Chapter 7

Lens Distortion Calibration Using Point Correspondences

7.1 Introduction

Radial lens distortion can be a significant factor in medium to wide angle lenses. The errors can be 10-100 pixels at the edges of the image [71]. These are typically the lenses used when performing image based 3D reconstruction of large objects or in a confined space. These are also the type of lenses which have been used in the experiments throughout this thesis. This chapter describes a new method for lens distortion calibration using point correspondences in multiple views and using the geometric constraints described in Section 2. There is no need to know either the 3D location of the points or the camera locations. In fact one can use the same point correspondence data which is to be used for subsequent vision tasks such as 3D reconstruction. The method is therefore suitable as an online preprocessing stage.

The simulation results in Section 7.5 show that the geometric configuration of the cameras is important for the success of the calibration. In summary, the cameras must be far apart with large angles between the camera views (i.e. large vergence angles). This result has practical implications for the estimation of structure and motion from image sequences (Chapter 3 and Chapter 5). In such applications the camera motion is typically small. On the one hand, this means that the radial distortion parameters cannot be estimated from the input images, and on the other hand, we find that the radial distortion does not affect the convergence. It only produces a nonlinear distortion in the final motion and structure estimates.

There are two variants of the new calibration method. The first uses two images and the second uses three images. Both methods will be presented but the major focus of this chapter (and the experimental results) will be on the three image method.

7.1.1 The Two Image Method

Given a point in one view we know that the corresponding point in the other view must line along an epipolar line. Given at least 8 point correspondences we can find the epipoles and epipolar lines in a linear manner [26]. With more than 8 points we can find the least squares solution.

The epipolar constraint holds for ideal pinhole cameras. Due to noise in the feature detection and due to lens distortion we get some error. With more than the minimum 8 points we can define a cost function as the RMS of the distances from the points to the corresponding epipolar lines. We then find lens distortion parameters that correct the image coordinates of the feature points to minimize this error. The error function is in general well behaved, and the distortion parameters can easily be found by nonlinear search techniques.

A variant of this method was proposed by Zhang [90] using two cameras from a stereo pair. He found that it was difficult to find the lens distortion parameters unless the distortion was severe or the points were found very accurately. The apparent contradiction between his results and ours can be resolved if we observe that in a typical stereo camera configuration the angle between the two views is very small. As we have noted, in order to get a distinct minimum in the error function, the angle between views must be large.

7.1.2 The Three Image Method

Corresponding points in three images are related by 4 independent trilinear equations. [57]. These equations have 27 parameters which can be found in a linear manner given at least 7 point correspondences. These parameters allow us to reproject corresponding points given in two of the images into the third image.

Due to noise and lens distortion, the reprojection is not perfect. We define our cost function as the RMS reprojection error and find lens distortion parameters that minimize it.

7.2 Related Work

Known world coordinates: The classic method for lens distortion calibration is the bundle adjustment method [66]. It uses one or more views of a calibration object with known 3D coordinates (control points). Using iterative methods it then finds both external parameters (position and orientation) and internal camera parameters. The internal camera parameters include the parameters of the pinhole camera model (principal point, principal distance) and the parameters of lens distortion. See also Weng et al. [86].

Projective constraints: Under perspective projection, straight lines in space project to straight lines in the image. With real lenses the lines appear instead to be slightly to moderately curved. By searching for lens distortion parameters which straighten the lines the *Plumb Line* method and its derivatives [8] [20] [71] [12] find the lens distortion without needing to find the external parameters or the other internal camera parameters. One or more images can be used.

Unknown world coordinates: Stein [71], and Du and Brady [13] use corresponding points or edges in images where the camera has undergone pure rotation to find the internal camera parameters including lens distortion. The 3D location of the points is not required.

7.3 Mathematical Background

We will review here the basic relevant material from Chapter 2. As in the preceding chapters a point in 3D projective space ($\mathcal{P}^3$) will be represented by P_i where the subscript i denotes the i'th point. The subscript i might often be dropped for clarity. Written in homogeneous coordinates: $P = (X, Y, Z, T)$. The point P_i projects onto the image plane of the j'th camera at point p_{ij} where $p = (x, y, 1)$.

The 4×3 camera projection matrix for the j'th camera will be denoted $\tilde{\mathbf{A}}_j$. Using homogeneous coordinates the perspective projection can be written as:

$$\rho_{ij} p_{ij} = \tilde{\mathbf{A}}_j P_i \tag{7.1}$$

for points $i = 1...n$ in images $j = 1...k$. ρ_{ij} is an unknown scale factor which is different for each point and each image.

7.3.1 The Epipolar Constraint

As shown in Section 2.3, For every pair of cameras (j,k) there exists a 3×3 *fundamental matrix* ($\mathbf{F}_{jk}$) such that:

$$p_{i,j}^T \mathbf{F}_{jk} p_{i,k} = 0 \tag{7.2}$$

The above equation is the epipolar constraint. Given a particular camera geometry, embodied in the matrix $\mathbf{F}$, each point in one image defines a line v on which the corresponding point in the second image must lie. The equation of that line is given by:

$$v^T \mathbf{F}_{jk} p_{i,k} = 0.$$

Given 8 or more point correspondences, the fundamental matrix ($\mathbf{F}$), can be determined up to a scale factor using the *eight point algorithm* which is described in [26] together with many important implementation details.

7.3.2 The Trilinear Tensor Constraint

In Section 2.4 we derive the trilinear equations of [57]. Given a set of 3D points there exists a set of trilinear equations between the projections of those points into any three perspective views. In total there exist 9 such equations for each point with at most 4 being independent.

Four of the nine equations are as follows (equation 2.42):

$$\begin{aligned}
x''\mathcal{T}_i^{13}p^i - x''x'\mathcal{T}_i^{33}p^i + x'\mathcal{T}_i^{31}p^i - \mathcal{T}_i^{11}p^i &= 0, \\
y''\mathcal{T}_i^{13}p^i - y''x'\mathcal{T}_i^{33}p^i + x'\mathcal{T}_i^{32}p^i - \mathcal{T}_i^{12}p^i &= 0, \\
x''\mathcal{T}_i^{23}p^i - x''y'\mathcal{T}_i^{33}p^i + y'\mathcal{T}_i^{31}p^i - \mathcal{T}_i^{21}p^i &= 0, \\
y''\mathcal{T}_i^{23}p^i - y''y'\mathcal{T}_i^{33}p^i + y'\mathcal{T}_i^{32}p^i - \mathcal{T}_i^{22}p^i &= 0.
\end{aligned} \tag{7.3}$$

where x', y' and x'', y'' are the image coordinates in the first and second images respectively. p is the image point in the third view. $\mathcal{T}_i^{jk}$ are the 27 coefficients arranged in a $3 \times 3 \times 3$ tensor. Seven point correspondences give 28 equations which are enough to recover the coefficients up to a scale factor.

Given 7 points in 3 images one can recover the tensor. Given the tensor and the locations of additional corresponding points in two of the images, one can calculate the locations of those points in the third image.

7.3.3 The Lens Distortion Model

These projective constraints assume a perfect pinhole model. This is a good model for long focal lengths, but medium to wide angle lenses have noticeable lens distortion. The standard model for lens distortion [66] is a mapping from the distorted image coordinates, (x_d, y_d), that are observable, to the undistorted image plane coordinates, (x_u, y_u), which are not physically measurable using the equation:

$$\begin{aligned} x_u &= x_d + x'_d(K_1 r'^2_d + K_2 r'^4_d + \cdots) \\ y_u &= y_d + y'_d(K_1 r'^2_d + K_2 r'^4_d + \cdots) \end{aligned} \tag{7.4}$$

where K_1 and K_2 are the first and second parameters of radial distortion and:

$$r'^2_d = x'^2_d + y'^2_d = (x^2_d - c_{xr})^2 + (y^2_d - c_{yr})^2 \tag{7.5}$$

It has been shown in [71] that allowing the center of radial distortion, (c_{xr}, c_{yr}) to be different from the principal point is a good approximation to adding the terms for decentering distortion as given in [66].

7.4 The Algorithm

The step by step algorithm is as follows:

1. Find point correspondences between 3 views.
2. Make an initial guess of the distortion parameters: an appropriate guess for (C_x, C_y) is the center of the image. Choose K_1 such that $K_1 \times r^3 \leq 0.1$ where r is the distance to the corner of the image.
3. Using the distortion parameters compute the undistorted feature locations (equation 7.4).
4. Compute the trilinear tensor. Then use the tensor to compute the reprojection error for each point (equation 7.3).
5. Adjust the distortion parameters to reduce the reprojection error. This is done using the Levenberg-Marquart algorithm (see Section 7.6.1).
6. Loop back to step 3 till convergence.

Comments: It is best to search initially for only one parameter of radial distortion (K_1). After finding the best K_1, one can use that value as a starting guess and try searching for other parameters as well. One can search for different values of K_1 for each camera. If one gets a much smaller RMS reprojection error than when enforcing the same value of K_1 for all the cameras then it is a good bet that the cameras are in fact different.

7.5 Simulation

Simulations were performed (using Matlab) to test the effect of camera configurations and noise on the shape of the cost function. Eighty 3D points were generated, uniformly distributed between -1 and 1 in the X,Y and Z directions. These points were projected using a projective camera model with optical axis aligned with the Z direction. The camera was located at $Z_0 = -4$. This simulates a camera with a viewing angle of about 55^o. The focal length was $f = 1000 pixels$.

The 3D points were rotated prior to projecting which is equivalent to the camera rotating around the center of the points at a given radius Z_0. Normally distributed random noise was added to the image coordinates which were then distorted using (equation 7.4) with only one radial distortion parameter, $K_1 = 2.5 \times 10^{-7}$, the value found in Section 7.6.

Consider the following 3 camera setup (Figure 7-1). Cameras 1 and 2 are rotated θ^o and $-\theta^o$ around the Y axis respectively. Camera 3 is rotated around the X axis with varying angles $\alpha = 0, 10^o, 20^o \ldots 80^o$. The radial distortion is varied from 0 to $2 \times K_1$ in $0.1 \times K_1$ increments and the reprojection error from camera images 1 and 2 to camera image 3 is calculated.

Due to lack of space we show only the simulation results for $\theta = 20^o$ and additive noise $\sigma = 0.3$ *pixel* (Figure 7-2). Note that for small α the error as a function of K_1 is very shallow and it would be hard to find the correct K_1 value. With larger noise values the minimum is sometimes outside the plotted range for $\alpha = 0$. Other simulation experiments show that increasing θ towards 40^o the curves become steeper and reducing θ makes them shallower. At $\theta = 2^o$ the results are unreliable even for large α values.

7.6 Experiments with Real Images

7.6.1 Experimental Details

The Images, Features and Feature Detector

Figure 7-3a, Figure 7-3b and Figure 7-3c show a typical image triplet used in the experiment (same camera). The camera motion included a large, unknown, degree of rotation and translation. Figure 7-3d was taken from approximately the same angle as Figure 7-3c but the camera was twice as far away and the focal length longer (zoom).

Three lines of feature points were marked on a flat sheet of metal. The features were two small triangles touching at a point. These points are saddle points in the gray level image and result in local minima in the determinant of the Hessian of the image. The point is located to subpixel accuracy by locally fitting the determinant of the Hessian image to a paraboloid and finding the minimum analytically.

Correspondence was performed manually. In total 75 feature points were detected: 7 points on the top surface of the metal cube, a total of 10 points on the other metal objects and 55 coplanar points in 3 lines plus one nearly coplanar point.

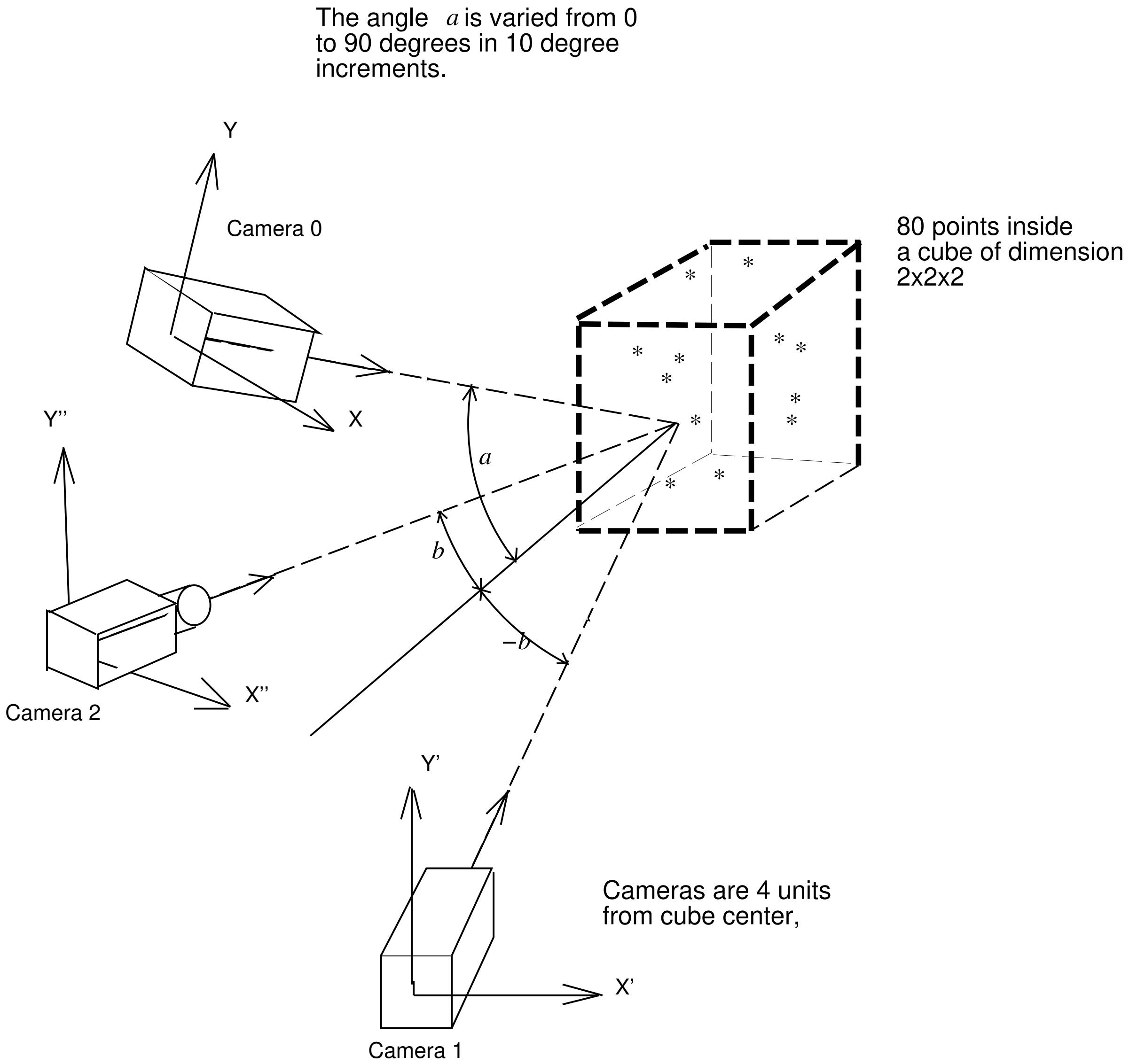
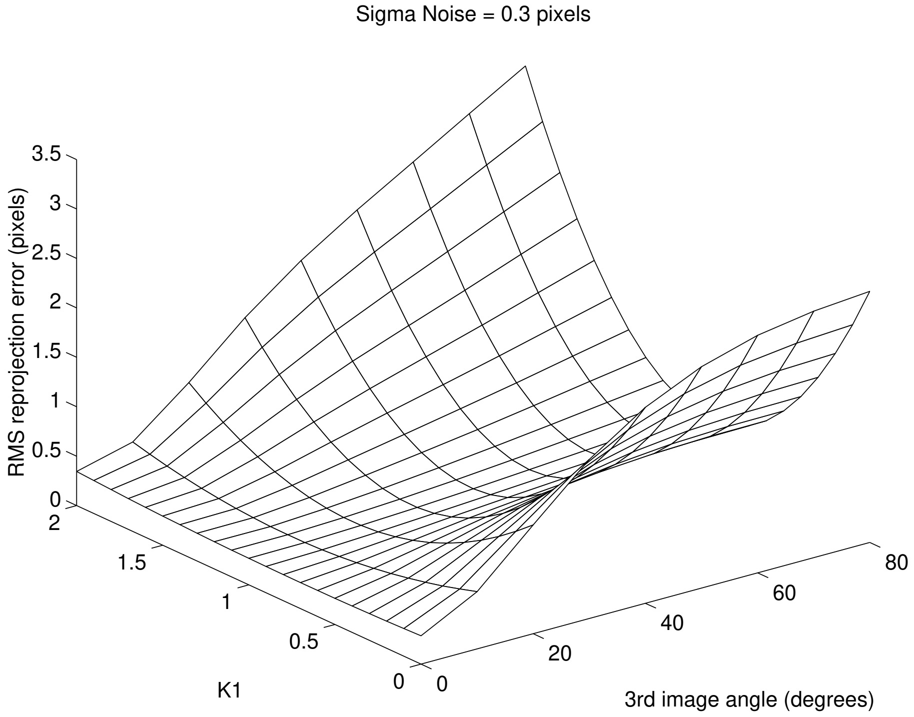


Figure 7-1: *Simulated camera setup.*



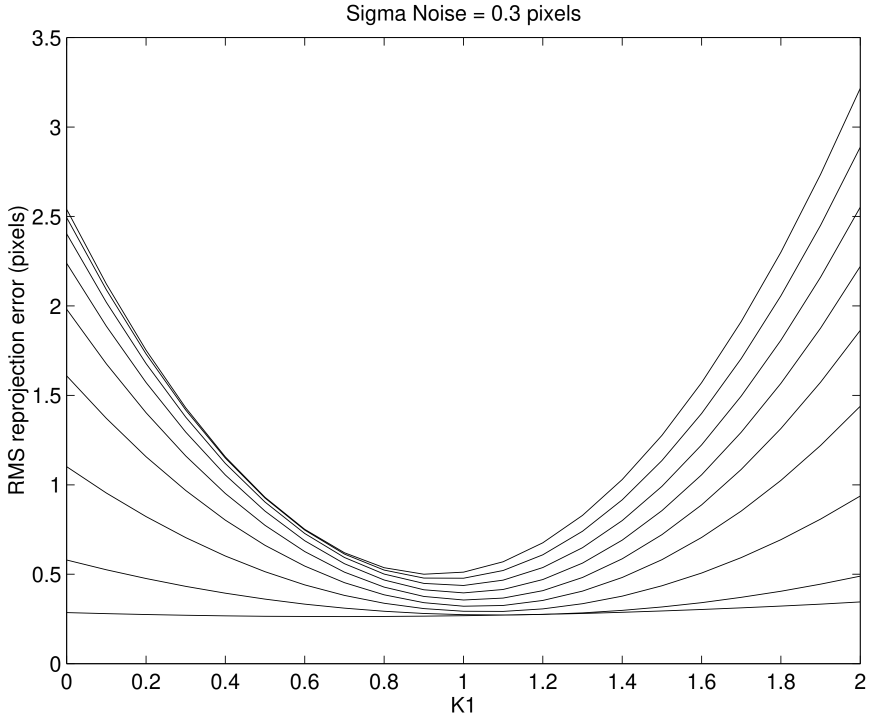
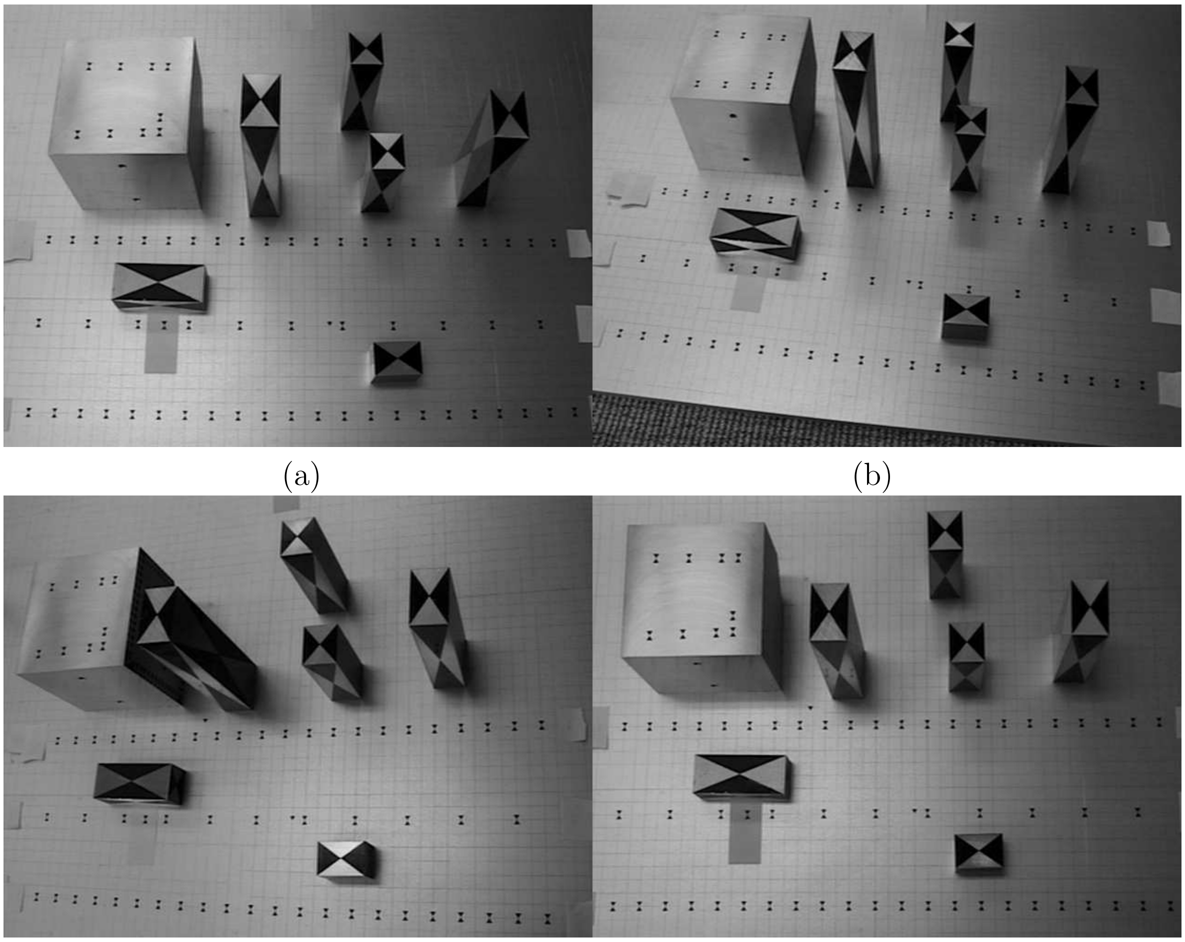


Figure 7-2: *Simulation results. Reprojection error while varying the radial distortion correction from 0 to 2 times the correct value and varying the viewing angle of the 3rd camera. Gaussian noise of sigma=0.3 was added to feature coordinates.*



(a) (b) (c) (d)

Figure 7-3: *(a,b,c) The images used for calibration. (d) Alternative third image taken from further away with a longer focal length. Note that in this case the lines are not curved.*

Hardware and Software

The camera was a SONY DCR-VX1000, a high quality digital handycam. The lens was the built-in zoom lens open to the widest angle giving a corner to corner viewing angle of approximately 55 degrees. Image capture and processing was performed on an SGI Indy workstation.

The camera parameters were found using a nonlinear optimization program based on the subroutine LMDIF from the software package MINPACK-1 [47]. This subroutine uses a modified Levenberg-Marquart algorithm.

7.6.2 Experiment 1: Finding the Lens Distortion Parameters

Using all 75 points in the 3 images (Figure 7-3a, Figure 7-3b, and Figure 7-3c) we computed the trilinear tensor (see Section 7.3) and then the reprojection error from Image 2 (Figure 7-3b) and Image 3 (Figure 7-3c) to Image 1 (Figure 7-3a). No robust estimation techniques were used. I will denote the radial distortion parameters found as K_{1best} and K_{2best}.

Convergence Range

One must supply the non-linear search routine with an initial guess. Many initial guesses were tried for the radial distortion parameter (K_1). The second parameter (K_2) was held at zero and the center of distortion at the image center (320,240). The non-linear search would converge to a value $K_{1best} = 2.58 \times 10^{-7}$ with a starting guess anywhere in the range $(2.5 \times 10^{-2} : 2.5 \times 10^{-12})$. (i.e. through at least 4 orders of magnitude in either direction). Convergence took between 3 and 10 iterations.

Assuming two radial distortion parameters (K_1, K_2), the non-linear search would converge to a value $K_{1best} = 4.755 \times 10^{-7} \pm 0.1\%$ and $K_{2best} = -1.153 \times 10^{-12} \pm 0.5\%$ with a starting guess anywhere in the range $K_1 = (2.5 \times 10^{-5} : 2.5 \times 10^{-9})$ and $K_2 = (-1 \times 10^{-7} : 1 \times 10^{-7})$. (i.e. through at least 2 orders of magnitude in either direction of K_1). Convergence took between 3 and 10 iterations.

The Shape of the Cost Function

One of the dangers of using non-linear methods is the presence of local minima. Figure 7-4a shows the cost function evaluated over a range of values of K_1 from 0 to $2 \times K_{1best}$ (twice the final estimated value) in $0.1 \times K_{1best}$ increments. The function is smooth, qualitatively parabolic in shape and with no local minima. These characteristics have repeatedly been seen with no exception in many similar experiments with varying ranges and resolutions.

Assuming Different Parameters for Each Camera

The calibration method can work for different cameras. We can search for 3 separate sets of calibration parameters, one set for each camera. Allowing for one radial distortion parameter for each camera we get:

Table 7.1: Radial distortion parameter K_1 obtained using various ordering of the images

Order	K_1	RMS error
a,b,c	2.58162e-07	e=6.773012e-02
a,c,b	2.5811e-07	e=6.773012e-02
c,a,b	2.4687e-07	e=7.554849e-02
c,b,a	2.46804e-07	e=7.554849e-02
b,a,c	3.19903e-07	e=6.590745e-02
b,c,a	3.19834e-07	e=6.590745e-02
mean	2.7498e-07	
std	3.9313e-08	
std/mean	0.14	

$K_{1best} = (2.55 \times 10^{-7}, 2.69 \times 10^{-7}, 2.19 \times 10^{-7},)$
The 3 images were taken by the same camera so the 3 values for K_{1best} should come out the same. In fact, they are within 20% of each other. Better results might be achieved with a larger number of points.

The third image (Fig: 7-3c) was then replaced with an image taken from further away using a longer focal length (Figure 7-3d). In other words it is a 'different camera'. Allowing for one parameter for radial distortion for each camera the results were:
$K_{1best} = (2.15 \times 10^{-7}, 2.20 \times 10^{-7}, -5.93 \times 10^{-8})$
The third camera appears to differ significantly and has much less distortion as we would expect from the longer focal length.

Which Image is Used for Reprojection

The method has an asymmetry in that two of the images are used to reproject into the third image. Thus, the third image is treated differently. All the possible permutations of the 3 images were tried. The results are summarized in Table 7.1. The standard deviation in the value of K_{1best} was 14% which is not large. Nevertheless, to avoid the ambiguity one can use as a cost function the sum of reprojection errors in the 3 distinct possibilities.

7.6.3 Experiment 2: Testing the Radial Distortion Parameters in the Projective Domain

We have no ground truth about the radial distortion parameters. In order to test the radial distortion parameters, a special set of feature points was used. 44 of the points are coplanar. They are arranged in 2 straight lines of 22 equally spaced points each.

Collinearity of Points

Using one of the images (Figure 7-3a), for each of the two sets of 22 points the image coordinates were fit with a best straight line. The RMS distance of the points from

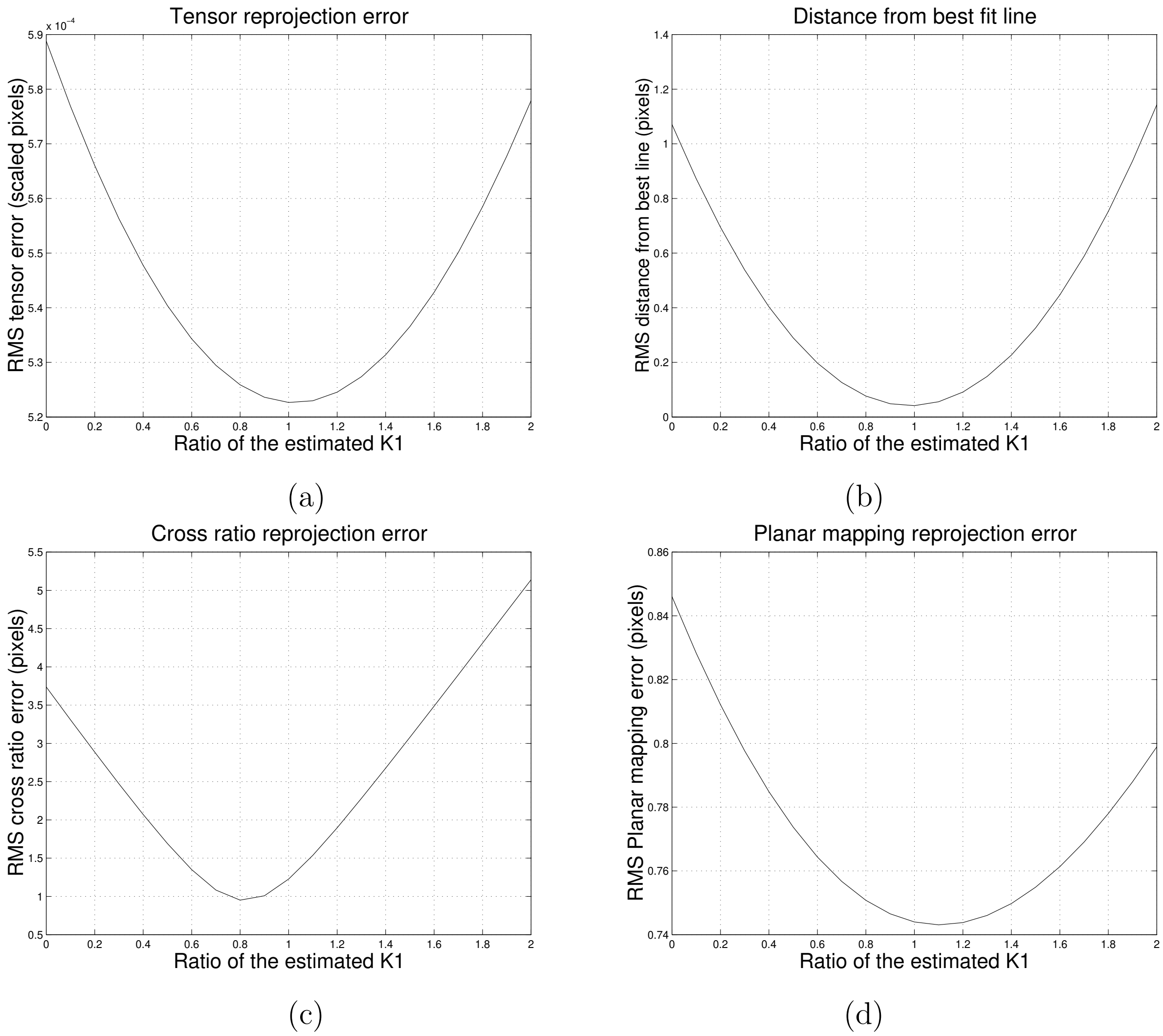


Figure 7-4: *(a) The tensor reprojection error as a function of the value of the radial distortion K_1. (b,c,d) The error in other projective measures as a function of the value of the radial distortion K_1.*

the corresponding lines was then computed. This is the measure used in the *plumb line method* described in Section 7.2.

Figure 7-4b shows the average distance as a function of the radial distortion parameter K_1, where K_1 varies from 0.0 to $2.0 \times K_{1best}$ in $0.1 \times K_{1best}$ steps. The function's minimum is at $K_1 = K_{1best}$. The average distance of points from the line drops from 1.15 pixels to 0.05 pixels after correction.

Cross Ratio of Points

The cross ratio of four points along a line is a projective invariant. Taking the 1st, 2nd and 22nd points along each of the lines, one can compute the cross ratio using any of the other points:

$$r_i = \frac{\frac{\alpha_1 - \alpha_2}{\alpha_1 - \alpha_i}}{\frac{\alpha_{22} - \alpha_2}{\alpha_{22} - \alpha_i}} \tag{7.6}$$

One can invert equation (7.6) to obtain:

$$\alpha_i = \frac{r_i \alpha_1 (\alpha_{22} - \alpha_2) - r_i \alpha_{22} (\alpha_1 - \alpha_2)}{r_i (\alpha_{22} - \alpha_2) - r_i (\alpha_1 - \alpha_2)} \tag{7.7}$$

Given that we know that the points along each line are equally spaced we can use equation (7.7) to estimate the location of points 3 through 21 given the location of points 1, 2 and 22. The errors are due to feature detection error and also radial distortion. Figure 7-4c shows the RMS error in the estimation of all the 38 points as a function of K_1. The minimum error is obtained at a value of K_1 close $(0.8 \times K_{1best})$ to that obtained using the new calibration method.

Planar Mapping of Points Between Two Images

Given two sets of image points obtained from two views of a planar object one can compute a transformation from one set of points to the other which takes the form:

$$p' = \mathbf{A}p \tag{7.8}$$

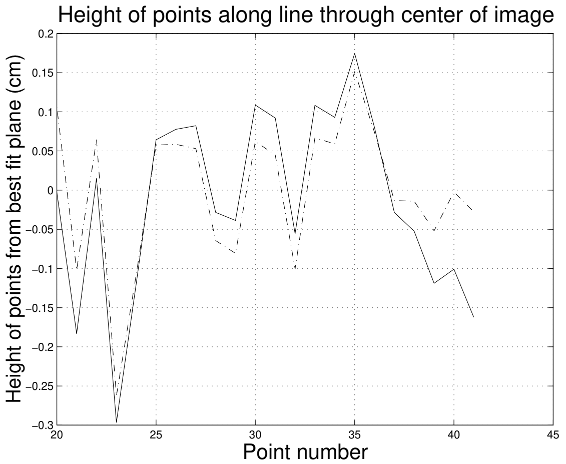
Where p and p' are the corresponding image points in the first and second images and $\mathbf{A}$ is a 3×3 homography matrix. Given more than 4 points one can compute $\mathbf{A}$ in a least squares manner. One can then use the matrix $\mathbf{A}$ to map all the planar points from one image to the second. Figure 7-4d shows the RMS reprojection error using the images in Figure 7-3a and Figure 7-3b. The minimum error is obtained near the value of K_1 found using the distortion calibration.

7.6.4 Experiment 3: Euclidean Reconstruction

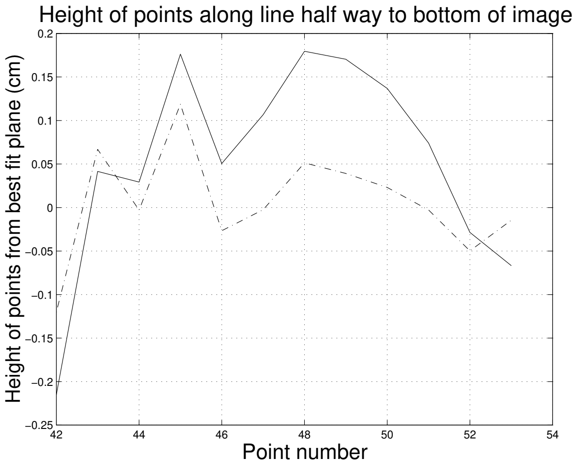
Projective reconstruction from the three images was performed according to [61]. Transformation to Euclidean 3D coordinates then required 5 control points. Three points were chosen from the planar surface and two other points were chosen such that the five were in *general position*, no 4 points coplanar. The world coordinate system was chosen such that the planar surface was the XY plane.

The reconstruction was performed twice, once with the original feature coordinates and once with the corrected coordinates using $K_1 = 2.58 \times 10^{-7}$. The best plane was found using the points which came from the planar surface. Figure 7-5 shows the height of the points from the best plane for the three lines. In Figure 7-5c one can see that for the line close to the edge of the image the line is distinctly curved in depth and that correcting for radial distortion removes the curvature. The curve along the whole length of the line is about $0.3cm$ which is about 1% of the total length of the line ($42cm$). The curvature of the lines closer to the center of the image is smaller.

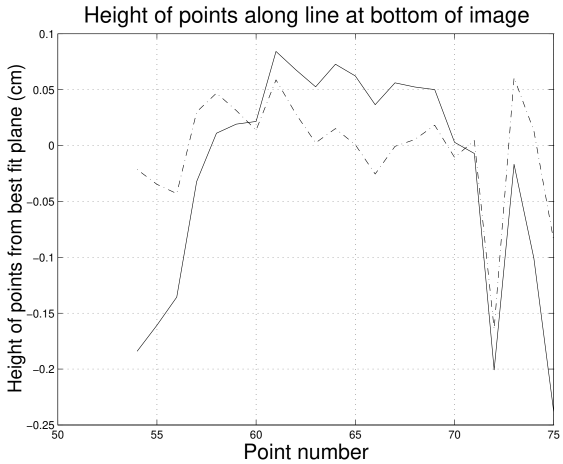
Point number 45 in Figure 7-5b is the point that was drawn on the piece of paper and was in fact not coplanar with the other points. In the case of the corrected points reconstruction (dot dash line) it is clearly stands apart.



(a)



(b)



(c)

Legend:	
———	Reconstruction using uncorrected points.
_ . _ . _	Reconstruction using corrected points.

Figure 7-5: *The height of the reconstructed 3D points from the best fit plane. Figures (a), (b) and (c) are the points along the lines which pass close to the center of the image, half way down and close to the bottom, respectively.*

7.7 Discussion

A new method for determining the lens distortion has been described. Although the method is iterative it has been shown that the method converges fast and that the range of convergence is very wide. The resulting distortion parameters are reasonable and have been shown to agree with those determined by other methods.

Since this method uses only point correspondences in multiple views it fits very naturally inside feature based 3D reconstruction systems. It has been shown here that in the case of a 55^o lens, when the distortion is hardly noticeable in the image, radial distortion can cause a small but noticeable error in reconstruction results. Since the computational overhead of this method is small it makes sense to incorporate this method in all 3D reconstruction systems which use medium and wide angle lenses and where accuracy is desired. Especially, since most of the computational machinery required for tasks such as finding the trilinear tensor or fundamental matrix probably already exists in the overall system.

The Manual of Photogrammetry [66] warns us that "the strong coupling that exists between interior elements of orientation [principal point, focal length] and exterior elements can be expected to result in unacceptably large variances for these particular projective parameters when recovered on a frame-by-frame basis". This is also applicable to lens distortion parameters and it was seen in Section 7.6.2 and Section 7.6.2. Note that the interior and exterior parameters are determined implicitly in the fundamental matrix or the trilinear tensor. To determine the actual lens distortion parameters one can use many images from a variety of locations and apply this method to many triplets of those images. The computational complexity of the method is linear in the number of triplets and number of feature points.

Chapter 8

Summary and Future Work

We have presented a new method for recovering motion and depth from images sequences. The method we presented, the *model-based brightness constraints*, combines photometric constraints in the form of the optical flow constraint equation, together with geometric constraints between corresponding points and lines in three views. Thus, following the 'direct methods' approach of Horn & Weldon [34], we avoid the need to use feature correspondences .

Results from experiments with real images demonstrate that the method can accurately recover the motion parameters even when the motion includes both translation and rotation, and the method can accurately recover true Euclidean depth if the internal camera parameters are known. The reconstruction results degrade gracefully if there are errors in the internal camera parameters.

We can recover the motions reliably if the two translations are not collinear. We have also shown that the method, as presented in this thesis, does not work in the case of collinear motion unless we can assume pure translation. Of course, this does not constitute a proof that direct estimation of motion from monocular sequences cannot be done in the collinear case. We have presented some success in motion estimation from collinear sequences in the case of pure translation but it is not clear how to extend this to the rotation case. The problem seems to lie in the first order approximation of the optical flow constraint equation.

The lack of a solution to the collinear motion limits some of the potential applications (for example, movie compression). This technique can still be used for site modeling from video streams, but care must be taken when capturing the image data. Alternatively, an active vision head could be used, and controlled to make sure that the camera translated in two distinct directions. The motion control of the active vision head need not be precise, as we have shown that the motion can be accurately recovered from the images. In general, collinear motion is not ideal for 3D site reconstruction since there will potentially be many lines in the image which lie on, or close to, epipolar lines, and cannot provide depth information. Using two distinct motion directions avoids this problem.

Monocular image sequences have other drawbacks for 3D site reconstruction which are independent on the method used for estimating the motion and depth. There is the scale ambiguity. This makes it hard to align 3D reconstructions from different parts of the sequence since the magnitude of translation might have changed and so the 3D reconstructions will be at different scales. If the actual translation is close to zero, as often happens with hand held cameras then the method becomes unstable. These drawbacks can be overcome by using stereo image sequences. We have presented a new method for recovering camera motion and 3D shape from a moving stereo rig. Using this method it is simple to bring 3D reconstructions from various view points into the same coordinate system.

8.1 Future Work

Affine Camera Models

It would be interesting to try and extend this framework to orthographic, weak perspective and affine camera models. Recently, Quan & Kanade [54] have shown that for affine cameras there exists a constraint between the *directions* of corresponding lines in three views which takes the form of a $2 \times 2 \times 2$ trilinear tensor. It is not clear how to combine this geometric constraint with the optical flow constraint equation which provides point-line correspondences.

Reconstruction Under Changing Lighting Conditions

In this thesis we use the constant brightness assumption to derive the optical flow constraint equation. This is a good approximation to the images of non-specular surfaces captured by a moving camera. It does not hold in general, in the case of a moving rigid object in front of stationary camera. Jourjine [37] provides a theoretical extension to this framework to the case of rigid object motion assuming a Lambertian surface model and parallel lighting. In this case both surface and albedo can be recovered.

Tracking and Reconstruction for Non-rigid Motion

How do we extend this framework to allow for more complex motion models such as smooth non-rigid motion. For example, suppose we are trying to reconstruct the head of someone talking or a person moving. We could use multiple cameras to provide us with synchronous images from multiple views (binocular video streams). How do we combine sequences of such views to form an extended scene reconstruction as was done in Chapter 5?

Where to Look Next?

Site reconstruction, the creation of 3D model of the surroundings, provides an exciting challenge beyond geometry and photometry, and deeper into the realm of artificial intelligence. An active vision system for site reconstruction has to reason about the current 3D model and decide where to look (and where to move) to efficiently add information to the 3D model.

In the 3D reconstruction from Chapter 5, shown again in Figure 8-1, there are noticeable holes: behind the trees and other obstacles. Some of these holes could have been filled in, if the camera was able to actively look down while passing overhead. Filling in other gaps would require actively changing the camera path. A complete system should also be able to detect regions of interest and sample those areas with higher resolution in order to correctly determine the shape of objects. It is an engineering detail to decide if this is to be done by a 'foveated' camera (a camera with non-uniform spatial sampling) or by a separate camera with a longer focal length.

The capability to reason about the 3D world and make active changes in the behavior is not only useful for site reconstruction but also for search and rescue operations. The system should be able to determine where an object of certain size might be hidden from view, and then determine how to move to search those areas.

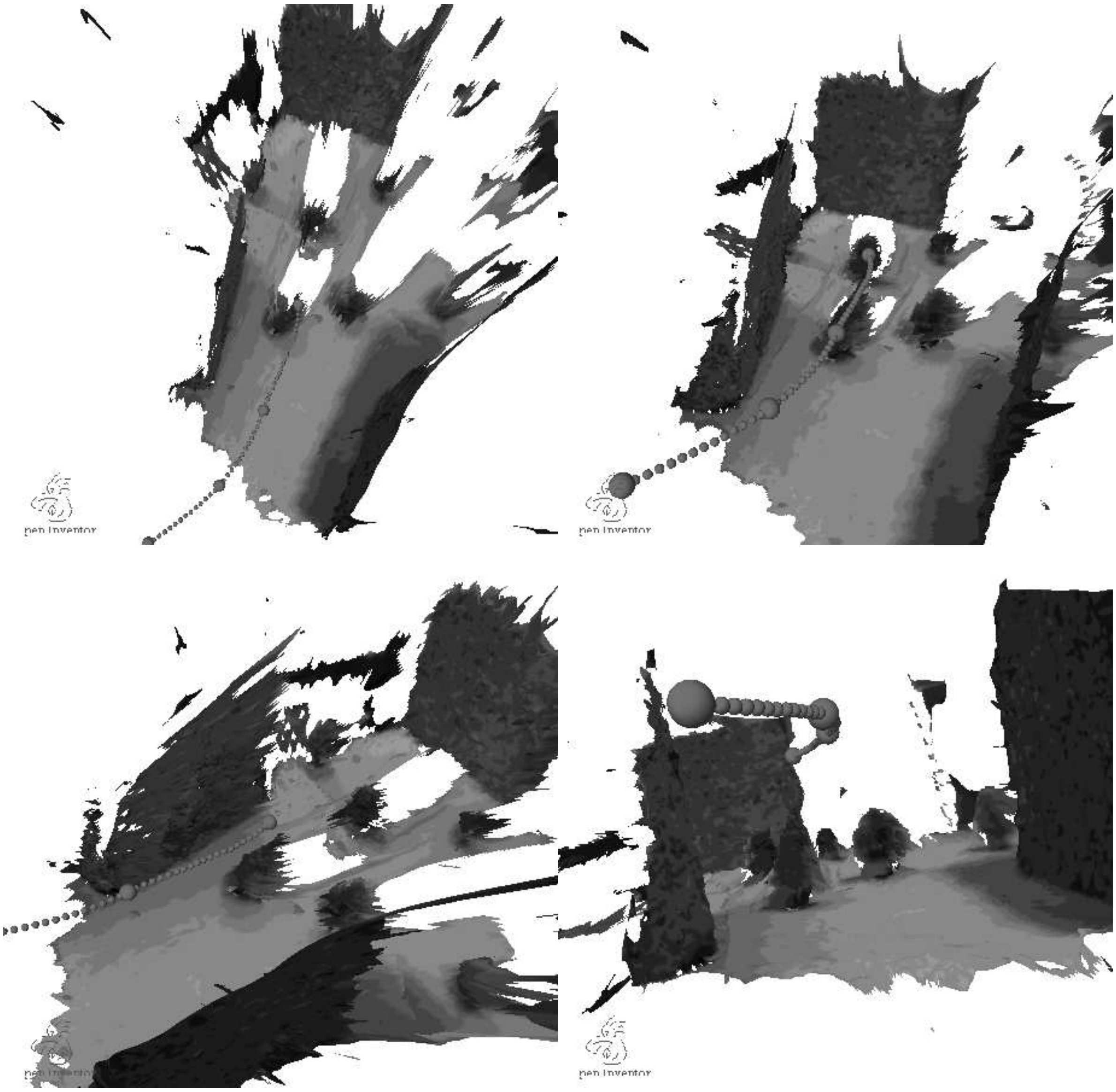


Figure 8-1: *Euclidean reconstruction of the extended scene created by aligning four separate reconstructions into one coordinate system using the recovered camera motion. The small spheres show the camera path. The large spheres show the camera positions of the four Euclidean reconstructions. There are noticeable holes in the reconstruction: behind the trees and other obstacles. Some of these holes could be filled in the camera was able to actively look down while passing overhead. Other holes would require changing the camera path.*

Bibliography

[1] M. Abdel-Mottaleb, R. Chellappa, and A. Rosenfield. Binocular motion stereo using map estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 321–327, New York, NY, June 1993.

[2] S. Avidan and A. Shashua. Tensorial transfer: On the representation of $n > 3$ views of a 3D scene. In *Proceedings of the ARPA Image Understanding Workshop*, Palm Springs, CA, February 1996.

[3] S. Avidan and A. Shashua. View synthesis in tensor space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.

[4] A. Azarbayejani and A. P. Pentland. Recursive estimation of motion, structure and focal length. *pami*, 17(6):562–575, June 1995.

[5] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, June 1992.

[6] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, June 1992.

[7] J. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center, 1990.

[8] D. C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37:855–866, 1971.

[9] T. Buchanan. On the critical set for photogrammetric reconstruction using line tokens in p3(c). *Geometriae Dedicata*, 44:223–232, 1992.

[10] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

[11] R. Deriche and G. Giraudon. Accurate corner detection: An analytical study. In *Proceedings of the International Conference on Computer Vision*, pages 66–70, Osaka, Japan, December 1990.

[12] F. Devernay and O. D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *Proceedings of SPIE Conference*, San Diego, CA, July 1995.

[13] F. Du and M. Brady. Self calibration of the intrinsic parameters of cameras for active vision systems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 477–482, New York, NY, June 1993.

[14] R. Dutta and M. Snyder. Robustness of correspondence based structure from motion. In *Proceedings of the International Conference on Computer Vision*, pages 106–110, Osaka, Japan, December 1990.

[15] O. Faugeras and T. Papadopoulo. A nonlinear method for estimating the projective geometry of three views. To be published ICCV 98, January 1998.

[16] O. D. Faugeras. *Three Dimensional Computer Vision: a Geometric Viewpoint*. The MIT Press, Cambridge, MA, 1993.

[17] O. D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proceedings of the International Conference on Computer Vision*, pages 951–956, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[18] C. Fermuller and Y. Aloimonos. Global rigidity constraints in image displacement fields. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[19] C. Fermuller and Y. Aloimonos. On the geometry of visual correspondence. *International Journal of Computer Vision*, 21(3):223–47, February 1997.

[20] J. G. Fryer and S. O. Mason. Rapid lens calibration of a video camera. *Photogrammetric Engineering*, 55(4):437–442, April 1989.

[21] P. Fua. Reconstructing complex surfaces from multiple stereo views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1078–1085. IEEE Computer Society Press, June 1995.

[22] E. Grosso and M. Tistarelli. Active/dynamic stereo vision. *Pattern Analysis and Machine Intelligence*, 17(11):1117–1128, November 1995.

[23] K. J. Hanna and N. E. Okamoto. Combining stereo and motion analysis for direct estimation of scene structure. In *Proceedings of the International Conference on Computer Vision*, Berlin, Germany, May 1993.

[24] R. Hartley. Lines and points in three views — a unified approach. In *Proceedings of the ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[25] R. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of the International Conference on Computer Vision*, pages 882–887, Cambridge, MA, June 1995.

[26] R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the International Conference on Computer Vision*, pages 1064–1070, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[27] R. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[28] D. J. Heeger and A. Jepson. Simple method for computing 3d motion and depth. In *Proceedings of the International Conference on Computer Vision*, pages 96–100, Osaka, Japan, December 1990.

[29] J. Heel. Direct estimation of structure and motion from multiple frames. AI Memo 1190, Massachusetts Institute of Technolgy Artificial Intelligence Labratory, March 1990.

[30] A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the International Conference on Computer Vision*, pages 1058–1063, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[31] B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, Mass., 1986.

[32] B. K. P. Horn. Relative orientation revisited. *Journal of the Optical Society of America*, 8:1630–1638, 1991.

[33] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[34] B. K. P. Horn and E. J. Weldon, Jr. Direct methods for recovering motion. *International Journal of Computer Vision*, 2:51–76, 1988.

[35] T. Huang and C. Lee. Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11:536–540, 1989.

[36] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization'. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, Washington, June 1994.

[37] A. N. Jourjine. Surface and albedo from a sequence of images. Siemens Corp, April 1997.

[38] T. Kanade, Okutomi, and Nakahara. A multiple-baseline stereo method. In *Proceedings of the ARPA Image Understanding Workshop*, pages 409–426. Morgan Kaufmann, San Mateo, CA, January 1992.

[39] R. Koch. 3-d surface reconstruction from stereoscopic image sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–114. IEEE Computer Society Press, June 1995.

[40] R. Kumar and P. Anandan. Direct recovery of shape from multiple views: A parallax based approach. In *Proceedings of the International Conference on Pattern Recognition*, Jerusalem, Israel, October 1994.

[41] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[42] H. Longuett-Higgins and K. Prazdny. The interpretation of a moving retinal image. In *Proceedings of the Royal Society of London B*, volume 208, pages 385–397. Royal Society of London, 1980.

[43] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings IJCAI*, pages 674–679, Vancouver, 1981.

[44] S. Maybank. The critical line congruence for reconstruction from three images. *Applicable Algebra in Engineering Communication and Computing (AAECC)*, 6:89–113, 1995.

[45] I. S. McQuirk. An analog vlsi chip for estimating the focus of expansion. AITR 1577, MIT, 1996.

[46] D. Michaels. *Exploiting Continuity-in-Time in Motion Vision*. PhD thesis, Department of Electrical engineering and Computer Science, Massachusetts Institute of Technology, May 1992.

[47] J. J. More, B. S. Garbow, and K. E. Hillstrom. User guide for minpack-1. Technical Report ANL-80-74, Argonne National Laboratory, Argonne, Illinois, August 1980.

[48] J. Mundy and A. Zisserman. Appendix — projective geometry for machine vision. In J. Mundy and A. Zisserman, editors, *Geometric invariances in computer vision*. MIT Press, Cambridge, 1992.

[49] H. H. Nagel. Direct estimation of optical flow and of its derivates. In G. A. Orban and H. H. Nagel, editors, *Artificial and Biological Vision Systems*, Basic Research Series, pages 193–224. Springer-Verlag, 1992.

[50] N. Navab, O. D. Faugeras, and T. Vieville. The critical sets of lines for camera displacement estimation: A mixed euclidean-projective and constructive approach. In *Proc. 4th International Conference on Computer Vision*, Berlin, Germany, May 1993. IEEE.

[51] N. Navab, R. Deriche, and O. D. Faugeras. Recovering 3d motion and structure from stereo and 2d token tracking cooperation. In *Proceedings of the International Conference on Computer Vision*, pages 513–516, Osaka, Japan, December 1990.

[52] S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *Pattern Analysis and Machine Intelligence*, 9(1):168–176, 1987.

[53] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, second edition edition, 1992.

[54] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *Pattern Analysis and Machine Intelligence*, 19(8):834–845, August 1997.

[55] J. Reiger and D. T. Lawton. Processing differential image motion. *Journal of the Optical Society of America*, 2:354–359, 1985.

[56] H. Sawhney. 3D geometry from planar parallax. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 929–934, Seattle, Washington, June 1994.

[57] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[58] A. Shashua. Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. Submitted for journal publication, June 1997.

[59] A. Shashua and S. Avidan. The rank4 constraint in multiple view geometry. In *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, April 1996.

[60] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 483–489, Seattle, Washington, 1994.

[61] A. Shashua and N. Navab. Relative affine structure: Canonical model for 3D from 2D geometry and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):873–883, 1996.

[62] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *Proceedings of the International Conference on Computer Vision*, June 1995.

[63] A. Shashua and P. Anandan. Trilinear constraints revisited: Generalized trilinear constraints and the tensor brightness constraint. In *Proceedings of the ARPA Image Understanding Workshop*, Palm Springs, CA, February 1996.

[64] A. Shashua and K. J. Hanna. The tensor brightness constraints: Direct estimation of motion revisited. Technical report, Technion, Haifa, Israel, November 1995.

[65] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *Proceedings of the International Conference on Computer Vision*, pages 920–925, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[66] C. Slama, editor. *Manual of Photogrammetry.* American Society of Photogrammetry, 4 edition, 1980.

[67] S. Soatto and P. Perona. Motion from fixation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 817–824, San Fransico, CA, June 1996.

[68] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.

[69] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proceedings of the ARPA Image Understanding Workshop*, 1990.

[70] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proceedings of the ARPA Image Understanding Workshop*, 1990.

[71] G. P. Stein. Internal camera calibration using rotation and geometric shapes. Technical Report AITR-1426, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1993.

[72] G. P. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *Proceedings of the International Conference on Computer Vision*, pages 230–236. IEEE Computer Society Press, June 1995.

[73] G. P. Stein. Lens distortion calibration using point correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.

[74] G. P. Stein and A. Shashua. Model based brightness constraints: On direct estimation of structure and motion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.

[75] R. Szeliski and S. Kang. Recovering 3d shape and motion from image streams using non-linear least squares. Technical Report CRL 93/3, DEC CRL, March 1992.

[76] R. Szeliski and S. B. Kang. Direct methods for visual scene reconstruction. In *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, June 1995.

[77] T. Y. Tian, C. Tomasi, and D. J. Heeger. Comparison of approaches to egomotion computation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 315–320, San Francisco, California, June 1996.

[78] A. P. Tirumalai, B. G. Schunk, and R. C. Jain. Dynamic stereo with self calibration. In *Proceedings of the International Conference on Computer Vision*, pages 466–470, Osaka, Japan, December 1990.

[79] C. Tomasi and T. Kanade. Shape and motion from image streams – a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[80] B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995. IEEE Computer Society Press, IEEE Computer Society Press.

[81] S. Ullman and R. Basri. Recognition by linear combination of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13:992—1006, 1991. Also in M.I.T AI Memo 1152, 1989.

[82] A. M. Waxman and J. H. Duncan. Binocular image flows: steps towards stereo-motion fusion. *Pattern Analysis and Machine Intelligence*, 8(6):715–729, November 1986.

[83] J. Weng, N. Ahuja, and T. Huang. Two view matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 64–73, Tampa, Florida, December 1989.

[84] J. Weng, T. Huang, and N. Ahuja. Motion and structure from line correspondences: Closed form solution, uniqueness and optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3), 1992.

[85] J. Weng, P. Cohen, and N. Rebibo. Motion and structure estimation from stereo image sequnces. *IEEE Transactions on Robotics and Automation*, 8(3):362–382, June 1992.

[86] J. Weng et al. Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence*, 14:965–980, 1992.

[87] G. Wolberg. *Digital Image Warping.* IEEE Computer Society Press, Los Alamitos, California, 1990.

[88] S. Xuening and M. E. Spetsakis. A comparison of weighted ls methods with ls methods in 3d motion estimation from stereo image sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 200–205, New York, NY, June 1993.

[89] J. W. Young. *Projective Geometry.* The Mathematical Society of America, 1930.

[90] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 407–411, Vienna, Austria, August a996.

[91] Z. Zhang and O. D. Faugeras. Three dimensional motion computation and segementation in a long sequnce of stereo frames. *International Journal of Computer Vision*, 7(3):211–241, 1992.