Camera Calibration for Video See-Through Head-Mounted Display

Mike Bajura

July 7, 1993

Abstract

This report describes a method for computing the parameters needed to model a television camera for video see-through head-mounted display applications.

1.0 Introduction

In a video see-through head-mounted display system, computer imagery is mixed with real imagery from a standard television camera mounted on a conventional head-mounted display. The combined imagery is shown on the head-mounted display allowing the user to ‘‘see through’’ the head-mounted display into his natural environment. As the user moves around his real surroundings, the computer-generated imagery is updated to give the illusion of virtual objects occupying real-world space. To make this illusion convincing, the real-world imagery and the computer-generated imagery must be correctly registered with respect to each other. Once the images are registered, they can be combined digitally on a pixel-by-pixel basis or with an analog video mixer.

Two things must be known to align real-world images from a head-mounted television camera with computer-generated images: 1) the position, orientation, and field of view of the television camera , and 2) the distortion function of the camera itself. These two sets of parameters are independent of one another and can be deduced separately. Section 2 describes a method for finding the position, orientation, and field of view of a video camera mounted on a tracked head-mounted display. Section 3 describes a method for method for finding a function to transform the distorted images of a video camera into ones from an ideal pinhole camera. The effects of camera distortion tend to be greater in wider angle lens systems and may possibly be ignored in lens systems with sufficently small distortion.

2.0 Finding Camera Position, Orientation and Field of View

2.1 Theory

In common head-mounted display systems, a tracking system is used to report the position and orientation of the head-mounted display. When a television camera is **firmly** attached to the head-mounted dipsplay, its position and orientation can be modelled by a constant translation and rotation from the reported position of the tracking system. The problem of finding a camera's position and orientation is thus one of once finding its fixed position and orientation relative to the reported position of the tracking system and then transforming the reported position of the tracking system by the fixed camera translation and rotation for each new camera position. The field of view of a camera is constant and need not be updated once it is known. These parameters can be found by a two-step process. First, a close guess can be computed automatically. Second, the parameters can be independently adjusted for a better result.

The idea for locating a camera's position relative to the origin of the head-mounted display tracking system is to use the tracking system itself as a measurement tool. The tracking system can be used to fix an absolute position and orientation in both real-world and tracking system coordinate spaces. The camera can then be placed in that fixed position and orientation at the same time the tracking system position and orientation is noted. The camera position and orientation relative to the tracking system can be computed as the difference between the reported position of the tracking system and the position and orientation of the camera. An initial field of view can be estimated from lens specifications for the television camera.

Corrections to the camera position, orientation, and field of view can be made by operating the see-through system using the approximate camera position and orientation. Ideally when a virtual target and a real target are positioned at the same coordinates in both real-world and virtual-world space, they will intersect when viewed through the see-through system. If they don't, that information can be used as feedback to adjust the fixed camera position, orientation, and field of view. In the extreme limit, these parameters are independent when different views of the real and virtual targets are taken. In practice, they are still relatively independent. The following rules apply to adjustments of the camera parameters:

> - When the camera is positioned relatively **far** from the intersecting targets, misregistration is most sensitive to errors in camera orientation. Rotation about the camera X and Y axes will align the target positions, rotation about the camera Z axis will align the target orientations.

-When the camera is positioned relatively **near** the intersecting targets, misregistration is most sensitive to errors in the camera position. Translations along the camera X and Y axes will align the target positions. Translations along the camera Z axis will move the camera either in front of, or behind the virtual target when the camera is very near the real target.

- When the targets are aligned along the optical axis (center of image) of the camera, errors in the field of view of the camera will result in a scaling of the virtual target but will not change its position. This means that camera position and orientation can be adjusted even if the camera field of view is not known.

- Once the position and orientation of the camera are correct, errors in target registration off the camera's optical axis are due to errors in the camera field of view. Off-axis registration can be adjusted by independently modifying either the vertical or horizontal field of view. This amounts to adjusting for a ‘‘1st order correction’’ in the distortion of the camera lens. If the lens has significant distortion, either the camera images or the computer-generated images must be corrected for the camera lens distortion (see Section 3) before the field of view can be accurately adjusted.

2.2 Implementation

The implementation follows the theory outlined above. An XYZ coordinate axis is defined and is used as the reference fixed point, reference orientation, and reference imaging target. An XYZ coordinate axis is defined in world space and tracking space by constraining the tracking system to rigid body rotations alternately about both the X and Y coordinate axes. From the positional and rotational information reported by the tracking system during rotation, the X and Y axes are computed. A Z axis is computed as normal to the X and Y axes at their intersection. Because a rigid body rotation is performed, the offset of the tracking system origin from the axis of rotation need not be explicitly known. Measurement of the coordinate axes is improved by rotating through larger angles closer to the tracking system origin and by taking more ovservations. It is important that the coordinate axis measurement be as accurate as possible since it is the basis for the rest of the calibration. By taking careful measurements, it is possible to locate this axis with nearly the same accuracy of the tracking system itself.

The initial guess for camera position and orientation is a two measurement process relative to the fixed coordinate axis. Empirically it is earier to separately place the camera focal point at a fixed position and then at a fixed orientation relative to the coordinate axis rather than trying to fix position and orientation simultaneously. The position of the camera focal point is found by positioning the camera on the assumed Z axis. This measurement can be quite accurate if the camera lens is placed directly on the

X-Y plane and the camera focal point is taken to be a measured positive value along the Z axis. When the camera is positioned, the difference between the reported position of the tracking system and the position of the camera focal point is computed. To compute the camera orientation, the camera is positioned to look down the positive Y axis with the Z axis pointed ‘‘up,’’ much like a boresight. When the camera is appropriately positioned, the difference in orientation between the tracking system and the positive Y axis is computed as the fixed camera orientation. An initial guess is made for the camera field of view based on data for the lens system.

The camera parameters are improved by an interactive process guided by the ‘‘rules’’ in sec 2.1. The video see-through system is run with a virtual XYZ coordinate axis displayed in the same position and the real XYZ coordinate axis. First, the rotational parameters for the camera are adjusted by positioning the camera far from the coordinate axis and modifying the camera rotation until the real and virtual axes are aligned in the center of the combined image. The translation parameters are adjusted similarly but with the camera positioned very near the coordinate axes. Once the camera position and orientation are fixed and the coordinate axes are aligned along the center of projection for both near and far views, the field of view is adjusted for off-center views of the coordinate axes.

3.0 Camera Distortion Correction

3.1 Theory

A perfect, distortion free, or pinhole, camera performs a simple projection from 3D points to a 2D camera image. The basic ‘‘rule’’ of projective geometry is that straight lines remain straight under projection. No guarantees about relative size or parallel lines not meeting are made. If an image can be corrected so that lines in it which are supposed to be straight, actually are, then the resulting image is a projection from a pinhole camera and the correction made approximates the distortion of the imaging system. There can be many sources of distortion in a lens system, but the dominant sources of error can be modelled as a radial stretching or shrinking function centered at a particular origin. This makes sense given that lenses are accurately ground to be radially symmetric.

3.2 Implementation

The camera is focused on a rectangular field of dots. An image processing algorithm is use to locate the centers of the dots. A line finding heuristic is used to connect the dots into a warped rectangular grid. Lines are fit to the warped lines of the grid. The error in the fit of the lines to the points is used as a metric for a multi-dimensional distortion parameter finding algorithm. The parameters of the algorithm are the coeffients of radial warping of the image and the offset to the center of radial warping. The parameters corresponding to the minimum error are taken to be the best approximation to the distortion of the camera.

The line finding heuristic starts by finding a point close to the center of the image. The point nearest to that point is taken to be in either the unit X or Y direction, whichever it is closest to. The distance between those points sets a search neighborhood for other points which can be connected to a given point. Either the Y or X direction is inferred as normal to the direction of the 1st pair of points located. The image is then searched for pairs of points within the search neighborhood which are within some tolerance of orientation (angle in degrees) to either the X or Y directions. These pairs of points, or line segments, are ‘‘grown’’ along their preferred directions to form lines which span the image.

Once the lines are located, a radial warping function using up to the 5th power of the radius at an arbitrary center is found which minimizes the error of the lines from being straight. A simplex algorithm converges relatively quickly and achieves good empirical results though no tests for absolute distortion correction or modelling of a lens distortion equation are attempted.

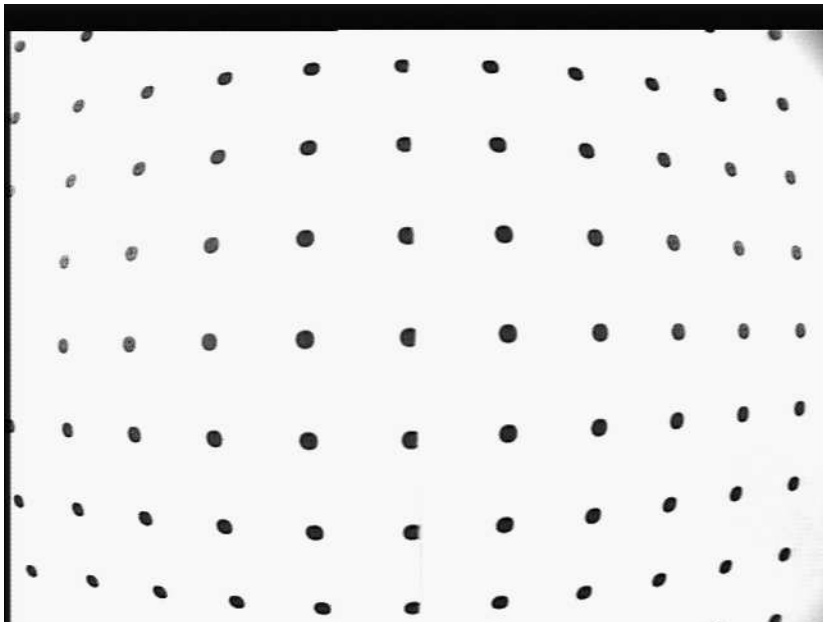

Fig. 1 Uncorrected distortion of 110 degree field of view lens

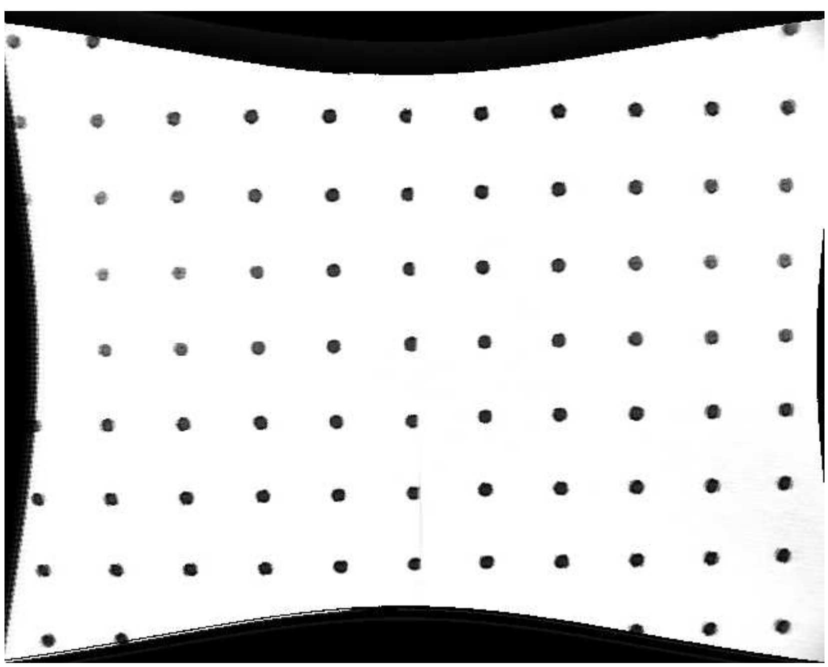

Fig. 2 Corrected distortion of 110 degree field of view lens