Maintaining Stereo Calibration by Tracking Image Points

James L. Crowley, Philippe Bobet, Cordelia Schmid

Appeared in 1993 CVPR, New York, June 1993.

Maintaining Stereo Calibration by Tracking Image Points

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

Cordelia Schmid,
University of Karlsruhe, Germany

Abstract

An important problem in active 3-D vision is updating the camera calibration matrix as the focus, aperture, zoom or vergence angle of the cameras changes dynamically. After an introduction we present a technique to compute the projection matrix from five and a half points in a scene without matrix inversion. We then present a technique for correcting the projective transformation matrix by tracking reference points. Our experiments show that a change of focus can be corrected by an affine transform obtained by tracking three points. For a change in camera vergence a projective correction, based on tracking four image points is slightly more precise than an affine correction matrix.

We also show how stereo reconstruction makes it possible to "hop" a reference frame from one object to another. Any set of four non-coplanar points in the scene may define such a reference frame. We show how to keep the reference frame locked onto a set of four points as a stereo head is translated or rotated. These techniques make it possible to reconstruct the shape of an object in its intrinsic coordinates without having to match new observations to a partially reconstructed description.

1 Introduction

This paper presents a method for updating the calibration matrices for an active stereo head. In an active 3-D vision system, the optical parameters of focus, aperture, zoom and vergence angle are constantly changing. The perspective transformation for a camera must be modified for each such change.

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 camera parameters makes it impossible to use classic 3D vision techniques based on pre-calibration of the intrinsic camera parameters.

Continually changing the focus of a camera made the use of reconstruction based on depth to scene points completely impractical. The "intrinsic parameters" are simply not constant [3]. In one experiment in our laboratory with a 25mm F1.8 lens, turning the focus over its full range changed size of an image by 5% and moved the center point in roughly a circular path whose largest displacement was approximately 20 pixels. We have also demonstrated image shifts of up to 4 pixels in the detection of edge-lines resulting uniquely from a change in aperture. This variability has posed a serious problem for the realization of a continuously operating 3D vision system. Our system required a method by which the calibration could be obtained and maintained by direct observation of objects in the scene.

We have found that a robust 3D vision system may be constructed using the objects in a scene to calibrate the cameras. With this technique, the cameras are calibrated by fixating on any known set of 6 points [8]. Calibration is then updated continually by tracking the image position of points as optical parameters are adjusted or as the camera is moved.

2 Calibrating to an affine reference frame

For 3D vision, the stereo cameras may be modeled by a 3 by 4 projective transformation matrix. An explicit separation of the camera parameters is not necessary.

2.1 The Transformation from scene to image

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

$$^{s}P = [x_S, y_S, z_S, 1]^T$$

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:

$$\begin{bmatrix} w\ i \\ w\ j \\ w \end{bmatrix} = {}^{i}_{s}\mathbf{M} \begin{bmatrix} x_S \\ y_S \\ z_S \\ 1 \end{bmatrix} \quad (1)$$

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

This work has been sponsored by the CEC DG XII Project ESPRIT BR 3038 / EP 7108 "Vision as Process"

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

We can recover 3-D position of points from stereo without explicitly extracting the intrinsic parameters [2]. Let ${}^{L}_{s}\mathbf{M}$ and ${}^{R}_{s}\mathbf{M}$ represent the matrices for the left and right cameras in a stereo pair. 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 two equations for each image. With a minimum of algebra, these equations can be rewritten as

$$\begin{aligned}
({}^{L}_{s}M_1 \bullet {}^{s}P) - i_L ({}^{L}_{s}M_3 \bullet {}^{s}P) = 0 \\
({}^{R}_{s}M_1 \bullet {}^{s}P) - i_R ({}^{R}_{s}M_3 \bullet {}^{s}P) = 0 \\
({}^{L}_{s}M_2 \bullet {}^{s}P) - j_L ({}^{L}_{s}M_3 \bullet {}^{s}P) = 0 \\
({}^{R}_{s}M_2 \bullet {}^{s}P) - j_R ({}^{R}_{s}M_3 \bullet {}^{s}P) = 0 \qquad (3)
\end{aligned}$$

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.

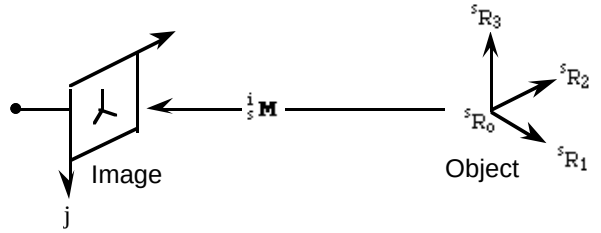

Figure 1 Four non-coplanar points define an affine reference frame.

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. 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 ${}^{i}P_o$, ${}^{i}P_1$, ${}^{i}P_2$, 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_o 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

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

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 equation 4 is the coefficients w_1, w_2 and w_3. It is useful to consider the physical interpretation of these coefficients. 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 and 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.

$$ {}^{i}_{s}\mathbf{M} \approx \begin{pmatrix} i1-i0 & i2-i0 & i3-i0 & i0 \\ j1-j0 & j2-j0 & j3-j0 & j0 \\ 0 & 0 & 0 & 1 \end{pmatrix} $$

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

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.

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

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

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.

The matrix ${}^{i}_{s}\mathbf{M}$ may be seen as composed of three parts:

$$ {}^{i}_{s}\mathbf{M} = {}^{i}_{r}\mathbf{C} \; {}^{r}_{c}\mathbf{P} \; {}^{c}_{s}\mathbf{T} $$

where

${}^{c}_{s}\mathbf{T}$ is a 4 by 4 matrix that describes the transformation from an arbitrary scene reference to a camera based projective reference,

${}^{r}_{c}\mathbf{P}$ is a 3 by 4 matrix which models the projection from the scene (in camera centered coordinates) to an ideal retina, and

${}^{i}_{r}\mathbf{C}$ is a 3 by 3 matrix which describes the projection from an ideal retina to the image.

In section 3, we show that a change to the optics or to the camera orientation can be modeled as a change in ${}^{i}_{r}\mathbf{C}$ and thus corrected by pre-multiplying ${}^{i}_{s}\mathbf{M}$ by a 3 by 3 correction matrix. This correction matrix may be exactly obtained by tracking four points. In some cases an affine approximation may be obtained by tracking three points. In section 4 we show that the same approach can be applied to the four by four matrix ${}^{c}_{s}\mathbf{T}$.

3 Dynamically correcting calibration by tracking image points

A modification to the lens of a camera will induce a transformation on the image position of points in the scene. Such transformations may be separated into two classes:

Class 1: Transformations which are independent of the distance to the point in the scene.

Class 2: Transformations which depend on the distance to the point in the scene.

In theory, any movement in the 3D position of the principal point (projection or stenope point) will result in an image transformation of the second class. In reality, when the movement is very small with respect to the focal length, the effects of the distance to the scene point may be much smaller than random pixel noise. In such a case, the transformation may be approximated as belonging to the first class. Transformations in the first class may be corrected by a 3 by 3 projective correction matrix obtained by tracking image coordinates of any three points. Transformations in the second class require tracking the scene position of four points.

In this chapter we experimentally measure the precision of the correction for a change in focus, aperture and vergence angle. Our results show that that convergence about an axis that passes through (or near) the projection point may be corrected by a 3 by 3 projective correction matrix obtained by tracking four points. Changes in focus may be corrected by a 3 x 3 affine transformation, obtained by tracking three points. Small changes in aperture are best not corrected, while large changes are best corrected with a projective correction.

3.1 Dynamic re-calibration of the projective transform

Modifications to the focus of a camera has the effect of shifting, rotating and scaling the image. Such a transformation may be recovered by tracking points in the

image before and after the modification.

Let us define the positions of points in image 1 (before the change) as ${}^1P_2$ and in the image 2 (after the change) as ${}^2P_2$. We seek to model the transformation of these points by a 3 by 3 homogeneous transformation ${}^2_1\mathbf{D}$.

$$\begin{bmatrix} w\, i_2 \\ w\, j_2 \\ w \end{bmatrix} = {}^2_1\mathbf{D} \begin{bmatrix} i_1 \\ j_1 \\ 1 \end{bmatrix}$$

Where w is the homogeneous variable. Let us choose three non-colinear points in the first image ${}^1P_0$, ${}^1P_1$ and ${}^1P_2$. These points can be used to compose a matrix, ${}^1P$.

$${}^1P = [{}^1P_1 \cup {}^1P_2 \cup {}^1P_0] = \begin{pmatrix} i_{11} & i_{21} & i_{01} \\ j_{11} & j_{21} & j_{01} \\ 1 & 1 & 1 \end{pmatrix}$$

Let us designate the corresponding points in the second image as ${}^2P_0$, ${}^2P_1$ and ${}^2P_2$ and use these to compose the matrix ${}^2P$.

$${}^2P_w = [{}^2P_1 w_1 \cup {}^2P_2 w_2 \cup {}^2P_0] = \begin{pmatrix} i_{12} w_1 & i_{22} w_2 & i_{02} \\ j_{12} w_1 & j_{22} w_2 & j_{02} \\ w_1 & w_2 & 1 \end{pmatrix}$$

The transformation from image 1 to image 2 is defined by the homogeneous matrix ${}^2_1\mathbf{D}$ and written as

$${}^2P_w = {}^2_1\mathbf{D}\ {}^1P$$

where the subscript w represents the inclusion of the scale factor coefficients. To solve for ${}^2_1\mathbf{D}$ we invert ${}^1P$ to obtain:

$${}^2_1\mathbf{D} = {}^2P_w\ {}^1P^{-1} \qquad (5)$$

In cases where the transformation from image 1 to image 2 is a composition of a translation, rotation and scale change, the homogeneous variables are given by $w_1 = w_2 = 1$ and ${}^2_1\mathbf{D}$ has the form of an affine transformation. We will call this an affine correction matrix ${}^2_1\mathbf{D}_a$. For such a transformation, ${}^2_1\mathbf{D}$, may be computed from any three non-linear points in image 1 for which a correspondence is known in image 2 by inverting the matrix ${}^1P$.

$${}^2_1\mathbf{D}_a = {}^2P\ {}^1P^{-1} \qquad (6)$$

When ${}^2_1\mathbf{D}$ is not affine, the values of w_1 and w_2 must be computed. For a solution in which no constraints are imposed on the points, we need the correspondence of an additional point. Let us note these this point in the first image as ${}^1P_3$ and with the corresponding point in the second image being ${}^2P_3$. We can use equation 5 to write two equations, one each for the i and j terms from

$${}^2P_3\, w_3 = {}^2_1\mathbf{D}\ {}^1P_3$$

Unfortunately this leaves us with 2 equations for the three unknowns w_1, w_2 and w_3. However, since we know that ${}^2_1\mathbf{D}$ is a transformation from one plane to another, we can write a constraint of the form:

$$w_3 + 1 = A\, w_1 + B\, w_2$$

where A and B are the affine coordinates of point ${}^2P_3$ in the affine basis defined by ${}^2P_0$, ${}^2P_1$, ${}^2P_2$ [5]. Alternatively, we can use the correspondence of the 4 points to write 8 simultaneous equations for the 8 unknown coefficients of the correction matrix ${}^2_1\mathbf{D}$.

3.2 Experimental results

Can focus be corrected with an affine correction matrix or is the full projective form required? To investigate, we placed our cube at a distance of about 105 cm and set the focus to the corresponding distance and calibrated using the technique described above.

After our initial calibration, we placed a paper with an x in front of the cube. We stepped the focus through 7 positions, while tracking 4 points on the "X". We corrected the calibration matrices using both the affine and projective correction methods, and then used the stereo images to reconstruct the upper three corners of the cube. As a control, we also reconstructed without correction (no-corr). We computed the average distance between the observed position of the three points and the true position. The results are shown in table 1.

Encoders	distance	no-corr	D_a	D_p
1	58 cm	0.092	0.044	**0.031**
3000	72 cm	0.105	**0.010**	0.014
6000	95 cm	0.032	**0.007**	0.008
7000	105 cm	0.010	–	–
8000	115 cm	0.030	**0.051**	0.066
10000	162 cm	0.066	**0.039**	0.061
14000	539 cm	0.155	0.093	**0.077**

Table 1 Results of reconstruction of the top three points of cube, using an affine correction matrix (D_a) and a projective correction matrix (D_p) to compensate for focus. The cube is placed a distance of approx 105 cm from the camera head. The most precise value is indicated in bold. The fourth line is the initial calibration position.

Researchers who have never worked with real images are sometimes surprised to learn the imprecision of stereo. The errors shown in table 1 are due to pixel quantization. Such errors are random and fundamental to stereo with digital images. Of course, the larger the image of the stereo points, the smaller the quantization error. Table 1 indicates that the affine and projective correction matrices give very similar precision in reconstruction. We were surprised to observe that the affine correction matrix was more precise for small changes in focus.

For changes in aperture, the situation is more delicate, as illustrated by the following experiment. Using the same scene of the cube, we stepped the aperture of the left camera through 5 position near the middle of its setting. We then performed a 3-D reconstruction for the seven visible corners of the cube and measured the average distance between the reconstructed and true points. The results are shown in table 2. For small changes in aperture, both the affine and projective correction matrices

are vanishingly close to identity. In fact, in such a case, the error due to quantization of the image position of the 4 points dominates, and it is better not to correct the transformation matrices. As the aperture opens, pixels begin to be shifted, and the projective correction began to give better results.

Encoders	no-correction	D_a	D_p
5400	**0.016**	-	-
5800	**0.016**	0.016	0.016
6000	**0.016**	0.016	0.016
6200	**0.022**	0.022	0.022
6400	0.023	0.032	**0.021**

Table 2 Average Reconstruction error for change in aperture, with no correction, affine correction and projective correction.

In our stereo head the vergence rotational axis is nearly aligned with the principal point. In this case, we have speculated that the transformation due to vergence is closely approximated by a simple translation in the image, modelled by an affine correction matrix. Experiments show that this expectation is incorrect.

Table 3 shows results from an experiment in which the head was calibrated to a cube at a distance of 120 cm, and the left camera was then slowly turned. The encoder values and angle of the left camera are shown in the first two columns. The other three columns show the average error for reconstructing the 7 visible corners of the cube without correction ("no-corr"), with an affine correction D_a and with a projective correction, D_p. The affine correction was computed by tracking corners 1, 2 and 4. The projective correction was computed by tracking points 1, 2, 4, and solving for w_1 and w_2 with corner 5.

Encoders	angle	no-corr	Da	Dp
2800	84.45°	0.010	-	-
2950	85.20°	0.513	0.019	**0.018**
3100	85.96°	1.405	0.022	**0.016**
3250	86.71°	1.774	0.031	**0.019**
3400	87.47°	2.117	0.034	**0.012**

Table 3 Average precision obtained for reconstruction of the 7 points of the cube when correcting for vergence with an affine transformation D_a and a projective transformation D_p.

Without a correction, the reconstruction error grows rapidly (21% for a change of 3°). Both the affine and projective correction matrices limit this error to a few percent. However, the average error for the affine transformation continues to grow as a function of the vergence angle, while the error is roughly constant for the projective correction. Our conclusion is that an affine transformation provides an acceptable correction for small vergence angles, but that the most precise correction requires a projective correction matrix.

One might ask how the choice of points influences the precision of the correction. To measure this, we have performed an experiment in which the left camera was converged while tracking all six points of the cube. We then computed the correction matrix with all possible sets of four of the six points. Average 3D reconstruction precision ranged from 1.3% error to 6.2% error. More importantly, the reconstruction precision was proportional to surface of the area enclosed by the four points. In the next section we show how a correction matrix obtained by tracking scene points can corrected for movements in the cameras.

4 Keeping scene coordinates locked on a reference object.

The projective transformations ${}^{L}_{S}\mathbf{M}$ and ${}^{R}_{S}\mathbf{M}$ are rigidly attached to the cameras. If the head moves, the reference system is translated and rotated. By tracking image correspondences and performing 3-D stereo reconstruction, it is possible to derive a 4 x 4 homogeneous correction matrix which keeps the reference frame locked on a scene object.

The projective transformation has the form of a 3 by 4 homogeneous matrix. The 3 dimensional side produces points in image coordinates while the 4 dimensional side refers to scene coordinates. A four by four correction matrix provides a transformation to the scene based reference frame. Such a transformation may be used to change the scene based reference frame.

Let us designate the current coordinate system as 1. Moving the cameras from position 1 to 2 can have the effect of translating and rotating the reference frame. We can express the resulting matrix as a product of the previous matrix and a transformation.

$${}^{L}_{2}\mathbf{M} = {}^{L}_{1}\mathbf{M}\ {}^{1}_{2}\mathbf{T} \qquad {}^{R}_{2}\mathbf{M} = {}^{R}_{1}\mathbf{M}\ {}^{1}_{2}\mathbf{T}$$

To shift the reference back to the object, we require the inverse transformation, ${}^{2}_{1}\mathbf{T}$. We reconstruct the reference points in the new reference frame, 2, using the current calibration matrices ${}^{L}_{2}\mathbf{M}$ and ${}^{R}_{2}\mathbf{M}$. We then compose the four points into a matrix

$${}^{2}R = \left[{}^{2}R_1 \cup {}^{2}R_2 \cup {}^{2}R_3 \cup {}^{2}R_0 \right]$$

We note that in the original reference frame, these same points represent the origin and the three unit vectors, expressed as:

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

To calculate ${}^{2}_{1}\mathbf{T}$ we write

$${}^{1}R = {}^{2}_{1}\mathbf{T}\ {}^{2}R$$

We obtain the correction matrix by inverting ${}^{2}R$.

$${}^{2}_{1}\mathbf{T} = {}^{1}R\ {}^{2}R^{-1} \qquad (7)$$

The following experiment demonstrates locking. We calibrated to our cube, and then translate and/or rotate the cube while tracking four of the corners. We reconstruct the new positions of the tracked corners and compute the

correction matrix. We then update the calibration matrices and compute the corrected position of the 3D points. Error is measured by average distance between the reconstructed points and the true positions. The cube was translated in the x,y plane and rotated about the z axis by carefully moving it around on a table. The correction matrix was computed with points R_0, R_2, R_4, and R_5 as shown in Figure 2. The average error in reconstructing all seven visible cube corners is shown in table 4.

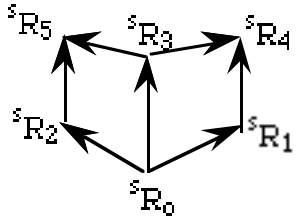

Figure 2 The reference points for a parallelpiped

scene	x, y (cm)	θ_z	No-Corr	Corrected
2	10.0, 0.0	0.0°	0.500	**0.010**
3	0.0, 10.0	0.0°	0.496	**0.013**
4	0.0, 10.0	16.7°	0.420	**0.065**

Table 4 The cube is displaced by 10 cm in x and then in y. It is then rotated by 16.7 degrees. The average error in reconstruction is shown before and after correction. An error of 0.5 represents half a length of the cube, or 10 cm.

To show that the technique works as well when the head is translated and rotated, we performed the same experiment while moving the head translated roughly 10cm and rotated about 10 degrees.

scene	No Correction	Corrected
2	0.487	**0.014**
3	0.267	**0.005**
4	0.430	**0.020**

Table 5 Correction for three head movements.

5 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 a technique to correct camera calibration as the optical parameters of focus and aperture are changed, as the camera is rotated, and as the camera is translated.

After some definitions, we have presented techniques which permit the calibration matrix to be updated by tracking points in the images. It is not necessary to know the 3-D scene position of the points. We have experimentally compared updating with both an affine and projective correction matrix. Our experimental results show that for changes in focus is best made with an affine correction, based on tracking 3 points. For camera convergence, the projective correction, provided by tracking four points, provides a slightly better reconstruction. For aperture, it is best not to correct the matrices after small changes, while large changes are best made with a projective correction. Experiments are currently underway to extend this technique to changes in zoom.

Finally we have shown how tracking four points for which the scene position is known can be used to hop the coordinate reference frame from one object to another. We have also shown how four reconstructed points can be used to keep the reference frame locked onto an object as the head (or object) is moved.

A final conclusion involves calibration. The current wisdom argues 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 including Roger Mohr and Thierry Vieville. Presentations by Jan Koenderinck and by Gunnar Sparr has also provided key inspirations.

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.