Appeared at ICCV 93, Berlin

Dynamic Calibration of an Active Stereo Head

James L. Crowley and Philippe Bobet
LIFIA(IMAG), Grenoble, France

Cordelia Schmid,
University of Karlsruhe, Germany

Abstract

This paper presents a method for using objects in a scene to define the reference frame for 3-D reconstruction. We first present a simple technique to calibrate an orthographic projection from four non-coplanar reference points. We then show how the observation of two additional known scene points can provide the complete perspective projection. When used with a known object, this technique permits a calibration of the full projective transformation matrix. For an arbitrary non-coplanar set of four points, this calibration provides an affine basis for the reconstruction of local scene structure. When the four points define three orthogonal vectors, the basis is orthogonal, with a metric defined by the lengths of the three vectors.

We demonstrate this technique for the case of a cube. We present results in which five and a half points on the cube are sufficient to compute the projective transformation for an orthogonal basis by direct observation (without matrix inversion). We then present experiments with a technique for reducing the imprecision due to pixel quantization and noise.

1 Introduction

Efforts to implement 3D vision systems have led numerous groups to confront the problem of calibrating cameras. The most widely used camera model is the "thin-lens" or pin-hole model, modelled by a perspective transformation represented in homogeneous coordinates. Reconstruction techniques tend to be extremely sensitive to the coefficients of this transformation. Of particular difficulty are are techniques which estimate distance to scene points and then attempt to reconstruct 3D shape using the so-called "intrinsic parameters" of the camera.

The intrinsic parameters are the parameters that are independent of camera position and orientation. They are typically listed as the "center" of the image, defined by the intersection of the optical axis with the retina, and the ratio of pixel size to focal length in the horizontal and vertical directions [1]. Reconstruction using depth is extremely sensitive to the precision of these parameters. This has led a number of investigators to develop techniques using estimation theory based on a large number of observations of a calibration pattern [2] [3]. Such techniques typically require careful set up and rather long computation times for precise location of the reference points.

It is often overlooked that the pin-hole model is only a rough approximation for the optics of a camera. For a real camera, there are typically a continuum of values for the intrinsic parameters providing reasonable approximations to the physical system. This continuum is extremely sensitive to the setting for focus and aperture and even to small perturbations in retina position due to vibration! Reconstruction techniques based explicit intrinsic camera parameters are extremely sensitive to the accuracy of these parameters. It is not surprising that most current 3-D vision systems only work for carefully set up laboratory demonstrations.

The techniques presented in this paper are the result of problems that we have encountered in the construction of a real-time active vision system [4]. Our system employs a binocular camera head mounted on a robot arm which serves as a neck. The system uses dynamically controlled vergence to fixate on objects. It is designed to track and servo on 2-D forms, to interpret such forms as objects, and to maintain a dynamically changing model of the 3D form of a scene. Focus and convergence of stereo cameras are maintained by low level reflexes. Constantly changing these parameters has posed difficult problems for 3D techniques based on classical calibration of the intrinsic camera parameters. Cumbersome and time consuming set-up means that calibration can not be performed “on the fly” as the system operates.

We have found that a robust 3D vision system may be constructed using the objects in a scene to calibrate the cameras. The simplest form of such calibration provides an orthographic transformation to an affine, scene based, reference defined by four non-coplanar points. A full perspective projection may be obtained from 6 known non-coplanar points.

2 Calibrating to an affine reference frame

In this section we show how an orthographic projection matrix can be computed by observation of four non-coplanar points. We then show how this transformation can be completed to form the perspective transformation by the observation of two additional points whose position is known relative to the first four points.

2.1 The Transformation from scene to image

In homogeneous coordinates, a point in the scene is expressed as a vector:

This work has been sponsored by the CEC DG XII under Project ESPRIT BR 3038 / 7108 “Vision as Process”

$${}^{s}P = [x_s, y_s, z_s, 1]^T$$

The index "s" raised in front of the letter indicates a "scene" based coordinate system for this point. The origin and scale for such coordinates are arbitrary. A point in an image is expressed as a vector:

$${}^{i}P = [i, j, 1]^T$$

The projection of a point in the scene to a point in the image can be approximated by a three by four homogeneous transformation ${}^{i}_{s}\mathbf{M}$. This transformation models the perspective projection with the equation:

$$\begin{bmatrix} w\ i \\ w\ j \\ w \end{bmatrix} = {}^{i}_{s}\mathbf{M} \begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix} \qquad (1)$$

The variable w captures the amount of "fore-shortening" which occurs for the projection of point ${}^{S}P$. This notation permits the pixel coordinates of ${}^{i}P$ to be recovered as a ratio of polynomials of ${}^{S}P$. That is

$$i = \frac{{}^{i}_{s}M_1 \cdot {}^{S}P}{{}^{i}_{s}M_3 \cdot {}^{S}P} \qquad j = \frac{{}^{i}_{s}M_2 \cdot {}^{S}P}{{}^{i}_{s}M_3 \cdot {}^{S}P} \qquad (2)$$

where ${}^{i}_{s}M_1$, ${}^{i}_{s}M_2$, and ${}^{i}_{s}M_3$ are the first, second and third rows of the matrix ${}^{i}_{s}\mathbf{M}$, and " $\cdot$ " is a scalar product.

2.2. Computing 3-D structure from stereo

Let ${}^{L}_{S}\mathbf{M}$ and ${}^{R}_{S}\mathbf{M}$ represent the transformations for the left and right cameras in a stereo pair. Let ${}^{L}_{S}\mathbf{M}_1$, ${}^{L}_{S}\mathbf{M}_2$, and ${}^{L}_{S}\mathbf{M}_3$ represent the first, second third rows of the ${}^{L}_{S}\mathbf{M}$, and ${}^{R}_{S}\mathbf{M}_1$, ${}^{R}_{S}\mathbf{M}_2$ and ${}^{R}_{S}\mathbf{M}_3$ represent the first, second third rows of the ${}^{R}_{S}\mathbf{M}$. Observation of a scene point, ${}^{S}P$, gives the image points ${}^{L}P = (i_L, j_L)$ and ${}^{R}P = (i_R, j_R)$. From equation 2 we can write [2]:

$$i_L = \frac{{}^{L}_{s}M_1 \cdot {}^{S}P}{{}^{L}_{s}M_3 \cdot {}^{S}P} \qquad i_R = \frac{{}^{R}_{s}M_1 \cdot {}^{S}P}{{}^{R}_{s}M_3 \cdot {}^{S}P}$$

$$j_L = \frac{{}^{L}_{s}M_2 \cdot {}^{S}P}{{}^{L}_{s}M_3 \cdot {}^{S}P} \qquad j_R = \frac{{}^{R}_{s}M_2 \cdot {}^{S}P}{{}^{R}_{s}M_3 \cdot {}^{S}P} \qquad (3)$$

This provides us with a set of four equations for recovering the three unknowns of ${}^{S}P$. Each equation describes a plane in scene coordinates that passes through a column or row of the image. Unfortunately, because of errors in pixel position due to sampling and image noise, the projection of these planes do not necessarily meet at a point. Thus we compute the point as the mean-square solution to the the four equations.

Because of quantization and the lever-arm effect, stereo reconstruction produces errors which are proportional to the distance from the origin. By placing the origin on the object to be observed, such error may be minimized. Computing the matrix ${}^{i}_{s}\mathbf{M}$ for a pair of cameras permits a very simply method to compute the position of points in the scene in a reference frame defined by the scene. Dynamically developing the transformations for the left and right images permit objects in the scene to be reconstructed independent of errors in the relative or absolute positions of the cameras.

2.3 Calibrating an orthographic projection

Any four points in the scene which are not in the same plane can be used to define an affine basis. Such a basis can be used as a scene based coordinate system (or reference frame). One of the four points in this reference frame will be taken as the origin. The other three points defines three axes, as shown in figure 1. On an arbitrary object, these axes are not necessarily orthogonal.

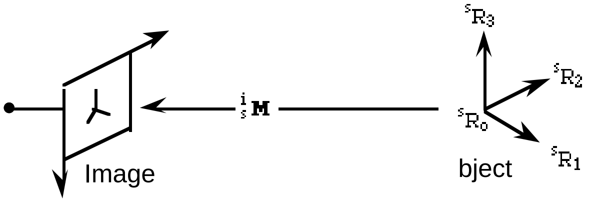

Figure 1 Four non-coplanar points define an affine reference frame.

A simple way to exploit this idea is to use any four non-coplanar points to define an orthographic projection from an affine reference frame in the scene to the image. Let us designate a point in the scene as the origin for a reference frame. By definition,

$${}^{S}R_o = [0, 0, 0, 1]^T$$

Three axes for an affine object-based reference frame may be defined by designating three additional scene points as:

$${}^{S}R_1 = [1, 0, 0, 1]^T \qquad {}^{S}R_2 = [0, 1, 0, 1]^T$$

$${}^{S}R_3 = [0, 0, 1, 1]^T$$

The vector from the origin to each of these points defines an axis for measuring distance. The length of each vector defines the unit distance along that vector. These three vectors are not required to be orthogonal. The four points may be used to define an affine basis by the addition of a constraint that the sum of the coefficients be constant [5]. We note that when the points are the corners on a right parallelpiped (a box), then they can be used to define an orthogonal basis and the additional constraint is unnecessary.

Let the symbol $\cup$ represent the composition of vectors as columns in a matrix. We can then represent our affine coordinate system by the matrix ${}^{S}R$.

$${}^{S}R = [{}^{S}R_1 \cup {}^{S}R_2 \cup {}^{S}R_3 \cup {}^{S}R_o] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

The projection on these four points to the image can be written as four image points ${}^{i}P_0$, ${}^{i}P_1$, ${}^{i}P_2$, and ${}^{i}P_3$. These image points form an observation of the reference system, represented by the matrix ${}^{i}P_w$, where the term w_0 has been set to 1.0.

$${}^{i}P_w = \begin{pmatrix} w_1\, i_1 & w_2\, i_2 & w_3\, i_3 & i_0 \\ w_1\, j_1 & w_2\, j_2 & w_3\, j_3 & j_0 \\ w_1 & w_2 & w_3 & 1 \end{pmatrix}$$

This allows us to write a matrix expression.

$${}^{i}P_w = {}^{i}_{s}\mathbf{M}\ {}^{s}R$$

The reference matrix ${}^{s}R$ has a simple inverse, which can be solved by hand.

$${}^{s}R^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 1 \end{pmatrix}$$

Inverting this matrix allows us to write the expression:

$${}^{i}_{s}\mathbf{M} = {}^{i}P_w\ {}^{s}R^{-1}$$

or

$${}^{i}_{s}\mathbf{M} = \begin{pmatrix} w_1 i_1 - i_0 & w_2 i_2 - i_0 & w_3 i_3 - i_0 & i_0 \\ w_1 j_1 - j_0 & w_2 j_2 - j_0 & w_3 j_3 - j_0 & j_0 \\ w_1 - 1 & w_2 - 1 & w_3 - 1 & 1 \end{pmatrix} \tag{4}$$

Having performed the inversion of ${}^{s}R$ by hand, there is no need to compute an inverse when the system is calibrated. The problem with equations 4 are the fore-shortening coefficients w_1, w_2, w_3. It is useful to consider the meaning of this vector. Each term “w” is a scale factor that describes the amount of “foreshortening” induced by perspective along each of the reference vectors. The units of this fore-shortening are (1/meters). Thus, if the scale factor is defined to be 1.0 at the reference point R_0, then vectors emanating from reference point R_1 will be “scaled” by a factor of w_1.

A simple solution is to set the coefficients w_1, w_2, w_3 to 1, yielding an orthographic projection. The magnitude of the error for such an approximation is proportional to the distance from the chosen origin, and inversely proportional to the focal length of the camera.

$${}^{i}_{s}\mathbf{M} \approx \begin{pmatrix} i_1 - i_0 & i_2 - i_0 & i_3 - i_0 & i_0 \\ j_1 - j_0 & j_2 - j_0 & j_3 - j_0 & j_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The orthographic approximation can provide a usable approximation for points near the reference object when the depth is large relative to the focal length. Alternatively, we may seek to determine the full perspective transformation by solving a set of linear equations to determine w_1, w_2, w_3. Solving for these coefficients requires three additional constraints or the observation of one and a half additional points whose position is known with respect to the first four points.

2.4 Obtaining the perspective projection

To obtain the perspective transformation to an affine basis from equation 4 we must solve for w_1, w_2, w_3. Solving for these three variables requires 3 independent equations, or the observation of the image coordinates for one and a half scene points. Let us define two scene points, ${}^{s}R_4$ and ${}^{s}R_5$, whose positions are known with respect to our affine basis.

$${}^{s}R_4 = [x_4, y_4, z_4, 1]^T$$

$${}^{s}R_5 = [x_5, y_5, z_5, 1]^T$$

Equation 4 permits us to use these points to write four equations with three unknowns.

$$i_4 = \frac{{}^{i}_{s}M_1 \cdot {}^{s}R_4}{{}^{i}_{s}M_3 \cdot {}^{s}R_4} \qquad j_4 = \frac{{}^{i}_{s}M_2 \cdot {}^{s}R_4}{{}^{i}_{s}M_3 \cdot {}^{s}R_4} \tag{5}$$

$$i_5 = \frac{{}^{i}_{s}M_1 \cdot {}^{s}R_5}{{}^{i}_{s}M_3 \cdot {}^{s}R_5} \qquad j_5 = \frac{{}^{i}_{s}M_2 \cdot {}^{s}R_5}{{}^{i}_{s}M_3 \cdot {}^{s}R_5} \tag{6}$$

Provided that no five of our six points are coplanar, these four equations can be solved to obtain the values of w_1, w_2 and w_3.

When the positions of the points ${}^{s}R_4$ and ${}^{s}R_5$ are known in advance, the solution can be structured to yield the full perspective transformation by direct observation, without matrix inversion. To illustrate this, let us consider the problem of calibrating ${}^{i}_{s}\mathbf{M}$ by observation of 6 vertices of cube.

3 Calibration by Direct Observation

A direct solution for calibrating the projective form of the matrix ${}^{i}_{s}\mathbf{M}$ is possible when the reference points are known in advance. This solution can be had without matrix inversion. Let us illustrate the technique by deriving the equations for calibrating the matrix ${}^{i}_{s}\mathbf{M}$ from the observation of 6 points on a parallelpiped.

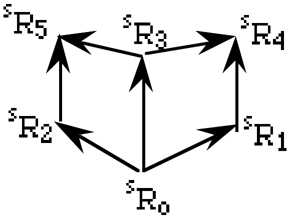

Figure 2 The reference points for a parallelpiped

3.1 Derivation

Consider a reference frame defined by six points on a cube, as shown in figure 2. Point ${}^{s}R_0$ defines the origin. Points ${}^{s}R_1$, ${}^{s}R_2$ and ${}^{s}R_3$ define the unit vectors for the X, Y and Z axes. Points ${}^{s}R_4$ and ${}^{s}R_5$ permit the full projective transformation to be recovered. Points ${}^{s}R_0$, ${}^{s}R_1$, ${}^{s}R_2$ and ${}^{s}R_3$ are defined as above as. Points ${}^{s}R_4$ and ${}^{s}R_5$ are given by:

$${}^{s}R_4 = [1, 0, 1, 1]^T$$

$${}^{s}R_5 = [0, 1, 1, 1]^T$$

Substituting ${}^{s}R_4$ and ${}^{s}R_5$ into equation 6 gives:

$$(i_4 – i_1)\, w_1 + (i_4 – i_3)\, w_3 – (i_4 – i_o) \;= 0 \tag{7}$$
$$(j_4 – j_1)\, w_1 + (j_4 – j_3)\, w_3 – (j_4 – j_o) \;= 0 \tag{8}$$
$$(i_5 – i_2)\, w_2 + (i_5 – i_3)\, w_3 – (i_5 – i_o) \;= 0 \tag{9}$$
$$(j_5 – j_2)\, w_2 + (j_5 – j_3)\, w_3 – (j_5 – j_o) \;= 0 \tag{10}$$

The coefficients w_1 and w_3 can be had from equations 7 and 8, that is from observation of ${}^{s}R_4$. The coefficients w_2 and w_3 can be had from equations 9 and 10, that is from observation of ${}^{s}R_5$. From the point ${}^{s}R_4$ we obtain:

$$w_1 = \frac{(i_4 – i_0)\,(j_4 – j_3) – (i_4 – i_3)\,(j_4 – j_0)}{(i_4 – i_1)\,(j_4 – j_3) – (i_4 – i_3)\,(j_4 – j_1)} \tag{11}$$

$$w_3 = \frac{(i_4 – i_0)\,(j_4 – j_1) – (i_4 – i_1)\,(j_4 – j_0)}{(i_4 – i_3)\,(j_4 – j_1) – (i_4 – i_1)\,(j_4 – j_3)} \tag{12}$$

While, from the point ${}^{S}R_5$ we obtain:

$$w_2 = \frac{(i_5 – i_0)\,(j_5 – j_3) – (i_5 – i_3)\,(j_5 – j_0)}{(i_5 – i_2)\,(j_5 – j_3) – (i_5 – i_3)\,(j_5 – j_2)} \tag{13}$$

$$w_3 = \frac{(i_5 – i_0)\,(j_5 – j_2) – (i_5 – i_2)\,(j_5 – j_0)}{(i_5 – i_3)\,(j_5 – j_2) – (i_5 – i_2)\,(j_5 – j_3)} \tag{14}$$

The fact that the equations are over-constrained poses a small problem. If the image position of points ${}^{s}R_4$ and ${}^{s}R_5$ are not perfectly measured, the resulting solution for w_1, w_2 and w_3 will not be consistent. However, it is inevitable that the position of the reference points will be corrupted by small random variations in position, if for no other reason, because of image sampling. We can improve the precision by exploiting the redundancy of the last half of a point to correct for random errors in the image position of the reference points.

3.2 Correcting for pixel quantization

The classic method for minimizing the inconsistency in reference point position is to compute a mean-squared solution. Faugeras and Toscani [2] present a direct method to minimize the sum of the error between the projection of calibration points and and their observation. From equation 3, for each calibration point ${}^{s}R_k$ and its image projection ${}^{i}P_k$, we can write:

$$({}^{i}_{s}M_1 \bullet {}^{s}R_k) – i_k\, ({}^{i}_{s}M_3 \bullet {}^{s}R_k) \;= 0$$
$$({}^{i}_{s}M_2 \bullet {}^{s}R_k) – j_k\, ({}^{i}_{s}M_3 \bullet {}^{s}R_k) = 0$$

For N non-coplanar calibration points we can write a linear system of 2N equations. We can then use Lagrange multipliers to obtain a least squares value for ${}^{i}_{s}\mathbf{M}$. We will refer to this below as the "mean square technique", denoted "msq" in the tables of experimental results below. In the following two sections we will compare the precision obtained from direct solution using 5 and 1/2 points to precision obtained using the least squares technique.

3.3 An Example of a calculation

In this section we present an example of the calibration using a aluminium cube with a side of 20cm. This example illustrates the method used in the experiments in the following sections. In our experiments, images of the cube are projected on the work-station screen and the pixel coordinates of the vertices ${}^{i}P_0$, ${}^{i}P_1$, ${}^{i}P_2$, ${}^{i}P_3$, ${}^{i}P_4$, and ${}^{i}P_5$ were selected with the mouse. The image size is 512 by 512 pixels. A standard left handed image coordinate system is used in which the origin is the upper left hand corner, positive i (columns) is to the left, and positive j (rows) is down. Reference points were indicated by pointing with a mouse, a technique which can sometimes result in an error of one or two pixels. For the left image, the vertices of the cube were detected at:

${}^{L}P_o = (228, 481)$	${}^{L}P_3 = (229, 223)$
${}^{L}P_1 = (347, 351)$	${}^{L}P_4 = (354, 107)$
${}^{L}P_2 = (77, 374)$	${}^{L}P_5 = (69, 125)$

Equations 7 through 9 give a solution for $\vec{w}$ of:

$$\vec{w} = (0.917610,\;\; 0.858158,\;\; 1.052614,\;\; 1)$$

By the direct method we then obtain ${}^{L}_{S}\mathbf{M}$ as

$$\begin{pmatrix} 147.589396 & –146.422112 & –11.048572 & 228.000000 \\ –101.081043 & –84.764543 & –269.732889 & 481.000000 \\ 0.082390 & 0.059453 & –0.052614 & 1.000000 \end{pmatrix}$$

Using the least squares technique, the matrix for the left image ${}^{L}_{S}\mathbf{M}$ is computed as:

$$\begin{pmatrix} 148.016122 & –146.716244 & –12.239302 & 228.149911 \\ -100.417731 & –85.159763 & –270.607106 & 481.003325 \\ 0.084301 & 0.058403 & –0.056504 & 1.000000 \end{pmatrix}$$

For the right image, the vertices of the cube were detected at:

${}^{R}P_o = (212, 464)$	${}^{R}P_3 = (197, 208)$
${}^{R}P_1 = (343, 332)$	${}^{R}P_4 = (337, 88)$
${}^{R}P_2 = (74, 360)$	${}^{R}P_5 = (52, 116)$

With the direct method, from equation 11 through 13 we obtain

$$\begin{pmatrix} 158.141055 & –132.711116 & –26.996216 & 212.000000 \\ –105.729358 & –78.270296 & –268.666055 & 464.000000 \\ 0.079128 & 0.071471 & –0.060894 & 1.000000 \end{pmatrix}$$

Using the least squares technique, the matrix for the right image ${}^{R}_{S}\mathbf{M}$ is computed as:

$$\begin{pmatrix} 158.066763 & –132.620333 & –26.745194 & 211.958839 \\ –105.863649 & –78.136621 & –268.493161 & 464.002612 \\ 0.078734 & 0.071856 & –0.060038 & 1.000000 \end{pmatrix}$$

As a check, we indicated the image positions of the point ${}^{s}R_6 = [1, 1, 1, 1]^T$ and construct the 3D position of this point by a stereo solution. Clicking on the corner corresponding to point 6 in the left and right images gives:

${}^{L}P_6 = (200, 23)$	${}^{R}P_6 = (193, 11)$

Solving for the 3-D position with the stereo technique using all four equations as described above gives

method	X	Y	Z	Dist
direct	1.004628	1.005564	0.997816	0.007559
msq	0.992905	0.993915	1.004042	0.010183

Reconstructed points are expressed in units defined by the side of the cube. One multiplies by 20 to obtain centimeters. We can observe that for this example, the direct calculation gives an error of about 0.7%, while the mean square technique gives about 1% error. The error is due to both the sampling interval of the pixels. Although the direct solution happened to perform best in this example, we will see in the experiments presented in the next section that the error is a random function. The mean square technique tends to give an error with the lowest average value.

3.4 Experiment with sensitivity to pixel noise

As a test of sensitivity to quantization, we modelled a stereo pair of cameras and then computed a stereo solution using the corrupted projection matrices to recover the 3-D position of a known scene point. We simulated our nominal experimental set-up composed a pair of cameras with a base line of 20cm, a focal length of 25mm and images with 512 x 512 pixels. The cameras are simulated to be looking at a cube 20 cm on each side at a distance of 1.2 meters. Three dimensional points were computed using least squares solution from all four stereo equations. Table 3.1 shows the reconstruction of scene point (1, 1, 1), as a function of random noise added to the calibration points. We can notice that the mean-square technique is more than twice as precise as a direct calculation.

std	0.0125	0.250	0.500	1.0	2.0	4.0
direct	0.0193	0.0388	0.0786	0.1613	0.3397	0.8395
msq	0.0092	0.0183	0.0364	0.0718	0.1395	0.3031

Table 3.1 3-D error for scene point (1,1,1) as a function of standard deviation of pixel error of calibration points.

Another question which one might ask is, what is the influence of the size of the cube in the image on the error of reconstruction. Equation 4 shows that the coefficients are calculated from the lengths of the vectors in the image. Thus, the larger the distance between the image of the calibration points, the less sensitive the coefficients are to an error in image position. None-the-less, one should ask: how sensitive is the 3-D reconstruction to the length of this vector?

Using a simulated cube, and the mean square correction method, we computed calibration matrices for a 20cm cube at distances of 100 cm to 200 cm in steps of 20 cm. At 100 cm, the cube fills the image. At 200 cm the cube is the size of a quarter of the image. For each pair of calibration matrices, we computed the stereo solutions for image projects at scene points (1,1,1). We used calibration matrices computed from pixel positions corrupted by Gaussian noise of standard deviation 0.125, 0.25, 0.5, 1, 2, and 4. For each point we performed a stereo reconstruction 100 times and computed the average and maximum errors, as shown in tables 3.2 and 3.3. The stereo solutions are computed, as above, using all four equations.

At a distance of 100 cm, the sides of the cube project to vectors of nearly the entire image. Interesting, in table 3.2, we see that in this case, the percentage of error in reconstruction is almost exactly proportional to the standard deviation of the pixel noise. That is, for a pixel error of 0.5 pixels the reconstruction error is 0.53%, for a pixel error of 1.0 the reconstruction error is 1.07%. The error percentages doubles when the cube occupies half the image at 140 cm, and double again when the cube reaches a quarter of the image at 200 cm.

dist	0.125	0.25	0.50	1.00	2.00	4.00
100	0.0013	0.0026	0.0053	0.0107	0.0214	0.0428
120	0.0019	0.0038	0.0076	0.0152	0.0303	0.0609
140	0.0025	0.0051	0.0102	0.0204	0.0410	0.0826
160	0.0033	0.0066	0.0132	0.0265	0.0532	0.1082
180	0.0041	0.0083	0.0167	0.0335	0.0675	0.1440
200	0.0051	0.0103	0.0206	0.0412	0.0832	0.1745

Table 3.2 The average 3-D error as a function of distance of the calibration cube from the camera (rows) and as a function of pixel noise (columns). Errors are expressed in units of the length of the side of the calibration cube (20cm). Projection was computed using the mean square technique. Scene points were computed using all four stereo equations.

dist	0.125	0.25	0.50	1.00	2.00	4.00
100	0.0065	0.0130	0.0260	0.0515	0.1008	0.2042
120	0.0092	0.0183	0.0364	0.0718	0.1395	0.3031
140	0.0123	0.0245	0.0485	0.0953	0.1893	0.4345
160	0.0158	0.0315	0.0623	0.1216	0.2529	0.6112
180	0.0198	0.0394	0.0776	0.1505	0.3306	3.5540
200	0.0243	0.0481	0.0945	0.1860	0.4257	1.8469

Table 3.3 The maximum reconstruction error, due to pixel noise, for the corners of the cube.

3.5 Experimental precision with real images

The following experiment was performed with live images produced by a Pulnix TM 560 camera equipped with a Cosmicar 25 mm F1.8 lens fed CCIR video signals to an Imaging Technologies FG100 Digitizer. Images were acquired with a resolution of 512 by 512.

Point	Real Position	direct	msq
S_0	(0, –0.325, 0)	**0.0175**	0.0185
S_1	(4.75, –0.325, 0)	0.1065	**0.0948**
S_2	(0, 0, 0)	**0.0131**	0.0523
S_3	(0, –0.325, 0.95)	**0.0240**	0.0549
S_4	(0.475, –0.325,.95)	0.1082	**0.0372**
S_5	(0, 0, 0.95)	0.0750	**0.0549**

Table 3.4 Errors for reconstructed corners of sugar box using three techniques. All distances are in units defined by the side of the calibration cube (20cm). The most precise values are indicated in bold.

Our 20 cm aluminium cube was painted such that two of its faces are white, two are gray and two are black. The cube was placed on a white table-cloth with a black face to the left, gray face to the right and the white face up. Stereo images of this cube from a distance of 1.2 meters were obtained. We then placed a box of sugar next to the calibration cube and reconstructed the corners of the box using the matrices determined by the two techniques. The six visible corners of the sugar box are listed as points S_0 through S_5. The 3-D error, measured as a percentage of the side of the cube, are shown for each of the 6 points.

The first thing that we can observe is that neither technique produces the best result for all six corners. The largest error was on the order of 10% for the the direct method, while the smallest was on the order of 1% for the direct method. None-the-less, our conclusion from these and many other experiments is that computing the calibration matrix using the mean-square technique gives a slight improvement in precision at a slight increase in computational cost. The direct method provides a 3D solution which is less precise but easier to program.

4 Discussion and Conclusions

The reliable operation of a 3D vision system depends on accurate calibration. Calibration procedures which require time consuming and cumbersome set-up are of little use when the optical parameters of the lenses are continually changing. In this paper we have presented the foundations for a technique in which camera calibration is determined and maintained using objects in the scene. These techniques permit objects in the scene to serve as the reference frame in which the scene is reconstructed. Because the object is reconstructed in its own reference frame, information about the shape of an object can be registered and fused without knowledge of the camera positions relative to the object.

This paper is concerned with the mathematics of calibration and reconstruction. We have not addressed the problem of locating the reference points. Yet the critical dependence of 3-D precision on image location shows that this is a fundamental problem for which a satisfactory solution still does not exist. What we can offer to this problem is the criteria for evaluating image analysis techniques for 3D reconstruction.

Our principal conclusion involves calibration. Some researchers argue for an initial calibration phase using a complex set up involving many reference points. The argument is that additional reference points permit improvement in precision through use of statistical methods. In a continuously operating vision system, calibration matrices must be continuously corrected for effects due to focus, aperture, vergence and camera zoom, as well as vibrations that can change the lens mounting. Thus, a more precise reconstruction of the scene requires continually updating the calibration.

Acknowledgements

The ideas developed in this paper first evolved in discussion with a number of people. The philosophical contribution of Roger Mohr is sufficiently important that we have even offered to name him as a co-author. Listening to presentations by Jan Koenderinck and by Gunnar Sparr has also provided key inspirations. Discussion with Thierry Vieville have also helped us to organise and better understand certain aspects of the technique.

Bibliography

[1] R. Y. Tsai, "A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using off the Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol 3 No. 4, August 1987.

[2] O. D.Faugeras and G. Toscani, "The Calibration Problem for Stereo. Computer Vision and Pattern Recognition, pp 15-20, Miami Beach, Florida, June 1986.

[3] P. Puget and T. Skordas, "Calibrating a Mobile Camera", Image and Vision Computing, Vol 8 No. 4, November 1990.

[4]] J. L. Crowley, "Towards Continuously Operating Integrated Vision Systems for Robotics Applications", SCIA-91, Seventh Scandinavian Conference on Image Analysis, Aalborg, DK, August 91.

[5] G. Sparr, "Depth Computations from Polyhedral Images", The Second European Conference on Computer Vision (ECCV-2), St. Margherita, Italy, May 1992.

[6] J. Koenderink and A. J. Van Doorn, "Affine Structure from Motion", Technical Report, Universtiy of Utrecht, Oct. 1989.

[7] R. Mohr, L. Morin and E. Grosso, "Relative Positioning with Poorly Calibrated Cameras", LIFIA-IMAG Technical Report RT 64, April 1991.

[8] J. L. Crowley, P. Bobet and C. Schmid , "Auto-Calibration of Cameras by Direct Observation of Objects", Journal of Image and Vision Computing, March 1993.