Realtime Camera Calibration for Enhanced Reality Visualization *

J.P. Mellor

Massachusetts Institute of Technology, Cambridge, MA 02139 USA
email: jpmellor@ai.mit.edu

Abstract. The problem which must be solved to make realtime enhanced reality visualization possible is basically the camera calibration problem. The relationship between the coordinate frames of the patient, the patient's internal anatomy scans and the image plane of the camera observing the patient must be established. This paper presents a new approach to finding this relationship and develops a system for performing enhanced reality visualization. Given the locations of a few fiducials our method is fully automatic, runs in nearly real-time, is accurate to a fraction of a pixel, allows both patient and camera motion, automatically corrects for changes to the internal camera parameters (focal length, focus, aperture, etc.) and requires only a single video image.

1 Introduction

Enhanced reality visualization is the process process of adding information to a real image. We take a model of the patient obtained from MR or CT data and overlay it on a video image of the patient. The enhanced image effectively provides X-ray vision by displaying those portions of the model which are not visible in the video image in the correct location.

We have developed a novel method for determining the relationship between model and image coordinates using two key insights which simplify the problem.

1. It is not necessary to separate the intrinsic and extrinsic camera parameters for enhanced reality visualization.
2. It is possible to recover depth information from a singe 2D image.

Many of the current enhanced reality visualization systems rely on active sensors such as magnetic trackers and require calibrated cameras. In contrast, our approach uses video information exclusively and does not require a calibrated camera making it particularly well suited for enhanced reality visualization. For a summary of current enhanced reality visualization techniques see [3].

* This work was supported in part by ARPA under Rome Laboratory contract F3060-94-C-0204 and ONR contract N00014-91-J-4038.

2 Our Method

2.1 A Perspective Transformation is Enough

The camera calibration problem is typically posed as follows:

$$I = M \begin{bmatrix} \mathcal{R} \\ \mathcal{T} \end{bmatrix} \mathcal{C} \tag{1}$$

Where I is a matrix of image points, M is a matrix of model points, $\mathcal{R}$ is an orthonormal rotation matrix, $\mathcal{T}$ is a translation vector and $\mathcal{C}$ is an internal camera calibration matrix. In general separating the parameters of $\mathcal{R}$, $\mathcal{T}$ and $\mathcal{C}$ is difficult. Unless the individual parameters are needed, there is no reason to separate them.

Camera calibration is often performed as a preliminary step in many applications. A set of camera calibration parameters is recovered and the intrinsic parameters are used for future images. This assumes that the intrinsic parameters are fixed. In general, they are not. They change with the focus and aperture settings. For example, the principle point can shift by 8 pixels or more with adjustments to focus [4]. The effective focal length f also varies with focus and aperture settings. Zoom lenses take this variability to an extreme enabling large changes to f. Lens distortion also varies with changes to focus and aperture [1].

In enhanced reality visualization, we are interested in the total transformation from model to image coordinates. We do not need to separate intrinsic and extrinsic parameters to generate an enhanced reality image. All of the parameters comprising both the intrinsic and extrinsic calibration parameters can be composed into a single 3×4 matrix. Equation (1) can be rewritten as:

$$I = M\mathcal{P} \tag{2}$$

Where M is a matrix of model points in the form $[X\ Y\ Z\ 1]$, I is a matrix of image points in the form $[x\ y\ z]$, x' and y' are the image coordinates of the projection of M onto the image plane and $x' = x/z$, $y' = y/z$. For our purposes finding values for the elements of $\mathcal{P}$ is sufficient. By formulating the problem in this manner we avoid the difficulties associated with decomposing the intrinsic and extrinsic camera parameters. We solve (2) for each image we obtain. Changes in the intrinsic parameters of the camera are inherently captured.[2]

2.2 Depth Information From a Single 2D Image

Even with the simplifications made so far, the problem is still nonlinear. Using spatial features rather than point features enables us to measure a local scale factor at each feature. We use circular fiducials as our spatial features. In essence, we are able to recover $2\frac{1}{2}$D information. The local scale factor s_i is equal to

[2] A linear approximation to radial distortion is implicitly modeled. Using 16mm and 25mm lenses of average quality we have found this approximation quite satisfactory. If lens distortion is more severe, a separate correction could be added

the focal length divided by the depth of the feature (f/z_i). The idea of using spatial features to recover a measure of depth is not new, however our use of this information is. Given the local scale factors, solving for $\mathcal{P}$ reduces to a set of three linear equations.

$$x'_i = s_i\,(p_{11}X_i + p_{21}Y_i + p_{31}Z_i + p_{41}) \tag{3}$$

$$y'_i = s_i\,(p_{12}X_i + p_{22}Y_i + p_{32}Z_i + p_{42}) \tag{4}$$

$$1 = s_i\,(p_{13}X_i + p_{23}Y_i + p_{33}Z_i + p_{43}) \tag{5}$$

Where x'_i, y'_i and $1/s_i$ are the location and local scale factor of the i^{th} image point and X_i, Y_i and Z_i are the components of the i^{th} model point. Given correspondences between four model and image points (3), (4) and (5) are exactly determined. If more than four correspondences exist the problem is solved in a least-squares fashion by minimizing the following error terms. Where $j = 1, 3$ and I_{ij} is the j^{th} component of the i^{th} image point $[x'_i\ y'_i\ 1]$

$$\|r_j\|_2 = \sum_{i=1}^{n} \left| \frac{I_{ij}}{s_i} - (p_{1j}X_i + p_{2j}Y_i + p_{3j}Z_i + p_{4j}) \right|^2 \tag{6}$$

2.3 Feature Detection

The location and local scale factor of our circular fiducials are recovered using moment calculations.[3] The projection of a circle is an ellipse. Under orthographic projection the centroid of the ellipse corresponds to the projection of the centroid of the circle and the major axis of the ellipse is related to the diameter of the circle by f/z. Under perspective projection these relationships are no longer exact. However for our configuration, a 1cm fiducial viewed from 1m using a 16mm or 25mm lens, the effects of perspective distortion are negligible compared to noise. A more detailed discussion can be found in [3].

2.4 Implementation

The current system is implemented in Lucid Common Lisp and runs on a Sun SparcStation 2 using a VideoPix frame grabber. The two most limiting components are the frame grabber (≤ 4 frames/sec) and the rendering/display system. Depending on the complexity of the model, the renderer may require a minute or more to perform the rendering. Using a simple model, frame rates of ~2hz can be achieved. If the image acquisition and rendering overheads were removed and the rest of the system was optimized, frames rates of ~10-20hz should be achievable without upgraded hardware. Using a C-30 digit signal processing board frame rates of >30hz should be achievable.

[3] The current implementation utilizes the aspect ratio of the image plane. Of all the camera parameters, the aspect ratio is truly fixed for a given camera/frame grabber combination. In theory, it is possible to perform a self-calibration to recover the aspect ratio when needed. We have not yet implemented this capability.

3 Results

Figure 1 shows some results typical of our method for a geometric object. The locations of the fiducials are accurately known. There are a total of seven fiducials. Three of the images have five in one plane and two in a plane displaced by 1cm. The fourth image has a single fiducial 4cm out of the plane. The image is obtained using a 16mm lens from ~1m. The pillar near the center of the object is 8cm high and is used exclusively for validation purposes. A wire frame corresponding to the edges of the pillar is overlayed on an image. The degree to which the wire frame matches the edges in the image is a measure of the accuracy of our method. The enhancement is accurate well outside the volume enclosed by the fiducials.

Figure 2 shows some results typical of our method for a plastic skull. The locations of the fiducials are determined using a laser scanner [2]. A discussion of this initial calibration can be found in [3]. Given the locations of the fiducials, a model of the skull obtained from CT imaging (shown as black dots) and a video image taken from ~ 1m away with a 25mm lens, the enhanced reality visualizations are generated. There are some slight inaccuracies near the edges of some of the images. The source of this error is currently under investigation. At least some of the error is introduced during the initial calibration.

4 Conclusion

A novel method for enhanced reality visualization has been presented. The method implicitly models the transformation from model to image coordinates and takes advantage of spatial features to obtain a linear system of equations. The method requires only a few fiducials and a single video image. The results are quick and accurate allowing camera parameters such as focus, aperture and zoom to vary dynamically without knowledge of the changes.

References

1. Duane C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 1965.
2. W.E.L. Grimson, T. Lozano-Pérez, G.J. Ettinger W.M. Wells III, S.J. White, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. In *Computer Vision and Pattern Recognition*, pages 430–436. IEEE, June 1994. Seattle, WA.
3. J.P. Mellor. Enhanced reality visualization in a surgical environment. Master's thesis, Massachusetts Institute of Technology, January 1995.
4. Reg G. Willson and Steven A. Shafer. What is the center of the image? Technical Report CMU-CS-93-122, Carnegie-Mellon University, Computer Science Department, April 1993.

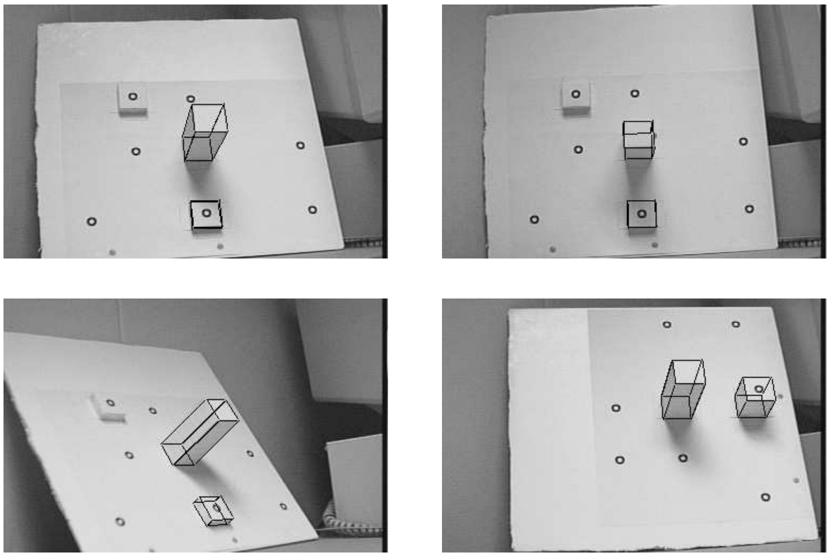

Fig. 1. Some Enhanced Reality Visualizations Using a Geometric Object.

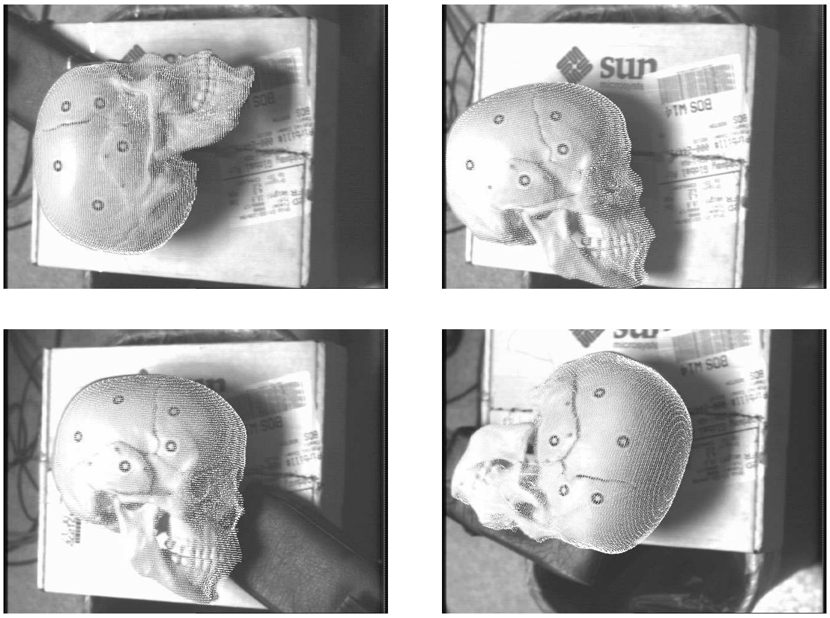

Fig. 2. Some Enhanced Reality Visualizations Using a Plastic Skull.