° operation is possible, but we are physically limited by the cables running to the display. This cables also restricted the maximum distance from the cones.

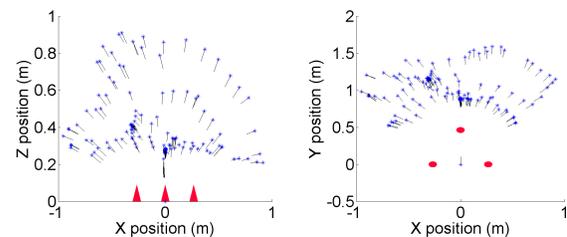


Figure 5: Three views of test trajectory with nominal configuration. The x-y plane is parallel to the ground. Estimated is indicated by the blue star, the direction with a line

ACKNOWLEDGEMENTS

This work was supported in part by Pathfinder Inc. and National Science Foundation grant ECS 01-34132.

REFERENCES

- [1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, pages 34–47, 2001.
- [2] R. Azuma, B. Hoff, H. Neely III, and R. Sarfaty. A motion-stabilized outdoor augmented reality system. In *Proc. IEEE Virtual Reality Conference (VR'99)*, page 252, 1999.
- [3] A. I. Comport, E. Manchard, and F. Chaumette. A real-time tracker for markerless augmented reality. In *Proc. ISMAR'03*, pages 36–45, 2003.
- [4] H. Kato, M. Billinghurst, B. Blanding, and R. May. ARToolKit. Technical report, Hiroshima City University, 1999.
- [5] U. Neumann and S. You. Natural feature tracking for augmented reality. *IEEE Transactions on Multimedia*, 1.
- [6] G. Reitmayr and T. W. Drummond. Going out: Robust model-based tracking for outdoor augmented reality. In *Proc. ISMAR'06*, 2006.
- [7] D. Schmalstieg, A. Fuhrmann, G. Hesina, and Z. Szalavári. The student-stube augmented reality project. *Presence*, 11(1):33–54, 2002.
- [8] D. Wagner and D. Schmalstieg. First steps towards handheld augmented reality. In *Proc. Seventh IEEE Symp. Wearable Computers*, pages 127–135, 2005.