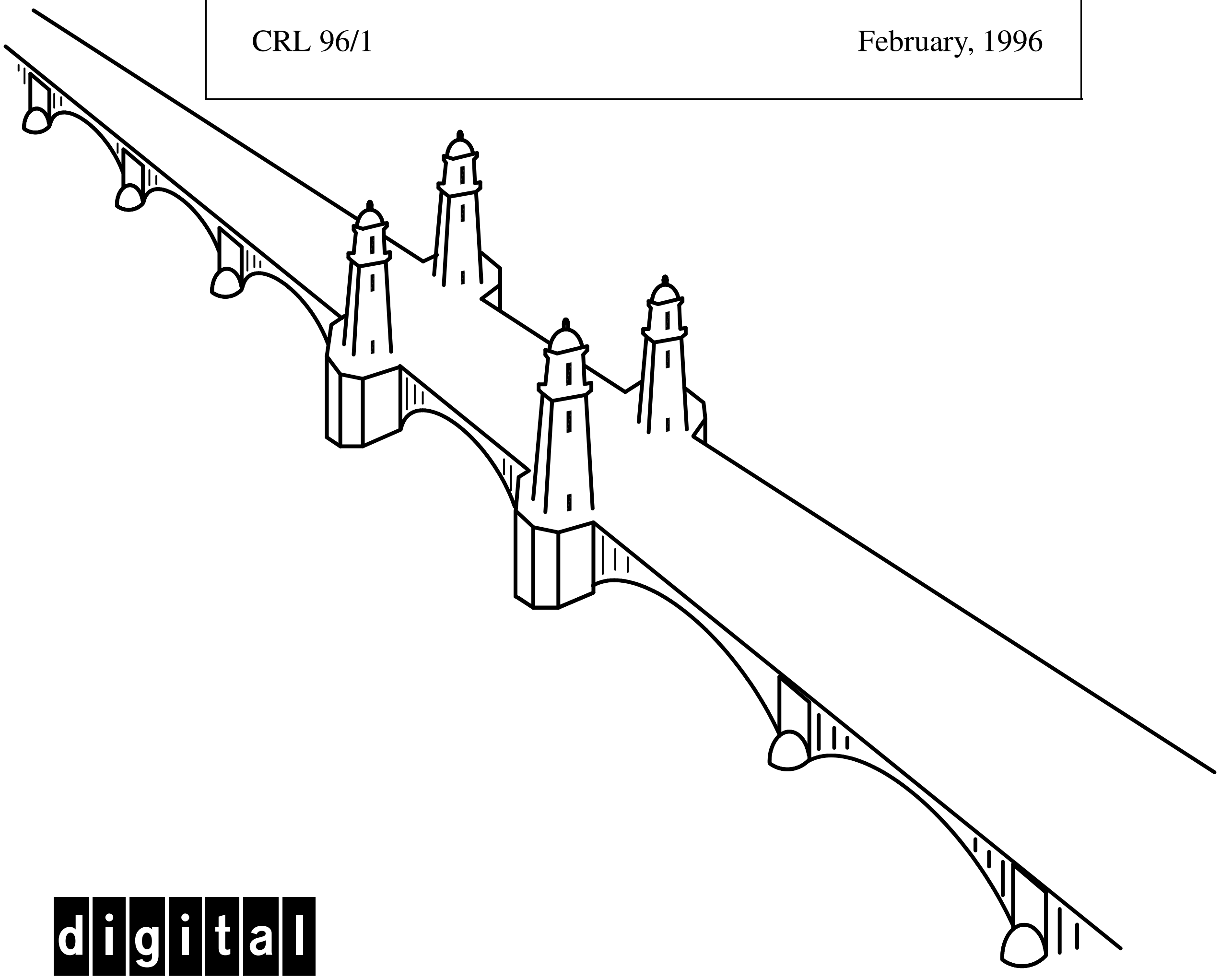
Shape Ambiguities in Structure from Motion

Richard Szeliski and Sing Bing Kang

Digital Equipment Corporation
Cambridge Research Lab

CRL 96/1 February, 1996



CAMBRIDGE RESEARCH LABORATORY
Technical Report Series

Cambridge Research Laboratory
One Kendall Square
Cambridge, Massachusetts 02139

Shape Ambiguities in Structure from Motion

Richard Szeliski[1] and Sing Bing Kang

Digital Equipment Corporation

Cambridge Research Lab

CRL 96/1 February, 1996

Abstract

This technical report examines the fundamental ambiguities and uncertainties inherent in recovering structure from motion. By examining the eigenvectors associated with null or small eigenvalues of the Hessian matrix, we can quantify the exact nature of these ambiguities and predict how they affect the accuracy of the reconstructed shape. Our results for orthographic cameras show that the bas-relief ambiguity is significant even with many images, unless a large amount of rotation is present. Similar results for perspective cameras suggest that three or more frames and a large amount of rotation are required for metrically accurate reconstruction.

Keywords: Structure from motion, ambiguities, uncertainty analysis

[1] Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399

Contents

List of Figures

List of Tables

1 Introduction

Structure from motion is one of the classic problems in computer vision and has received a great deal of attention over the last decade. It has wide-ranging applications, including robot vehicle guidance and obstacle avoidance, and the reconstruction of 3-D models from imagery. Unfortunately, the quality of results available using this approach is still often very disappointing. More precisely, while the qualitative estimates of structure and motion look reasonable, the actual quantitative (*metric*) estimates can be significantly distorted.

Much progress has been made recently in identifying the sources of errors and instabilities in the structure from motion process. It is now widely understood that the arbitrary algebraic manipulation of the imaging equations to derive closed-form solutions (e.g., [LH81]) can lead to algorithms that are numerically ill-conditioned or unstable in the presence of measurement errors. To overcome this, statistically optimal algorithms for estimating structure and motion have been developed [SA89; WAH89; Hor90; TK92b; SK94]. It is also understood that using more feature points and images results in better estimates, and that certain configurations of points (at least in the two frame case) are pathological and cannot be reconstructed.

An example of an algorithm which generates very good results is the factorization approach of Tomasi and Kanade [TK92b]. This algorithm assumes orthography and is implemented using an object-centered representation and singular value decomposition. It uses many points and frames, and for most sequences, a large amount of object rotation (usually $360°$). However, when only a small range of viewpoints is present (e.g., the "House" sequence in [TK92b], Figure 7), the reconstruction no longer appears metric (the house walls are not perpendicular).

In this technical report, we demonstrate that it is precisely this last factor, i.e., the overall rotation of the object, or equivalently, the variation in viewpoints, which critically determines the quality of the reconstruction. The ambiguity in object shape due to small viewpoint variation often looks like it might be a *projective* deformation of the Euclidean shape, which is interesting—several researchers have argued recently in favor of trying to recover only this projective structure [Fau92; HGC92; MQVB92; Sha93]. In fact, we show that the major ambiguity in the reconstruction is a simple depth scale uncertainty, i.e., the classic *bas-relief* ambiguity which exists for two-frame structure from motion under orthographic projection [LH86].[1]

[1] The bas-relief ambiguity is even more pronounced in shape from shading, and forms the basis of classical friezes and bas-relief sculptures.

To derive our results, we use eigenvalue analysis of the covariance matrix for the structure and motion estimates. This assumes that we can compute a near optimal solution, and that the error in the solution is due to linear perturbations arising from small amounts of image noise (feature point mislocalization). This kind of analysis has not previously been applied to structure from motion, and yet it is a very powerful way to predict the ultimate performance of structure from motion algorithms.

Our results are significant for two reasons. First, we show how to theoretically derive the expected ambiguity in a reconstruction, and also derive some intuitive guidelines for selecting imaging situations which can be expected to produce reasonable results. Second, since the primary ambiguities are very well characterized by a small number of modes, this information can be used to construct better on-line (recursive) estimation algorithms.

Our technical report is structured as follows. After reviewing previous work, we present our formulation of the structure from motion problem and develop our technique for analyzing ambiguities using eigenvector analysis of the information (Hessian) matrix. We then present the results of our analysis for a series of camera models: 1-D and 2-D orthographic cameras, and 1-D and 2-D perspective cameras. We conclude with a discussion of the main sources of errors and ambiguities, and directions for possible future work.

2 Previous work

Structure from motion has been extensively studied in computer vision. Early papers on this subject [LH81; TH84] develop algorithms to compute the structure and motion from a small set of points matched in two frames using an *essential parameter* approach. The performance of this approach can be significantly improved using non-linear least squares (*optimal estimation*) techniques [WAH89; WAH93; SA89; Hor90; SA91].

Recent research focuses on extraction of shape and motion from longer image sequences [KTJ89; DA90; CWC90; TK92b; CT92]. Cui, Weng, and Cohen [CWC90] use an optimal estimation technique (non-linear least squares) between each pair of frames, and an extended Kalman filter to accumulate information over time (see also [THO93; SPFP93]). Azarbayejani *et al.* [AHP93] also use a Kalman filter-based approach to recover rigid (object-centered) depth and motion directly from the sequence of image measurements. Tomasi and Kanade [TK92b] use a factorization method which extracts shape and motion from an image stream without computing camera-centered depth. Their approach formulates the shape from motion problem in object-centered coordinates, assumes orthography, and processes all of the frames simultaneously. Chen and Tsuji [CT92] relax the assumption of orthography by analyzing the image sequence through its temporal and spatial subparts. Taylor and Kriegman [TKA91; TK92a] formulate the shape from motion task as a non-linear least squares problem in which the Euclidean distance between the estimated and actual positions of the points in the image sequence is minimized using the Levenberg-Marquardt algorithm. Szeliski and Kang [SK94] extend this approach approaches to general 3-D structure and also to projective structure and motion recovery.

Another line of research has addressed recovering affine [KvD91; SZB93] or projective [Fau92; HGC92; HG93; MVQ93] structure estimates. Most of these techniques rely on identifying and tracking a small number of feature points in the image sequence, using these points to form a basis set for the geometric description, and also only use 2 frames to recover the geometry. However, Mohr *et al.* [MVQ93] and Szeliski and Kang [SK94] use as many points and frames as possible to recover the geometry and motion, thus producing more reliable estimates.

The nature of structure and motion errors, which is the main focus of this technical report, has also previously been studied. Weng *et al.* perform some of the earliest and most detailed error analyses of the two-frame essential parameter approach [WAH89; WAH93]. Adiv [Adi89] and Young and Chellappa [YC92] analyze continuous-time (optical flow) based algorithms using the concept of the Cramer-Rao lower bound. Oliensis and Thomas [OT91; THO93] show how modeling the motion error can significantly improve the performance of recursive algorithms.

In this technical report, we extend these previous results using an eigenvalue analysis of the covariance matrix. This analysis can pinpoint the exact nature of structure from motion ambiguities and the largest sources of reconstruction error. We also focus on multi-frame optimal structure from motion algorithms, which have not been studied in great detail.

3 Problem formulation and uncertainty analysis

Structure from motion can be formulated as the recovery of a set of 3-D structure parameters $\mathbf{p}_i$ and time-varying motion parameters $\mathbf{m}_j$ from a set of observed image features $\mathbf{u}_{ij}$. In this section, we present the forward equations, i.e., the rigid body and perspective transformations which map 3-D points into 2-D image points. We also show how the Jacobians of the forward equation can be used to estimate the inverse covariance matrix for the parameters being recovered, how this can be used to quantify expected reconstruction errors, and how our results relate to classical structure from motion ambiguities.

3.1 Problem formulation

The equation which projects the ith 3-D point $\mathbf{p}_i$ into the jth frame at location $\mathbf{u}_{ij}$ is

$$\mathbf{u}_{ij} = \mathcal{P}\left(T(\mathbf{p}_i, \mathbf{m}_j)\right). \tag{1}$$

The perspective projection $\mathcal{P}$ (defined below) is applied to a rigid transformation

$$T(\mathbf{p}_i, \mathbf{m}_j) = \mathbf{R}_j \mathbf{p}_i + \mathbf{t}_j, \tag{2}$$

where $\mathbf{R}_j$ is a rotation matrix and $\mathbf{t}_j$ is a translation applied after the rotation. A variety of alternative representations are possible for the rotation matrix [Aya91]. In this technical report, we primarily use a quaternion $\mathbf{q} = [w, (q_0, q_1, q_2)]$ representation, with a corresponding rotation matrix

$$\mathbf{R}(\mathbf{q}) = \begin{pmatrix} 1 - 2q_1^2 - 2q_2^2 & 2q_0q_1 + 2wq_2 & 2q_0q_2 - 2wq_1 \\ 2q_0q_1 - 2wq_2 & 1 - 2q_0^2 - 2q_2^2 & 2q_1q_2 + 2wq_0 \\ 2q_0q_2 + 2wq_1 & 2q_1q_2 - 2wq_0 & 1 - 2q_0^2 - 2q_1^2 \end{pmatrix} \tag{3}$$

since this representation has no singularities. The rotation parameters q_0, q_1, q_2 also have a natural interpretation (for small values) as the half-angles of rotation around the x, y, and z axes. For our one-dimensional examples, we use the rotation angle around the vertical axis.

The standard perspective projection equation used in computer vision is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathcal{P}_1 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \equiv \begin{pmatrix} f\frac{x}{z} \\ f\frac{y}{z} \end{pmatrix}, \tag{4}$$

where f is a product of the focal length of the camera and the pixel scale factor (assuming that pixels are square). An alternative object-centered formulation, which we introduced in [SK94] is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathcal{P}_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \equiv \begin{pmatrix} s\frac{x}{1+\eta z} \\ s\frac{y}{1+\eta z} \end{pmatrix}. \tag{5}$$

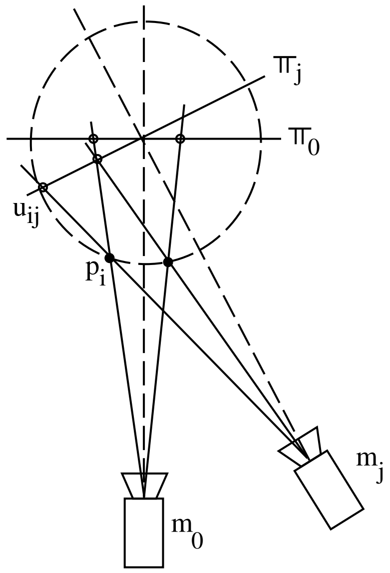


Figure 1: Sample configuration of cameras ($\mathbf{m}_j$), 3-D points ($\mathbf{p}_i$), image planes(Π_j), and screen locations ($\mathbf{u}_{ij}$)

Here, we assume that the (x, y, z) coordinates before projection are with respect to a reference frame Π_j that has been displaced away from the camera by a distance t_z along the optical axis, with $s = f/t_z$ and $\eta = 1/t_z$ (Figure 1). The projection parameter s can be interpreted as a *scale factor* and η as a *perspective distortion factor*. Our alternative perspective formulation allows us to model both orthographic and perspective cameras using the same model.

A variety of techniques (reviewed in Section 2) can be used to estimate the unknowns $\{\mathbf{p}_i, \mathbf{m}_j\}$ from the given image measurements $\{\mathbf{u}_{ij}\}$. In our previous work [SK94], we used the iterative Levenberg-Marquardt algorithm, since it provides a statistically optimal solution [WAH89; SA89; TK92a; SK94]. The Levenberg-Marquardt method is a standard non-linear least squares technique [PFTV92] which directly minimizes a merit or objective function

$$\mathcal{C}(\mathbf{a}) = \sum_i \sum_j c_{ij} \left| \tilde{\mathbf{u}}_{ij} - \mathbf{f}_{ij}(\mathbf{a}) \right|^2, \tag{6}$$

where $\tilde{\mathbf{u}}_{ij}$ is the observed image measurement, $\mathbf{f}_{ij}(\mathbf{a}) = \mathbf{u}(\mathbf{p}_i, \mathbf{m}_j)$ is given in (1), and the vector $\mathbf{a}$ contains all of the unknown structure and motion parameters, including the 3-D points $\mathbf{p}_i$, the motion parameters $\mathbf{m}_j$, and any additional unknown calibration parameters. The weight c_{ij} in (6) describes the confidence in measurement $\mathbf{u}_{ij}$, and is normally set to the inverse variance σ_{ij}^{-2} (it can be set to zero for missing measurements).

3.2 Uncertainty analysis

Regardless of the solution technique, the uncertainty in the recovered parameters—assuming that image measurements are corrupted by small Gaussian noise errors—can be determined by computing the inverse covariance or *information* matrix $\mathbf{A}$ [Sor80]. This matrix is formed by computing outer products of the *Jacobians* of the measurement equations

$$\mathbf{A} = \sum_i \sum_j c_{ij} \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{a}} \frac{\partial \mathbf{f}_{ij}}{\partial \mathbf{a}^T}. \tag{7}$$

For notational succinctness, we use the symbol

$$\mathbf{H}_{ij} = \begin{bmatrix} \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{p}_i} \\ \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{m}_j} \end{bmatrix}$$

to denote the non-zero portion of the full Jacobian $\frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{a}}$.

If we list the structure parameters $\{\mathbf{p}_i\}$ first, followed by the motion parameters $\{\mathbf{m}_j\}$, the $\mathbf{A}$ matrix has the structure

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{p}} & \mathbf{A}_{\mathbf{pm}} \\ \hline \mathbf{A}_{\mathbf{pm}}^T & \mathbf{A}_{\mathbf{m}} \end{array} \right]. \tag{8}$$

The matrices $\mathbf{A}_{\mathbf{p}}$ and $\mathbf{A}_{\mathbf{m}}$ are block diagonal, with diagonal entries

$$\mathbf{A}_{\mathbf{p}_i} = \sum_j \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{p}_i} \frac{\partial \mathbf{f}_{ij}}{\partial \mathbf{p}_i^T} \quad \text{and} \quad \mathbf{A}_{\mathbf{m}_j} = \sum_i \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{m}_j} \frac{\partial \mathbf{f}_{ij}}{\partial \mathbf{m}_j^T}, \tag{9}$$

respectively (assuming $c_{ij} = 1$), while $\mathbf{A}_{\mathbf{pm}}$ is dense, with entries

$$\mathbf{A}_{\mathbf{p}_i \mathbf{m}_j} = \frac{\partial \mathbf{f}_{ij}^T}{\partial \mathbf{p}_i} \frac{\partial \mathbf{f}_{ij}}{\partial \mathbf{m}_j^T}. \tag{10}$$

The information matrix has previously been used in the context of structure from motion to determine *Cramer-Rao lower bounds* on the parameter uncertainties by taking the inverse of the diagonal entries [Adi89; YC92]. The Cramer-Rao bounds, however, can be arbitrarily weak, especially when $\mathbf{A}$ is singular or near-singular. In this technical report, we use eigenvector analysis of $\mathbf{A}$ to find the dominant directions in the uncertainty (covariance) matrix and their magnitudes, which gives us more insight into the exact nature of structure from motion ambiguities.

3.3 Estimating reconstruction errors

An important benefit of uncertainty analysis is that we can easily quantify the expected amount of reconstruction (and motion) error for an optimal structure from motion algorithm. For example, the expected sum of squared error in reconstructed 3-D point positions is

$$S_{pos}^2 \equiv \left\langle \sum_i \|\tilde{\mathbf{p}}_i - \mathbf{p}_i^*\|^2 \right\rangle, \tag{11}$$

where $\tilde{\mathbf{p}}_i$ are the estimated (recovered) positions and $\mathbf{p}_i^*$ the true positions. The positional uncertainty matrix $\mathbf{C}_\mathbf{p}$ can be computed by inverting $\mathbf{A}$ and looking at its upper left block (the block corresponding to the $\mathbf{p}_i$ variables).[2] If we perform an eigenvalue analysis of $\mathbf{C}_\mathbf{p}$, we obtain

$$\mathbf{C}_\mathbf{p} = \mathbf{E}_\mathbf{p}^T \mathbf{\Lambda}_\mathbf{p} \mathbf{E}_\mathbf{p}, \tag{12}$$

where $\mathbf{E}_\mathbf{p}$ is the matrix of eigenvectors, and $\mathbf{\Lambda}_\mathbf{p}$ is the diagonal matrix containing the eigenvalues of $\mathbf{C}_\mathbf{p}$. Since S_{pos}^2 is a Euclidean norm, its value is unaffected by orthogonal coordinate transformations such as $\mathbf{E}_\mathbf{p}$. The value of S_{pos}^2 can thus be computed as either the trace of $\mathbf{C}_\mathbf{p}$ or the trace of $\mathbf{\Lambda}_\mathbf{p}$, i.e., the sum of the eigenvalues of $\mathbf{C}_\mathbf{p}$.

In practice, we do not need to compute $\mathbf{C}_\mathbf{p}$. Instead, the sum of squared reconstruction and motion error,

$$S_{all}^2 \equiv \left\langle \sum_i \|\tilde{\mathbf{p}}_i - \mathbf{p}_i^*\|^2 + \sum_j \|\tilde{\mathbf{m}}_j - \mathbf{m}_j^*\|^2 \right\rangle, \tag{13}$$

can be computed directly summing the *inverse* eigenvalues of the information matrix $\mathbf{A}$. By choosing an appropriate scaling for the parameters being estimated (say scaling positions to be in the range $[-100 \ldots 100]$ and rotations in the range $[-\pi \ldots \pi]$), we can make the mean of S_{all} be close to the mean of S_{pos}. Note that for general 3-D camera motion, positional errors in the motion estimates will be on the same scale as 3-D reconstruction errors, and may sometimes dominate (if the absolute distance of the camera is ill determined).

What is the advantage of this approach, if computing eigenvalues is just as expensive as inverting matrices? First, we can compute the first few eigenvalues more cheaply (and in less space) than the matrix inverse, and these tend to dominate the overall reconstruction error. Second, it justifies the approach in the technical report, which is to look at the minimum eigenvalue as the prime indicator of reconstruction error. We can therefore study how much certain ambiguities (such as the

[2]Note that this is *not* the same as simply inverting $\mathbf{A}_\mathbf{p}$.

bas-relief ambiguity) contribute to the overall reconstruction error. We can also obtain much tighter lower bounds on the reconstruction error than would be possible by using the Cramer-Rao bounds.

3.4 Ambiguities in structure from motion

Because structure from motion attempts to recover both the structure of the world and the camera motion without any external (prior) knowledge, it is subject to certain ambiguities. The most fundamental (but most innocuous) of these is the coordinate frame (also known as pose, or Euclidean) ambiguity, i.e., we can move the origin of the coordinate system to an arbitrary place and pose and still obtain an equally valid solution.

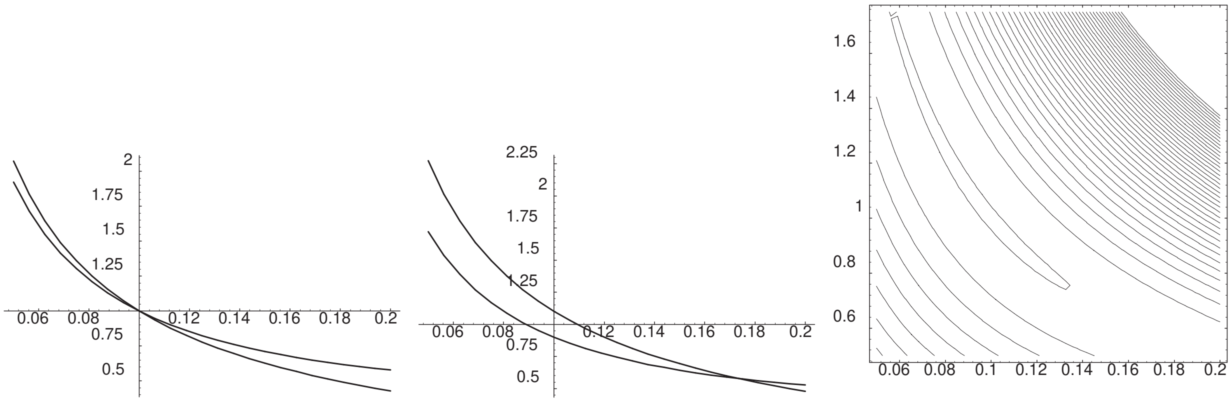
The next most common ambiguity is the scale ambiguity (for a perspective camera) or the depth ambiguity (for an orthographic camera). This ambiguity can be removed with a small amount of additional knowledge, e.g., the absolute distance between camera positions.

A third ambiguity, and the one we focus on in this technical report, is the *bas-relief ambiguity*. In its pure form, this ambiguity occurs for a two frame problem with an orthographic camera, and is a confusion between the *relative depth* of the object and the amount of object rotation. In this technical report, we focus on the *weak* form of this ambiguity, i.e., the very large *bas-relief uncertainty* which occurs with imperfect measurements even when we use more than two frames and/or perspective cameras. A central result of this technical report is that the bas-relief ambiguity captures the largest uncertainties arising in structure from motion. However, when examined in detail, it appears that a larger class of deformations (i.e., projective) more fully characterizes the errors which occur in structure from motion.

To characterize these ambiguities, we will use eigenvector analysis of the information matrix, as explained in Section 3.2. Absolute ambiguities will show up as zero eigenvalues (unless we add additional constraints or knowledge to remove them), whereas weak ambiguities will show up as small eigenvalues.

4 A two parameter example

To develop an intuitive understanding of the basic bas-relief ambiguity, we start with a simple two-parameter example. Assume that we have an orthographic scanline camera which measures the x component of 2-D points (x, z). Furthermore, assume that we already know the shape up to a scale



(a) Noise-free (b) Noise = 0.01 (c) Contour plot

Figure 2: Constraint lines and energy surface for simple two-parameter example. The x-axis is the angle $\Delta\theta$ and the y-axis is the scale factor a.

factor in depth,

$$\mathbf{p}_i = (x_i, az_i)$$

and that the rotation angles are uniform,

$$\theta_j = j\Delta\theta.$$

The projection equation is then

$$u_{ij} = c_j x_i - s_j a z_i \tag{14}$$

with $c_j = \cos\theta_j$ and $s_j = \sin\theta_j$.

What happens when we try to estimate the scale factor a and the angle $\Delta\theta$ from a set of noisy measurements $\{u_{ij}\}$? First, let's examine the very simplest case, which is a single point, say at $(x, z) = (1, 1)$. Each new image gives us a constraint of the form

$$c_j - as_j = c_j^* - a^* s_j^* + n_j \tag{15}$$

where c_j^*, s_j^*, and a^* are the true values and n_j is random noise. Figure 2a shows the two constraint lines for $j = \pm 1$ assuming the noise-free case (with $a = 1$ and $\Delta\theta = 0.1$ rad). Figure 2b shows the constraint lines for $n_{-1} = n_1 = 0.01$. As can be seen, the estimate for $(\Delta\theta, a)$ is very sensitive to noise. This can also be seen in the contour plot of the energy surface (Figure 2c) which can be computed by summing the constraints in (15).

To characterize the shape of the error surface near its minimum, we compute the information matrix $\mathbf{A}$. The Jacobian for $(a, \Delta\theta)$ is straightforward,

$$\mathbf{H}_{ij} = \begin{bmatrix} \frac{\partial u_{ij}}{\partial a} \\ \frac{\partial u_{ij}}{\partial \Delta\theta} \end{bmatrix} = \begin{bmatrix} -s_j z_i \\ -j(ac_j z_i + s_j x_i) \end{bmatrix} \approx -j \begin{bmatrix} \Delta\theta z_i \\ a z_i + j\Delta\theta x_i \end{bmatrix} \tag{16}$$

if we assume small rotation angles, $|\theta_j| \ll 1$, so that $s_j \approx j\Delta\theta$ and $c_j \approx 1$. The inverse covariance (information) matrix is then

$$\mathbf{A} \approx J_2 Z \begin{bmatrix} \Delta\theta^2 & a\Delta\theta \\ a\Delta\theta & a^2 + \Delta\theta^2 \frac{J_4 X}{J_2 Z} \end{bmatrix} \tag{17}$$

where $J_2 = \sum_j j^2$, $J_4 = \sum_j j^4$, $X = \sum_i x_i^2$, and $Z = \sum_i z_i^2$ (assuming that $\sum_j j = 0$). Assuming that $\Delta\theta^2 \ll a^2$, we can compute (Appendix A) the approximate eigenvalues of $\mathbf{A}$ as

$$\lambda_{\min} \approx \Delta\theta^4 J_4 X / a^2 \quad \text{and} \quad \lambda_{\max} \approx J_2 Z a^2. \tag{18}$$

The eigenvalues of the information matrix describe an "elliptic" approximation to the error surface (and hence posterior probability distribution), which matches the true "banana shaped" surface near the optimal solution but not far away from it. To determine if the additional nonlinearities in the reconstruction process result lower or higher overall uncertainties than those predicted by the information matrix, we would have to resort to numerical simulations. In practice, we expect these secondary effect to be much smaller than the large variations in eigenvalues which explain most of the uncertainties (ambiguities) associated with structure from motion.

5 Orthography: single scanline

Let us now turn to a true structure from motion problem where both the structure and motion are unknown. For simplicity, we analyze the orthographic scanline camera first, where the unknowns are the 2-D point positions $\mathbf{p}_i = (x_i, z_i)$ and the rotation angles θ_j.[3] The imaging equations are

$$u_{ij} = c_j x_i - s_j z_i \tag{19}$$

with $c_j = \cos\theta_j$ and $s_j = \sin\theta_j$.

[3] We do not estimate the horizontal translation since it can be determined from the motion of the centroid of the image points [TK92b].

The Jacobian for the 1-D orthographic camera is

$$\mathbf{H}_{ij} = \left[\begin{array}{cc|c} \frac{\partial u_{ij}}{\partial x_i} & \frac{\partial u_{ij}}{\partial z_i} & \frac{\partial u_{ij}}{\partial \theta_j} \end{array} \right]^T = \left[\begin{array}{cc|c} c_j & -s_j & -(c_j z_i + s_j x_i) \end{array} \right]^T, \tag{20}$$

and the entries in the information matrix are

$$\mathbf{A}_{\mathbf{p}_i} = \begin{bmatrix} \sum_j c_j^2 & -\sum_j c_j s_j \\ -\sum_j c_j s_j & \sum_j s_j^2 \end{bmatrix} = \begin{bmatrix} C & -D \\ -D & S \end{bmatrix}, \tag{21}$$

$$\mathbf{A}_{\mathbf{p}_i \mathbf{m}_j} = \begin{bmatrix} -c_j^2 z_i - c_j s_j x_i \\ c_j s_j z_i + s_j^2 x_i \end{bmatrix}, \tag{22}$$

$$\mathbf{A}_{\mathbf{m}_j} = \left[\sum_i (c_j z_i + s_j x_i)^2 \right] = \left[c_j^2 Z + 2 c_j s_j W + s_j^2 X \right], \tag{23}$$

with $C = \sum_j c_j^2$, $D = \sum_j c_j s_j$, $S = \sum_j s_j^2$, $Z = \sum_i z_i^2$, $W = \sum_i z_i x_i$, and $X = \sum_i z_i^2$.

Before analyzing the complete information matrix, let us look at the two subblocks $\mathbf{A}_\mathbf{p}$ and $\mathbf{A}_\mathbf{m}$. If we know the motion, the structure uncertainty is determined by $\mathbf{A}_{\mathbf{p}_i}$ and is simply the triangulation error, i.e., $\sigma_x^2 \propto C^{-1}$ and $\sigma_z^2 \propto S^{-1}$ (note that for small rotations, σ_x^2 is generally much smaller than σ_z^2). If we know the structure, the motion accuracy is determined by $\mathbf{A}_{\mathbf{m}_j}$ and is inversely proportional to the variance in depth along the viewing direction (s_j, c_j).

What about ambiguities in the solution? Under orthography, the traditional scale ambiguity does not exist. However, translations along the optical axis cannot be estimated, and an overall pose (coordinate frame) ambiguity still exists. Unless we add some additional constraints, we can always rotate the coordinate system by a $\Delta\theta$ and add the same amount to the $\{\theta_j\}$. This manifests itself as the null (zero eigenvalue) eigenvector

$$\mathbf{e}_0 = \left[\begin{array}{cccccc|ccc} z_0 & -x_0 & \cdots & z_N & -x_N & 1 & \cdots & 1 \end{array} \right]^T.$$

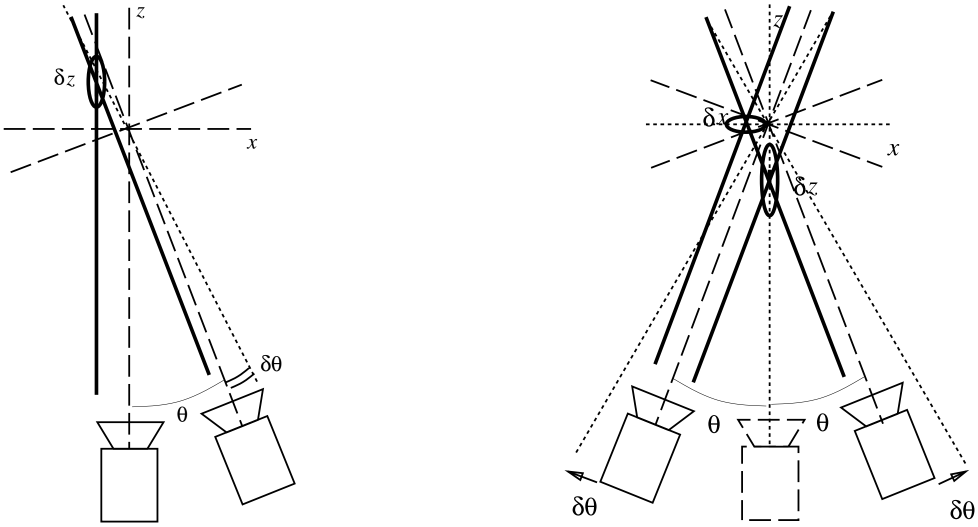
5.1 Two frames: the bas-relief ambiguity

Let us say we only have two frames, and we have fixed $\theta_0 = 0, c_0 = 1, s_0 = 0, \theta_1 = \theta, c_1 = c, s_1 = s$ (Figure 3). Then

$$\mathbf{A}_{\mathbf{p}_i} = \begin{bmatrix} 1 + c^2 & -cs \\ -cs & s^2 \end{bmatrix} \tag{24}$$

$$\mathbf{A}_{\mathbf{p}_i \mathbf{m}} = \begin{bmatrix} -c^2 z_i - c s x_i \\ c s z_i + s^2 x_i \end{bmatrix} \tag{25}$$

$$\mathbf{A}_{\mathbf{m}} = \left[c^2 Z + 2csW + s^2 X \right]. \tag{26}$$



(a) stationary first frame (b) antisymmetric cameras

Figure 3: Orthographic projection, two frames.
The solid lines indicate the viewing rays, while the thin lines indicate the optical axes and image planes. The diagonal dashed lines are the displaced viewing rays, while the ellipses indicate the positional uncertainty in the reconstruction due to uncertainty in motion (indicated as $\delta\theta$).

The bas-relief ambiguity manifests itself as a null eigenvector

$$\mathbf{e}_0 = \left[\begin{array}{ccccc|c} 0 & cz_0 + sx_0 & 0 & \cdots & cz_N + sx_N & -s \end{array} \right]^T,$$

as can be verified by inspection. This is as we expected, i.e., the primary uncertainty in the structure is entirely in the depth (z) direction, and is a scale uncertainty (proportional to z). Note however that this uncertainty is proportional to $cz + sx$ rather than z, as can be seen by inspecting Figure 3a.

An alternative parameterization of the two-frame problem is to set $\theta_0 = -\theta_1$ (Figure 3b), in which case we have

$$\mathbf{A}_{\mathbf{p}_i} = \begin{bmatrix} 2c^2 & 0 \\ 0 & 2s^2 \end{bmatrix} \tag{27}$$

$$\mathbf{A}_{\mathbf{p}_i\mathbf{m}} = \begin{bmatrix} -2csx_i \\ 2csz_i \end{bmatrix} \tag{28}$$

$$\mathbf{A}_{\mathbf{m}} = \begin{bmatrix} 2c^2 Z + 2s^2 X \end{bmatrix}. \tag{29}$$

In this case, the null eigenvector is

$$\mathbf{e}_0 = \left[\begin{array}{ccccc|c} s^2 x_0 & -c^2 z_i & \cdots & s^2 x_N & -c^2 z_N & cs \end{array}\right]^T. \tag{30}$$

This is also very illuminating. It shows that the primary effect of the bas-relief ambiguity is a "squashing" of the z values for a small increase in motion, with a much smaller "bulging" in the x values (at least for small inter-frame rotations).[4] This squashing and bulging is an affine deformation of the true structure.

5.2 More than two frames, equi-angular motion constraint

To simplify the analysis, we assume for the moment that we know we have an equi-angular image sequence, i.e., that the rotation angles are given by $\theta_j = j\Delta\theta$, $j \in \{-J, \ldots, J\}$, $J = \frac{F+1}{2}$, where F is the total number of frames (imagine Figure 3b with more cameras). In this case, we have

$$\begin{aligned}
\mathbf{H}_{ij}^T &= \left[\begin{array}{cc|c} c_j & -s_j & -j(c_j z_i + s_j x_i) \end{array}\right] && (31)\\
\mathbf{A}_{\mathbf{p}_i} &= \begin{bmatrix} \sum_j c_j^2 & 0 \\ 0 & \sum_j s_j^2 \end{bmatrix} = \begin{bmatrix} C & 0 \\ 0 & S \end{bmatrix}, && (32)\\
\mathbf{A}_{\mathbf{p}_i\mathbf{m}} &= \begin{bmatrix} -\sum_j j c_j s_j x_i \\ \sum_j j c_j s_j z_i \end{bmatrix} = \begin{bmatrix} -E x_i \\ E z_i \end{bmatrix}, && (33)\\
\mathbf{A}_{\mathbf{m}} &= \left[\begin{array}{c} \sum_j j^2 c_j^2 Z + \sum_j j^2 s_j^2 X \end{array}\right] = \left[\begin{array}{c} C'Z + S'X \end{array}\right], && (34)
\end{aligned}$$

with $E = \sum_j j c_j s_j$, $C' = \sum_j j^2 c_j^2$, $S' = \sum_j j^2 s_j^2$, and C, D, S, Z, W, X defined as in (22–23). In this case, the smallest eigenvalue eigenvector has the form

$$\mathbf{e}_0 = \left[\begin{array}{ccccc|c} \alpha x_0 & -\beta z_0 & \cdots & \alpha x_N & -\beta z_N & 1 \end{array}\right]^T. \tag{35}$$

This will be an eigenvector if we can satisfy the matrix equation $\mathbf{A}\mathbf{e} = \lambda\mathbf{e}$, i.e.,

$$\left[\begin{array}{c|c} \mathbf{A}_\mathbf{p} & \mathbf{A}_{\mathbf{pm}} \\ \hline \mathbf{A}_{\mathbf{pm}}^T & \mathbf{A}_\mathbf{m} \end{array}\right] \left[\begin{array}{c} \alpha x_0 \\ -\beta z_0 \\ \vdots \\ -\beta z_N \\ \hline 1 \end{array}\right] = \lambda \left[\begin{array}{c} \alpha x_0 \\ -\beta z_0 \\ \vdots \\ -\beta z_N \\ \hline 1 \end{array}\right],$$

[4] Note that compared to the previous example where frame 0 was fixed, the total interframe rotation is now 2θ.

which reduces to the following three equations:

$$\begin{aligned} \alpha C - E &= \alpha\lambda \\ \beta S - E &= \beta\lambda \\ (S' - \alpha E)X + (C' - \beta E)Z &= \lambda. \end{aligned}$$

Substituting $\alpha = \frac{E}{C-\lambda}$ and $\beta = \frac{E}{S-\lambda}$ into the third equation, we obtain a cubic in λ,

$$(S-\lambda)(S'(C-\lambda) - E^2)X + (C-\lambda)(C'(S-\lambda) - E^2)Z - (S-\lambda)(C-\lambda)\lambda = 0, \quad (36)$$

which can be solved analytically using a package such as *Mathematica*$_{®}$ [Wol91].

Assuming that the smallest eigenvalue is very small, we can use the approximation $\alpha \approx \frac{E}{C}$ to obtain a quadratic in λ,

$$(S-\lambda)(S'C - E^2)X + C(C'(S-\lambda) - E^2)Z - (S-\lambda)C\lambda = 0. \quad (37)$$

Furthermore, using the small angle approximations, $C \approx \sum_j 1 \equiv J_0$, $S \approx \Delta\theta^2 J_2$, $E \approx \Delta\theta J_2$, $C' \approx J_2$, and $S' \approx \Delta\theta^2 J_4$, we obtain after some manipulation (Appendix A)

$$\lambda_{\min} \approx \frac{\Delta\theta^4 X J_2(J_0 J_4 - J_2^2)}{J_0 J_2 Z + \Delta\theta^2[X(J_0 J_4 - J_2^2) + J_0 J_2]}. \quad (38)$$

Notice that the minimum eigenvalue is related to the fourth power of $\Delta\theta$, i.e., doubling the inter-frame rotation reduces the RMS (root mean square) error by a factor of 4 (assuming that $Z \gg \Delta\theta^2$). Increasing the extent of the x_i compared to the z_i directly increases the minimum eigenvalue, i.e., it decreases the structure uncertainty. This result is somewhat surprising, and suggests that flatter objects can be reconstructed better.

We can numerically compute the values of λ for a range of J and $\Delta\theta$ values (Figure 4). For example, with $J = 1, \Delta\theta = 0.1$ rad $\approx 6°$, and $X = Z = 1$, we have $\lambda = \{0.0000664436, 1.98064, 3.0193\}$. For the smallest eigenvalue, $\lambda = 0.0000664436$, we have a corresponding $\alpha = 0.0666676$ and $\beta = 10.0001$.

Once the smallest eigenvalue and eigenvector have been computed, we can easily determine some additional eigenvectors. Any vector which consists purely of x_i or z_i values which is also orthogonal to $\mathbf{A}_{\mathrm{pm}}$ is an eigenvector, e.g.,

$$\mathbf{e} = \left[\begin{array}{cccccc|c} x_1 & 0 & -x_0 & 0 & \cdots & 0 & 0 \end{array}\right].$$

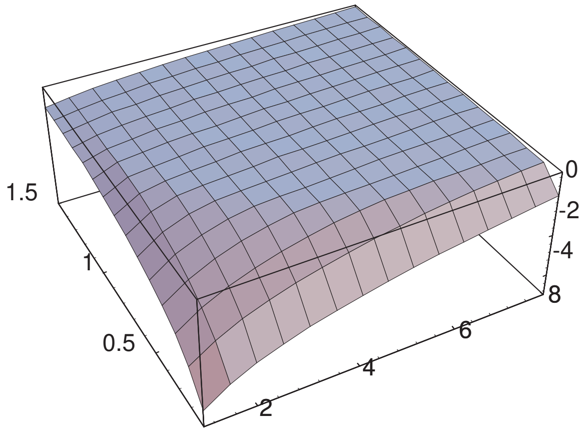


Figure 4: Plot of $\log_{10} \lambda_{\min}$ as a function of $J \in [1, 8]$ and $\Delta\theta \in [0.1, 1.5]$.

$\lambda_{\min}$	$F = 2$	$F = 3$	$F = 4$	$F = 5$	$F = 6$	$F = 7$	$F = 8$
$\theta_{\text{tot}} = 11.5°$	0.000000	0.000067	0.000079	0.000088	0.000096	0.000104	0.000112
$\theta_{\text{tot}} = 22.9°$	0.000000	0.001087	0.001283	0.001418	0.001547	0.001677	0.001810
$\theta_{\text{tot}} = 34.4°$	0.000000	0.005618	0.006597	0.007277	0.007931	0.008594	0.009269
$\theta_{\text{tot}} = 45°$	0.000000	0.016854	0.019688	0.021673	0.023596	0.025552	0.027547
$\theta_{\text{tot}} = 60°$	0.000000	0.054679	0.063442	0.069678	0.075782	0.082017	0.088389
$\theta_{\text{tot}} = 90°$	0.000000	0.272977	0.316453	0.348500	0.380039	0.412200	0.444997

Table 1: Minimum eigenvalues for 1-D orthographic known equi-angular motion

The eigenvalues corresponding to the pure x eigenvectors are C, while the z eigenvalues are S. In other words, once the global bas-relief uncertainty has been accounted for (squashing in z and smaller bulging in x), the variance in x position estimates is proportional to C^{-1} and in z positions is proportional to S^{-1}, i.e., exactly the expected triangulation error for known camera positions.

For the above example with $J = 1$ (3 frames), $\Delta\theta = 0.1$ rad $\approx 6°$, and $X = Z = 1$, the values for C and S are 2.98 and 0.0199, respectively. From this, we see that the correlated depth uncertainty due to the motion uncertainty is a factor of $0.0199/0.00006644 = 300$ times greater than the individual depth uncertainties. A full table of $\lambda_{\min}$ as a function of $F = 2J + 1$ (the number of frames) and $\theta_{\text{tot}} = (F - 1)\Delta\theta$ (the total rotation angle) is shown in Table 1.

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}}=11.5°$	0.000000	0.000067	0.000079	0.000087	0.000095	0.000103	0.000111
$\theta_{\text{tot}}=22.9°$	0.000000	0.001080	0.001263	0.001391	0.001513	0.001636	0.001762
$\theta_{\text{tot}}=34.4°$	0.000000	0.005537	0.006377	0.006971	0.007549	0.008136	0.008731
$\theta_{\text{tot}}=45°$	0.000000	0.016450	0.018596	0.020163	0.021721	0.023311	0.024924
$\theta_{\text{tot}}=60°$	0.000000	0.052521	0.057558	0.061612	0.065825	0.070179	0.074598
$\theta_{\text{tot}}=90°$	0.000000	0.254859	0.261589	0.273769	0.288362	0.303857	0.319541

Table 2: Minimum eigenvalues for 1-D orthographic equi-angular motion with no constraint

5.3 More than two frames, without motion constraint

If we take the same data set as above, but remove the additional knowledge of equi-angular steps, we end up solving for each motion (angle) estimate separately. The equations for $\mathbf{A}_{\mathbf{p}_i}$, $\mathbf{A}_{\mathbf{p}_i\mathbf{m}_j}$, and $\mathbf{A}_{\mathbf{m}_j}$ are given in (22–23), with $D=0$. Let us guess that the bas-relief ambiguity eigenvector has the form

$$\mathbf{e}_0 = \left[\begin{array}{cccc|ccc} \alpha x_0 & -\beta z_0 & \cdots & -\beta z_N & -J & \cdots & J \end{array}\right]^T. \tag{39}$$

The requirements for this to be an eigenvector are similar to those we derived before,

$$\begin{aligned} \alpha C - E &= \alpha\lambda & (40)\\ \beta S - E &= \beta\lambda & (41)\\ c_j^2(jZ-\alpha W) + c_j s_j(2jW - \alpha X - \beta Z) + s_j^2(jX-\beta W) &= \lambda j. & (42) \end{aligned}$$

In this case, we do not have a closed form solution, since we have $2J+3$ equations in 3 unknowns. However, if we assume a small angle approximation and $W=0$ (i.e., that the 3-D point cloud is rotationally symmetric with respect to the middle frame), then the $2J+1$ equations of the form (42) are equivalent and we get the same eigenvectors as with the known equiangular motion constraint.

This behavior can be verified numerically (Table 2), where the results are quite similar to those shown in Table 1. To obtain these results, we computed the $\mathbf{A}$ matrix explicitly using a set of 9 points sampled on the unit square, i.e., $\{(x,z), x, z \in \{-1,0,1\}\}$, and then computed the eigenvalues. Note, however, that for an example where $W \neq 0$, i.e., by adding one additional point at $(2,2)$ to the previous example, we get an eigenvector which is not of the form hypothesized in (39). It is, however, an affine transform of the (x_i, z_i) coordinates.

S_{all}	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}} = 11.5°$	∞	123.61	113.80	108.18	103.50	99.34	95.60
$\theta_{\text{tot}} = 22.9°$	∞	31.81	29.46	28.02	26.80	25.71	24.73
$\theta_{\text{tot}} = 34.4°$	∞	14.88	13.88	13.21	12.62	12.09	11.62
$\theta_{\text{tot}} = 45°$	∞	9.32	8.74	8.30	7.92	7.58	7.27
$\theta_{\text{tot}} = 60°$	∞	6.01	5.65	5.35	5.08	4.85	4.64
$\theta_{\text{tot}} = 90°$	∞	3.94	3.62	3.37	3.16	2.99	2.84

Table 3: S_{all} estimates for 1-D orthographic equi-angular motion with no constraint, $X = Z = 100$, $\sigma = 1$.

We can also estimate the expected reconstruction error S_{all} by summing the inverse eigenvalues. Using the same parameters as for Table 2, but with $X = Z = 100$ to make structure errors dominate, we obtain the results in Table 3. This table shows how the bas-relief ambiguity dominates the reconstruction error. At small viewing angles, doubling the angle results in a fourfold reduction in the sum of squared error S_{all}. Adding more frames is much less effective than increasing the effective baseline of the system.

6 Orthography: full 3-D reconstruction

The situation with a regular orthographic camera (2-D retina, 3-D world) is quite similar to the scanline camera. In this case, we use unit quaternions to represent the rotation matrices,

$$u_{ij} = r_{00j}x_i + r_{01j}y_i + r_{02j}z_i \tag{43}$$

$$v_{ij} = r_{10j}x_i + r_{11j}y_i + r_{12j}z_i, \tag{44}$$

where the entries in the rotation matrix r_{kl} are given in (3).

To obtain a qualitative feel for the bas-relief ambiguity, let us examine the known equiangular motion case with a small amount of rotation around a fixed axis (say in the y-z plane),

$$\mathbf{q}_j \approx [1, (0, jq_1, jq_2)], \tag{45}$$

where q_1 is the incremental rotation around the y axis, and q_2 is the rotation about the z (optical) axis. As before, we ignore camera translations under orthography, since these can be recovered from the motion of the point centroid.

The Jacobian matrix is now

$$\mathbf{H}_{ij}^T = \left[\begin{array}{ccc|cc} \frac{\partial u_{ij}}{\partial x_i} & \frac{\partial u_{ij}}{\partial y_i} & \frac{\partial u_{ij}}{\partial z_i} & \frac{\partial u_{ij}}{\partial q_1} & \frac{\partial u_{ij}}{\partial q_2} \\ \frac{\partial v_{ij}}{\partial x_i} & \frac{\partial v_{ij}}{\partial y_i} & \frac{\partial v_{ij}}{\partial z_i} & \frac{\partial v_{ij}}{\partial q_1} & \frac{\partial v_{ij}}{\partial q_2} \end{array}\right] \tag{46}$$

$$\approx \left[\begin{array}{ccc|cc} 1 & 2jq_2 & -2jq_1 & -4j^2q_1x_i - 2jz_i & -4j^2q_2x_i + 2jy_i \\ -2jq_2 & 1 & 2j^2q_1q_2 & 2j^2q_2z_i & -2jx_i - 4j^2q_2y_i + 2j^2q_1z_i \end{array}\right]. \tag{47}$$

The entries in the information matrix are

$$\mathbf{A}_{\mathbf{p}_i} \approx \begin{bmatrix} J_0 & 0 & 0 \\ 0 & J_0 & -2J_2q_1q_2 \\ 0 & -2J_2q_1q_2 & 4J_2q_1^2 \end{bmatrix}, \tag{48}$$

$$\mathbf{A}_{\mathbf{p}_i\mathbf{m}_j} \approx \begin{bmatrix} -4J_2q_1x_i & -2J_2q_2x_i \\ -2J_2q_2z_i & 2J_2q_1z_i \\ 4J_2q_1z_i & -4J_2q_1y_i \end{bmatrix}, \tag{49}$$

$$\mathbf{A}_{\mathbf{m}_j} = \begin{bmatrix} 4J_2\sum_i z_i^2 & -4J_2\sum_i y_iz_i \\ -4J_2\sum_i y_iz_i & 4J_2\sum_i(x_i^2+y_i^2) \end{bmatrix} = 4J_2\begin{bmatrix} Z & W' \\ W' & X+Y \end{bmatrix}, \tag{50}$$

with $Y = \sum_i y_i^2$, $W' = \sum_i y_iz_i$, and other terms as defined before.

These equations are similar to those for the orthographic scanline camera (22–23), with $C \approx J_0$, $S \approx J_2q_1^2$, $E \approx J_2q_1$, and $C' \approx J_2$. In the absence of positional uncertainty, the accuracies of the q_1 and q_2 estimates ($\mathbf{A}_{\mathbf{m}_j}^{-1}$) are inversely proportional to Z and $X+Y$, respectively, as is to be expected. Similarly, with known motion, the triangulation error ($\mathbf{A}_{\mathbf{p}_i}^{-1}$) are inversely proportional to the number of frames J_0 for x and y, and proportional to the squared rotation angle $J_2q_1^2$ for z. Notice that a non-zero tilt of the rotation axis ($q_2 \neq 0$) confounds some of the y and z positional uncertainties.

Instead of trying to find an analytical solution to the eigenvalue problem, we present a brief example showing the dependence of $\lambda_{\min}$ on q_1 and q_2 (Table 4). For this example, we used a 15-point data set consisting of the 8 corners of a unit cube, the 6 cube faces, and the origin. The eigenvalues for the no-tilt case ($q_2 = 0$) are almost identical to the results of 1-D analysis (Table 2). The eigenvalues for the tilted case ($q_2/q_1 = \tan 30°$) are similar in shape but show the effect of the overall decrease in q_1 values. By examining the eigenvectors (not shown here), we observe that for *both* cases, the minimum eigenvector has no y components.

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}}=11.5°$	0.000000	0.000067	0.000079	0.000088	0.000096	0.000104	0.000112
$\theta_{\text{tot}}=22.9°$	0.000000	0.001092	0.001267	0.001410	0.001531	0.001665	0.001792
$\theta_{\text{tot}}=34.4°$	0.000000	0.005677	0.006405	0.007189	0.007747	0.008452	0.009065
$\theta_{\text{tot}}=45°$	0.000000	0.017153	0.018653	0.021226	0.022638	0.024838	0.026500
$\theta_{\text{tot}}=60°$	0.000000	0.056333	0.056757	0.067148	0.069948	0.078044	0.082245
$\theta_{\text{tot}}=90°$	0.000000	0.287619	0.203405	0.320241	0.268727	0.343410	0.318149

Table 4: Minimum eigenvalues for 2-D orthographic equi-angular motion with no constraint, rotation around y axis ($q_1=\sin\frac{\Theta_j}{2}$, $q_2=0$).

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}}=11.5°$	0.000000	0.000046	0.000055	0.000061	0.000066	0.000072	0.000077
$\theta_{\text{tot}}=22.9°$	0.000000	0.000750	0.000873	0.000971	0.001056	0.001148	0.001236
$\theta_{\text{tot}}=34.4°$	0.000000	0.003857	0.004392	0.004919	0.005316	0.005795	0.006224
$\theta_{\text{tot}}=45°$	0.000000	0.011507	0.012731	0.014410	0.015451	0.016919	0.018101
$\theta_{\text{tot}}=60°$	0.000000	0.036927	0.038640	0.044940	0.047420	0.052530	0.055737
$\theta_{\text{tot}}=90°$	0.000000	0.170400	0.150632	0.200555	0.196403	0.233575	0.235277

Table 5: Minimum eigenvalues for 2-D orthographic equi-angular motion with no constraint, rotation around y axis tilted $30°$ ($q_1=\cos 30°\sin\frac{\Theta_j}{2}$, $q_2=\sin 30°\sin\frac{\Theta_j}{2}$).

7 Perspective: single scanline

Before analyzing the perspective camera in 3-D, let us briefly look at a perspective scanline (1-D) camera. We can use this model to develop some intuitions, but unfortunately we cannot use it to predict the performance of the full two-frame algorithm, since even under perspective, the scanline camera has a bas-relief ambiguity. This can be shown by a simple parameter counting argument: there are $2N$ unknowns for the 2-D coordinates $\{(x_i, z_i)\}$ and 1 (or more) unknowns for the motion, but only $2N$ measurements. In other words, we can place the cameras arbitrarily, and the intersections of the optical rays will determine the location of the 2-D points. This argument obviously does not carry over to 3-D, but it is suggestive of why two-frame structure from motion may be poorly

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}} = 11.5°$	0.000000	0.000080	0.000094	0.000104	0.000114	0.000124	0.000133
$\theta_{\text{tot}} = 22.9°$	0.000000	0.001274	0.001498	0.001655	0.001807	0.001960	0.002116
$\theta_{\text{tot}} = 34.4°$	0.000000	0.006421	0.007489	0.008257	0.009006	0.009768	0.010544
$\theta_{\text{tot}} = 45°$	0.000000	0.018670	0.021580	0.023745	0.025885	0.028072	0.030305
$\theta_{\text{tot}} = 60°$	0.000000	0.057351	0.065494	0.071906	0.078373	0.085026	0.091834
$\theta_{\text{tot}} = 90°$	0.000000	0.255136	0.288877	0.317718	0.347360	0.377933	0.409211

Table 6: Minimum eigenvalues for 1-D perspective projection, equi-angular rotation, $\eta = 0.2$.

conditioned.

The projection equation for a scanline camera, using the new projection model introduced in (5), is

$$u_{ij} = \frac{c_j x_i - s_j z_i + t_{xj}}{1 + \eta(s_j x_i + c_j z_i + t_{zj})} = \frac{N_{ij}}{D_{ij}}. \tag{51}$$

The Jacobian matrix is

$$\begin{aligned} \mathbf{H}_{ij}^T &= \left[\begin{array}{cc|ccc} \frac{\partial u_{ij}}{\partial x_i} & \frac{\partial u_{ij}}{\partial z_i} & \frac{\partial u_{ij}}{\partial \theta_j} & \frac{\partial u_{ij}}{\partial t_{xj}} & \frac{\partial u_{ij}}{\partial t_{zj}} \end{array} \right] \\ &= \frac{1}{D_{ij}} \left[\begin{array}{cc|ccc} c_j - \eta s_j \tilde{u}_{ij} & -(s_j + \eta c_j \tilde{u}_{ij}) & -(s_j x_i + c_j z_i + \eta(c_j x_i - s_j z_i)\tilde{u}_{ij}) & 1 & -\eta \tilde{u}_{ij} \end{array} \right] \end{aligned} \tag{52}$$

where $\tilde{u}_{ij}$ is the predicted value of u_{ij} computed by (51). In addition to the usual coordinate frame ambiguity, we have a scale ambiguity, i.e., the (x_i, z_i) and t_{xj} can be multiplied by a factor a, and t_{zj} can be set to $at_{zj} + (a-1)/\eta$, without affecting the solution. As mentioned above, a full bas-relief ambiguity also exists for 2 frames.

Rather than continuing our analysis with the construction of the Hessian matrix $\mathbf{A}$, let us just look briefly at the form of $\mathbf{H}_{ij}$. In addition to the terms already present under orthography (20), we have the extra terms involving η, as well as the partial with respect to t_{zj}. These additional terms are what will, in full 3-D, enable the two-frame problem to be solved by removing the bas-relief ambiguity.

To see the effects of using a perspective camera instead of an orthographic camera, we show in Table 6 the minimum eigenvalue as a function of total viewing angle and number of frames. Compared to Table 2, we see that there is a small, but not dramatic, improvement in the size of $\lambda_{\min}$.

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$	$F=7$	$F=8$
$\theta_{\text{tot}}=11.5°$	0.000175	0.000214	0.000239	0.000269	0.000299	0.000331	0.000364
$\theta_{\text{tot}}=22.9°$	0.000690	0.001289	0.001462	0.001633	0.001803	0.001981	0.002158
$\theta_{\text{tot}}=34.4°$	0.001512	0.004372	0.004972	0.005491	0.006009	0.006510	0.007024
$\theta_{\text{tot}}=45°$	0.002512	0.009905	0.011282	0.012020	0.012959	0.013460	0.014070
$\theta_{\text{tot}}=60°$	0.004234	0.020246	0.022853	0.021650	0.021870	0.020495	0.019727
$\theta_{\text{tot}}=90°$	0.008381	0.032074	0.032623	0.027976	0.026149	0.023367	0.021596

Table 7: Minimum eigenvalues for 3-D perspective projection, equi-angular rotation around y axis, $\eta = 0.1$.

8 Perspective in 3-D

Let us finally analyze the most interesting case, that of a perspective camera operating in a 3-D environment. Here, we know that the two-frame problem has a solution, although our results on the simpler camera models suggest that the reconstructions may be particularly sensitive to noise.

The forward imaging equations are given in (1–3) and (5). We will not bother deriving the Jacobian and Hessian matrices here, as they are complex and not particularly informative. Instead, we present some numerical results on $\lambda_{\min}$ and RMS_{pos} and discuss their significance. (Note that $RMS_{pos} = S_{pos}/\sqrt{n}$, where n is the number of points.) These results were obtained using the *Mathematica*® package [Wol91], by analytically differentiating the forward projection equations, and then substituting in the known structure and motion parameters. Numerical eigenvalue analysis was then used to obtain our results. For these examples, we used the 15 points sampled on the unit cube described in Section 6.

We present results for two special cases: pure object-centered rotation (which in camera-centered coordinates is actually both rotation and translation), and pure forward translation. Ignoring the effects of motion across the retina, these two cases capture the basic motion cues available to structure from motion.

8.1 Pure object-centered rotations

To compute the minimum eigenvalue results, we used the same approach as for the orthographic 3-D camera (Section 6). The computed eigenvalues are shown in Table 7. Compared to the orthographic

$\lambda_{\min}$	$\eta = 0.025$ $\phi = 3°$	$\eta = 0.05$ $\phi = 6°$	$\eta = 0.1$ $\phi = 12°$	$\eta = 0.2$ $\phi = 28°$	$\eta = 0.3$ $\phi = 46°$	$\eta = 0.4$ $\phi = 67°$	$\eta = 0.5$ $\phi = 90°$
$\theta_{\text{tot}} = 11.5°$	0.000010	0.000041	0.000175	0.000899	0.002648	0.003899	0.002947
$\theta_{\text{tot}} = 22.9°$	0.000040	0.000161	0.000690	0.003505	0.010216	0.015504	0.011702
$\theta_{\text{tot}} = 34.4°$	0.000087	0.000354	0.001512	0.007578	0.021758	0.034461	0.025941
$\theta_{\text{tot}} = 45°$	0.000145	0.000591	0.002512	0.012402	0.035035	0.057861	0.043377
$\theta_{\text{tot}} = 60°$	0.000247	0.001002	0.004234	0.020494	0.056570	0.097234	0.072229
$\theta_{\text{tot}} = 90°$	0.000492	0.001993	0.008381	0.039718	0.105540	0.144799	0.111384

Table 8: Minimum eigenvalues for 3-D perspective projection, equi-angular rotation around y axis, two frames ($F = 2$), varying η. ϕ is the camera's field of view.

case (Table 4), we see some striking differences. First, the two-frame problem is now soluble (up to a scale ambiguity, of course). Second, for small viewing angles, there is marked improvement even for multiple frames. Third, the results for large viewing angles with small η's are significantly inferior to the orthographic results. This appears to be caused by ambiguities in camera motion along the optical axis (t_z), which are neglected in the orthographic case.

This table only shows us the results for a particular value of η. The dependence of $\lambda_{\min}$ on η is presented in Tables 8 and 9 for the two and three frame problems. In these tables, the fields of view equivalent to each η were computed from the horizontal spread of the data points on the unit cube and the distance of the cube from the camera η^{-1} using the formula $\phi = 2\tan^{-1}\frac{\eta}{1-\eta}$. As can be seen for the two-frame case, doubling the amount of perspective distortion η results in a fourfold increase in $\lambda_{\min}$ (and hence a halving of the RMS error). For the three-frame case, the results are less sensitive to η.

What does a typical minimum eigenvector look like? Figure 5 shows the eigenvector corresponding to the three-frame problem with $\eta = 0.1$ and $\theta_{tot} = 11.5°$. As we can see, the majority of the ambiguity is indeed a depth scaling. Notice, however, that the eigenvector is not a pure affine transform of the 3-D coordinates, since the tips of the vectors in a given row do not form a straight line (this has also been verified numerically). Our conjecture is that the minimum eigenvector may be a *projective* transformation of the 3-D points, i.e., that the main ambiguity is projective, but we have not yet found a proof for this conjecture.

How do the 3-D (position) errors RMS_{pos} vary with the number of frames and viewing angle?

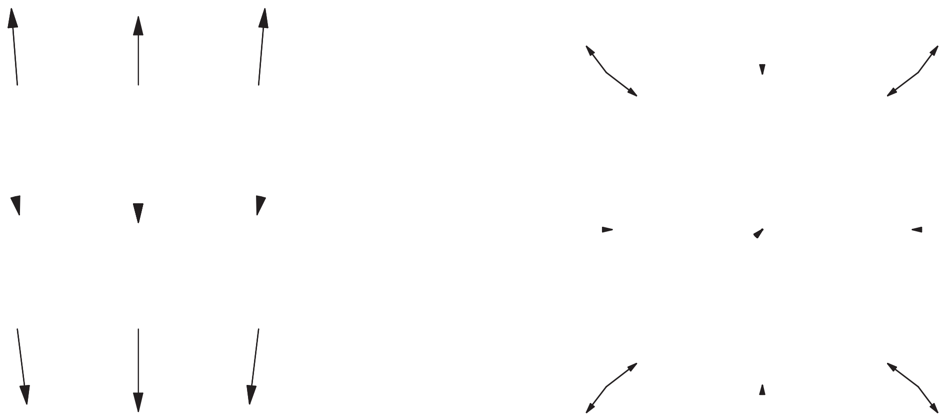
$\lambda_{\min}$	$\eta = 0.025$ $\phi = 3°$	$\eta = 0.05$ $\phi = 6°$	$\eta = 0.1$ $\phi = 12°$	$\eta = 0.2$ $\phi = 28°$	$\eta = 0.3$ $\phi = 46°$	$\eta = 0.4$ $\phi = 67°$	$\eta = 0.5$ $\phi = 90°$
$\theta_{\text{tot}} = 11.5°$	0.000043	0.000075	0.000214	0.000956	0.002736	0.003908	0.002958
$\theta_{\text{tot}} = 22.9°$	0.000502	0.000688	0.001289	0.004384	0.011565	0.015655	0.011874
$\theta_{\text{tot}} = 34.4°$	0.001399	0.002606	0.004372	0.011820	0.028129	0.035277	0.026838
$\theta_{\text{tot}} = 45°$	0.001825	0.005074	0.009905	0.023998	0.052204	0.060488	0.046154
$\theta_{\text{tot}} = 60°$	0.002009	0.007177	0.020246	0.051574	0.103096	0.107273	0.082110
$\theta_{\text{tot}} = 90°$	0.002098	0.008302	0.032074	0.121672	0.205362	0.215310	0.181425

Table 9: Minimum eigenvalues for 3-D perspective projection, equi-angular rotation around y axis, three frames ($F = 3$), varying η. ϕ is the camera's field of view.

RMS_{pos}	$F = 2$	$F = 3$	$F = 4$	$F = 5$	$F = 6$	$F = 7$	$F = 8$
$\theta_{\text{tot}} = 11.5°$	20.78	19.08	18.04	17.02	16.12	15.32	14.62
$\theta_{\text{tot}} = 22.9°$	10.51	8.09	7.61	7.19	6.83	6.51	6.23
$\theta_{\text{tot}} = 34.4°$	7.13	4.64	4.38	4.13	3.94	3.75	3.60
$\theta_{\text{tot}} = 45°$	5.57	3.24	3.06	2.89	2.76	2.63	2.53
$\theta_{\text{tot}} = 60°$	4.35	2.32	2.19	2.07	1.98	1.89	1.82
$\theta_{\text{tot}} = 90°$	3.25	1.70	1.59	1.49	1.43	1.37	1.32

Table 10: RMS_{pos} for 3-D perspective projection, equi-angular rotation around y axis, $\eta = 0.1$.

By computing the full covariance matrix (inverting $\mathbf{A}$) and taking the trace of the positional covariance matrix $\mathbf{C}_{\mathrm{p}}$ (as described in Section 3.2), we obtain the results shown in Table 10. These numbers indicate the relative errors in reconstruction for a unit retina and unit noise. For example, if the retina is actually 200 pixels wide ($s = 100$ in (5)) and the positional error in the tracked points is $\sigma = 0.1$, then the 3-D reconstruction errors would be 1000 times smaller than the values given in Table 10. We see that this error decreases linearly with total viewing angle (for small viewing angles), and varies only slightly with the total number of frames. This is similar to the results obtained when computing $\lambda_{\min}$ in Table 4 (remember that RMS error should be proportional to the square root of the inverse eigenvalues).



(a) (b)

Figure 5: Minimum eigenvector for a three-frame perspective reconstruction problem: (a) top-down view (x-z), (b) frontal view (x-y). While the main ambiguity is a z scaling, the vector is not exactly an affine transform of the 3-D points on the unit cube.

8.2 Looming

The motion of a camera forward in a 3-D world creates a different kind of parallax, which can also be exploited to compute structure from motion. To compute the ambiguities in this kind of motion, we used the same approach as before, except with no rotation and pure forward motion ($t_z \neq 0$).

Using our usual 15-point data set results in some unexpected behavior: four of the eigenvalues are zero. This is because the z coordinates of the three points on the optical axis cannot be recovered as they lie on the focus of expansion. This is a severe limitation of recovering structure from looming: points near the focus of expansion are recovered with extremely poor accuracy. For the experiments in this section, we use a 12-point data set instead, i.e., the 15-point set with the three points $(x, y) = (0, 0)$ removed.

Table 11 shows $\lambda_{\min}$ as a function of the number of frames F and the total extent of forward motion t_z (the object being viewed is a unit cube with coordinates $[-1, 1]^3$). These results are for a camera with $\eta = 0.3$, i.e., a camera placed about 3.3 units away from the cube origin. As we can see, the two-frame results are almost as good at the three frame results with the same extent of motion. The value of $\lambda_{\min}$ appears to depend quadratically on the total extent of motion. Overall, however, these results are much worse than those available with object-centered rotation.

Table 12 shows $\lambda_{\min}$ as a function of η, i.e., the distance of the central frame to the object. It

$\lambda_{\min}$	$F=2$	$F=3$	$F=4$	$F=5$	$F=6$
$t_z=0.1$	0.000007	0.000007	0.000007	0.000008	0.000009
$t_z=0.2$	0.000027	0.000027	0.000030	0.000033	0.000037
$t_z=0.3$	0.000060	0.000060	0.000067	0.000075	0.000084
$t_z=0.4$	0.000107	0.000107	0.000119	0.000134	0.000150
$t_z=0.5$	0.000168	0.000168	0.000187	0.000210	0.000235

Table 11: Minimum eigenvalues for 3-D perspective projection, pure forward translation, $\eta = 0.3$.

$\lambda_{\min}$	$\eta=0.1$	$\eta=0.2$	$\eta=0.3$	$\eta=0.4$	$\eta=0.5$
	$\phi=12°$	$\phi=28°$	$\phi=46°$	$\phi=67°$	$\phi=90°$
$t_z=0.1$	0.000000	0.000002	0.000007	0.000013	0.000020
$t_z=0.2$	0.000001	0.000009	0.000027	0.000051	0.000078
$t_z=0.3$	0.000002	0.000020	0.000060	0.000115	0.000176
$t_z=0.4$	0.000004	0.000036	0.000107	0.000205	0.000314
$t_z=0.5$	0.000006	0.000057	0.000168	0.000320	0.000490

Table 12: Minimum eigenvalues for 3-D perspective projection, pure forward translation, $F = 2$, varying η.

appears that $\lambda_{\min}$ depends cubically on η, at least for small t_zs. To obtain reasonable estimates, therefore, it is necessary to both use a wide field of view and a large amount of motion relative to the scene depth.

Figure 6. shows the structural part of the minimum eigenvectors in particular for $\eta = 0.3$, $J = 1$ ($F = 3$), and $\Delta t_z = 0.2$. eigenvector whose 3-D structure is shown in Figure 6. By inspection of the complete eigenvector (not shown here), we can see that the ambiguity is between the amount of x and y yaw and x and y translation, i.e., it is a classic bas-relief ambiguity.

9 Experimental results

To verify if the positional errors predicted by our analysis coincide with the errors observed in practice, we ran our iterative non-linear least squares algorithm on a 24-point sample data set [SK94].

(a) (b)

Figure 6: Minimum eigenvector for a three-frame perspective reconstruction problem with pure z translation: (a) top-down view (x-z), (b) frontal view (x-y). The main ambiguity is a rocking confusion between sideways camera translation and rotation, which affects the points furthest back.

The 24 points were four points at $(\pm 0.4685, \pm 0.4685)$ on the six faces of a unit $[-1, +1]^3$ cube. The points were projected onto a 200 pixel wide retina ($s = 100$ in (5)) and 2-D noise with $\sigma = 0.1$ was added to each projected point.[5] The algorithm was then initialized with the correct 3-D structure and run to completion.

The 3-D positional errors are shown in Tables 13 and 14. Three kinds of error are shown: the Euclidean error, after registering the recovered and true 3-D data sets under the best possible similarity transform (rigid + scaling); the affine error (computing the best affine transform); and the projective error (computing the best 4×4 homography). These errors were scaled by a factor of 1000 to make them "dimensionless" (i.e., unit retina, unit image noise). The RMS error predicted by our uncertainty analysis (the trace of the positional covariance matrix) is also shown.

From these results, we can see that the uncertainty analysis predicts the general variation of reconstruction error with viewing angle, perspective distortion, and number of frames. Unfortunately, there remains a small but fairly consistent discrepancy between our predicted figures and the measured errors, which we have not been able to track down. We also see that the affine error is about 2 to 3 times lower than the Euclidean error (actually, this factor increases with decreasing viewing

[5]The results scale linearly with σ up to about $\sigma = 1$, after which they increase sub-linearly (i.e., they less than double when σ is doubled).

RMS_{pos}	$\eta = 0.1$				$\eta = 0.2$			
$F = 2$	predicted	Euclidean	affine	projective	predicted	Euclidean	affine	projective
$\theta_{\text{tot}} = 8°$	35.02	58.98	20.41	19.02	19.68	34.43	21.68	20.48
$\theta_{\text{tot}} = 16°$	18.21	35.70	10.27	9.39	9.93	16.63	10.39	9.75
$\theta_{\text{tot}} = 32°$	9.28	15.70	5.10	4.78	5.13	9.15	5.34	4.98
$\theta_{\text{tot}} = 60°$	5.24	8.47	2.89	2.72	3.02	4.69	3.01	2.82
$\theta_{\text{tot}} = 90°$	3.85	5.36	2.03	1.93	2.37	3.32	2.15	2.04

Table 13: RMS errors (predicted and observed) for 3-D perspective projection, equi-angular rotation around y axis, two frames, 24 point data set.

RMS_{pos}	$\eta = 0.1$				$\eta = 0.2$			
$F = 3$	predicted	Euclidean	affine	projective	predicted	Euclidean	affine	projective
$\theta_{\text{tot}} = 6°$	41.94	61.17	20.21	18.76	25.79	40.45	22.21	20.26
$\theta_{\text{tot}} = 12°$	19.83	26.90	10.31	9.69	12.55	18.12	10.39	9.71
$\theta_{\text{tot}} = 24°$	7.42	11.34	4.99	4.76	5.75	8.08	5.23	4.91
$\theta_{\text{tot}} = 48°$	2.76	3.70	2.50	2.43	2.59	3.63	2.72	2.61
$\theta_{\text{tot}} = 90°$	1.59	1.96	1.54	1.50	1.57	1.90	1.59	1.53

Table 14: RMS errors (predicted and observed) for 3-D perspective projection, equi-angular rotation around y axis, three frames, 24 point data set.

angle, as predicted by our analysis). The projective error is *not* significantly lower than the affine error, which further supports our hypothesis that most of the error is in the bas-relief ambiguity.[6]

10 Discussion

The results presented in this technical report suggest that in many situations where structure from motion might be applied, the solutions are extremely sensitive to noise. In fact, despite dozens of algorithms having been developed, very few results of convincing quality are available. Those

[6]It is not suprrising that the projective error is always smaller than the affine error, as there are 3 more degrees of freedom (15 vs. 12) in the projective fit used before the error computation.

cases where metrically accurate results have been demonstrated almost always use a large amount of rotation [TK92b].

This raises the obvious question: are any of the many structure from motion algorithms developed in the computer vision community of practical significance? Or, when we wish to perform metrically accurate reconstructions from images, should be adopt the photogrammetrists' approach of using control points at known locations? This essentially reduces structure from motion to camera pose estimation (and possibly calibration) followed by stereo reconstruction.

The situation is perhaps not that bad. For large object rotations, we can indeed recover accurate reconstructions. Furthermore, for scene reconstruction, using cameras with large fields of view, several camera mounted in different directions, or even panoramic images, should remove most of the ambiguities. In any case, it would appear prudent to carefully analyze the expected ambiguities and uncertainties in any structure from motion problem (or any other image-based estimation task) before actually putting a method into practice.

The general approach developed in this technical report, i.e., eigenvalue analysis of the Hessian (information) matrix appears to explain most of the known ambiguities in structure from motion. However, there are certain ambiguities (e.g., depth reversals under orthography, or multiplicities of solutions with few points and frames) which will not be detected by this analysis because they correspond to multiple local minima of the cost function in the parameter space. Furthermore, analysis of the information matrix can only predict the sensitivity of the results to *small* amounts of image noise. Further study using empirical methods is required to determine the limitations of our approach.

Using the minimum eigenvalue to predict the overall reconstruction error may fail when the dominant ambiguities are in the motion parameters (e.g., what appears to be happening under perspective for large motions). Computing the RMS_{pos} error directly from the covariance matrix $\mathbf{A}^{-1}$ is more useful in these cases.

10.1 Future work

In future work, we plan to compare results available with object-centered and camera-centered representations (Equations 4–5). Our guess is that the former will produce estimates of better quality. Similarly, we would like to analyze the effects of mis-estimating internal calibration parameters such as focal length, and to study the feasibility of estimating them as part of the reconstruction process. The results presented here have assumed for now that feature points are visible in all images. Our approach generalizes naturally to missing data points. In particular, we would like to study the effects feature tracks with relatively short lifetimes.

Finally, it appears that the portion of the uncertainty matrix which is correlated can be accounted for by a small number of modes. This suggest that an efficient recursive structure from motion algorithm could be developed which avoids the need for using full covariance matrices [THO93] but which performs significantly better than algorithms which ignore such correlations.

11 Conclusions

This technical report has developed new techniques for analyzing the fundamental ambiguities and uncertainties inherent in structure from motion. Our approach is based on examining the eigenvalues and eigenvectors of the Hessian matrix in order to quantify the nature of these ambiguities. The eigenvalues can also be used to predict the overall accuracy of the reconstruction.

Under orthography, the bas-relief ambiguity dominates the reconstruction error, even with large numbers of frames. This ambiguity disappears, however, for large object-centered rotations. For perspective cameras, two-frame solutions are possible, but there must still be a large amount of object rotation for best performance. Using three of more frames avoids some of the sensitivities associated with two-frame reconstructions. Translations towards the object are an alternative source of shape information, but these appear to be quite weak unless large fields of views and large motions are involved.

When available, prior information about the structure or motion (e.g., absolute distances, perpendicularities) can be used to improve the accuracy of the reconstructions. Whether 3-D reconstruction errors (for modeling) or motion estimation errors (for navigation) are most significant for a given application determines the conditions which produce acceptable results. In any case, careful error analysis is essential in ensuring that the results of structure from motion algorithms are sufficiently reliable to be used in practice.

References

[Adi89] G. Adiv. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):477–490, May 1989.

[AHP93] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pages 294–299, New York, New York, June 1993.

[Aya91] N. Ayache. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. MIT Press, Cambridge, Massachusetts, 1991.

[CT92] Q. Chen and S. Tsuji. A hierarchical method that solves the shape and motion from an image sequence problem. In *IEEE/RSJ Int'l Conference on Intelligent Robots and Systems*, pages 2131–2138, July 1992.

[CWC90] N. Cui, J. Weng, and P. Cohen. Extended structure and motion analysis from monocular image sequences. In *Third International Conference on Computer Vision (ICCV'90)*, pages 222–229, Osaka, Japan, December 1990. IEEE Computer Society Press.

[DA90] C. H. Debrunner and N. Ahuja. A direct data approximation based motion estimation algorithm. In *10th Int'l Conference on Pattern Recognition*, pages 384–389, 1990.

[Fau92] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Second European Conference on Computer Vision (ECCV'92)*, pages 563–578, Santa Margherita Liguere, Italy, May 1992. Springer-Verlag.

[HG93] R. Hartley and R. Gupta. Computing matched-epipolar projections. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pages 549–555, New York, New York, June 1993. IEEE Computer Society.

[HGC92] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'92)*, pages 761–764, Champaign, Illinois, June 1992. IEEE Computer Society Press.

[Hor90] B. K. P. Horn. Relative orientation. *International Journal of Computer Vision*, 4(1):59–78, January 1990.

[KTJ89] R. V. R. Kumar, A. Tirumalai, and R. C. Jain. A non-linear optimization algorithm for the estimation of structure and motion parameters. In *IEEE Computer Society Con-*

ference on Computer Vision and Pattern Recognition (CVPR'89), pages 136–143, San Diego, California, June 1989. IEEE Computer Society Press.

[KvD91] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8:377–385538, 1991.

[LH81] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[LH86] H. C. Longuet-Higgins. Visual motion ambiguity. *Vision Research*, 26(1):181–183, 1986.

[MQVB92] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG-12, LIFIA — IRIMAG, Grenoble, France, June 1992.

[MVQ93] R. Mohr, L. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pages 543–548, New York, New York, June 1993.

[OT91] J. Oliensis and J. I. Thomas. Incorporating motion error in multi-frame structure from motion. In *IEEE Workshop on Visual Motion*, pages 8–13, Princeton, New Jersey, October 1991. IEEE Computer Society Press.

[PFTV92] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, second edition, 1992.

[SA89] M. E. Spetsakis and J. Y. Aloimonos. Optimal motion estimation. In *IEEE Workshop on Visual Motion*, pages 229–237, Irvine, California, March 1989. IEEE Computer Society Press.

[SA91] M. E. Spetsakis and J. Y. Aloimonos. A multiframe approach to visual motion perception. *International Journal of Computer Vision*, 6(3):245–255, August 1991.

[Sha93] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. In *Fourth International Conference on Computer Vision (ICCV'93)*, pages 583–590, Berlin, Germany, May 1993. IEEE Computer Society Press.

[SK94] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using nonlinear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, March 1994.

[Sor80] H. W. Sorenson. *Parameter Estimation, Principles and Problems*. Marcel Dekker, New York, 1980.

[SPFP93] S. Soatto, P. Perona, R. Frezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pages 428–433, New York, New York, June 1993.

[SZB93] L. S. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. OUEL 1994/93, Oxford University Robotics Research Group, April 1993.

[TH84] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(1):13–27, January 1984.

[THO93] J. I. Thomas, A. Hanson, and J. Oliensis. Understanding noise: The critical role of motion error in scene reconstruction. In *Fourth International Conference on Computer Vision (ICCV'93)*, pages 325–329, Berlin, Germany, May 1993. IEEE Computer Society Press.

[TK92a] C. J. Taylor and D. J. Kriegman. Structure and motion from line segments in multiple images. In *IEEE International Conference on Robotics and Automation*, pages 1615–1621, Nice, France, May 1992. IEEE Computer Society Press.

[TK92b] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[TKA91] C. J. Taylor, D. J. Kriegman, and P. Anandan. Structure and motion in two dimensions from multiple images: A least squares approach. In *IEEE Workshop on Visual Motion*, pages 242–248, Princeton, New Jersey, October 1991. IEEE Computer Society Press.

[WAH89] J. Weng, N. Ahuja, and T. S Huang. Optimal motion and structure information. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'89)*, pages 144–152, San Diego, California, June 1989. IEEE Computer Society Press.

[WAH93] J. Weng, N. Ahuja, and T. S. Huang. Optimal motion and structure estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884, September 1993.

[Wol91] S. Wolfram. *Mathematica*$^{\mathrm{TM}}$*, A System for Doing Mathematics by Computer*. Addison-Wesley, 1991.

[YC92] G.-S. Y. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):995–1013, October 1992.

A Approximate minimum eigenvalue computation

The eigenvalues of a matrix of the form

$$\begin{bmatrix} a^2 & ab \\ ab & b^2+c^2 \end{bmatrix}$$

are the solutions to

$$\lambda^2 - \lambda(a^2+b^2+c^2) - a^2c^2 = 0,$$

i.e.,

$$\lambda = \frac{1}{2}(a^2+b^2+c^2 \pm \sqrt{(a^2+b^2+c^2)^2 - 4a^2c^2})$$

or for $c^2 \ll a^2+b^2$

$$\lambda_{\min} \approx \frac{a^2c^2}{a^2+b^2}$$

$$\lambda_{\max} \approx a^2+b^2$$

Similarly, for a quadratic of the form

$$a\lambda^2 - b\lambda + c = 0$$

with $ac \ll b^2$,

$$\lambda_{\min} = \frac{b-\sqrt{b^2-4ac}}{2a} \approx \frac{c}{b}. \tag{53}$$

To find the approximate minimum eigenvalue for the equiangular orthographic scanline camera, we substitute the values $C \approx \sum_j 1 \equiv J_0$, $S \approx \theta^2 J_2$, $E \approx \theta J_2$, $C' \approx J_2$, and $S' \approx \theta^2 J_4$, into (37),

$$\begin{aligned} 0 &= C\lambda^2 - (SC + (S'C - E^2)X + CC'Z)\lambda + S(S'C - E^2)X + C(C'S - E^2)Z \\ &\approx J_0\lambda^2 - (J_0 J_2(\theta^2 + Z) + \theta^2(J_0 J_4 - J_2^2)X)\lambda + \theta^4 J_2(J_0 J_4 - J_2^2)X + \theta^2 J_0(J_2^2 - J_2^2)Z. \end{aligned}$$

Using the approximation in (53), we obtain

$$\lambda_{\min} \approx \frac{\theta^4 X J_2(J_0 J_4 - J_2^2)}{J_0 J_2 Z + \theta^2[X(J_0 J_4 - J_2^2) + J_0 J_2]}. \tag{54}$$