Reconstruction of Linearly Parameterized Models from Single Images with a Camera of Unknown Focal Length

David Jelinek
GRASP Laboratory, CIS Department
University of Pennsylvania
3401 Walnut Street, Rm 329C
Philadelphia, PA, 19104-6228
email: davidj2@seas.upenn.edu
Phone: (215) 898 0340
Fax: (215) 573 2048

Camillo J. Taylor
GRASP Laboratory, CIS Department
University of Pennsylvania
3401 Walnut Street, Rm 335C
Philadelphia, PA, 19104-6228
email: cjtaylor@central.cis.upenn.edu
Phone: (215) 898 0376
Fax: (215) 573 2048

September 5, 2000

Abstract

This paper deals with the problem of recovering the dimensions of an object and its pose from a single image acquired with a camera of unknown focal length. It is assumed that the object in question can be modeled as a polyhedron where the coordinates of the vertices can be expressed as a linear function of a dimension vector, λ. The reconstruction program takes as input a set of correspondences between features in the model and features in the image. From this information the program determines an appropriate projection model for the camera (scaled orthographic or perspective), the dimensions of the object, its pose relative to the camera and, in the case of perspective projection, the focal length of the camera. This paper describes how the reconstruction problem can be framed as an optimization over a compact set with low dimension - no more than four. This optimization problem can be solved efficiently by coupling standard non-linear optimization techniques with a multistart method which generates multiple starting points for the optimizer by sampling the parameter space uniformly. The result is an efficient, reliable solution system that does not require initial estimates for any of the parameters being estimated.

Keywords: 3D Reconstruction, uncalibrated imagery, numerical optimization

1 Introduction

This paper deals with the problem of recovering the dimensions of an object and its pose from a single image acquired with a camera of unknown focal length. It is assumed that the object in question can be modeled as a polyhedron where the coordinates of the vertices can be expressed as a linear function of a dimension vector, λ. That is, if λ is an $n \times 1$ vector, then there are a set of $3 \times n$ matrices, $K_1, K_2, \ldots, K_m$, where the position of the ith vertex is given by $K_i\lambda$. Consider, for example, the model shown in Figure 1. For this model the following expressions detail how the coordinates of the vertices labeled, P_1 and P_2 can be expressed as linear functions of the dimension vector $\lambda = (LWHh)^t$.

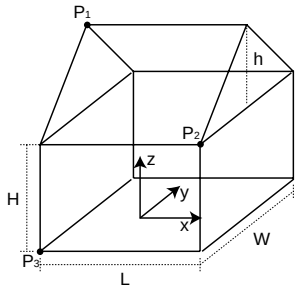


Figure 1: A simple example of a linearly parameterized polyhedral model. The coordinates of each of the vertices in this figure can be expressed as a linear function of the parameter vector $\lambda = (LWHh)^t$

$$P_1 = \begin{pmatrix} -L/2 \\ 0 \\ (H+h) \end{pmatrix} = \begin{pmatrix} -0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \lambda; P_2 = \begin{pmatrix} L/2 \\ -W/2 \\ H \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \lambda \tag{1}$$

In most situations the entries in the parameter vector, λ, will refer to dimensions that are only meaningful when positive. Note that generally speaking any polyhedron can be expressed by this model simply by choosing λ to be a vector of dimension $3N$ where N is the number of vertices in the model. In practice, most man-made objects, such as buildings, contain symmetries which allow the model to be expressed with far fewer parameters. For the model shown in Figure 1, the positions of 10 vertices can be characterized using only four parameters. This makes it possible to recover the model dimensions from measurements in a single image.

The input to the reconstruction program takes the form of a set of correspondences between features in the model, lines and points, and features in the image. From this information the program determines an appropriate projection model for the camera, scaled orthographic or perspective, the dimensions of the object, its pose relative to the camera and, in the case of perspective projection, the focal length of the camera.

The principal difficulties in solving this problem stem from the non-linearities associated with the unknown rotation, $R \in SO(3)$, that represents the orientation of the camera with respect to the objects frame of reference. In some situations it is possible to recover information about this rotation from vanishing points in the imagery. A number of systems have been proposed which exploit this cue [1, 7]. Less attention has been directed to cases where the vanishing point information is inconclusive or non-existent. The principal contribution of this paper is to describe a framework which is able to handle the full range of situations that can occur in practice including cases where *no* vanishing points are available.

Additionally, this paper describes how the reconstruction problem can be framed as an optimization over a compact set with low dimension - no more than four. This optimization problem can be solved efficiently by coupling standard non-linear optimization techniques with a multistart method which generates multiple starting points for the optimizer by sampling the parameter space uniformly. The result is an efficient, reliable solution system that does not require initial estimates for any of the parameters being estimated.

In [4] and [2] the problem of reconstructing models from one or more images taken with calibrated cameras was addressed. This paper improves on those results by proposing efficient techniques to deal with situations where the imagery was acquired with an incompletely calibrated camera and describes how the computational effort required to solve for all the unknown parameters can be reduced by taking advantage of the structure of the projection equations.

Tomasi and Kanade [8] and Pollefeys, Van Gool and Proesmans [5, 6] describe effective techniques for recovering the structure of a rigid scene from a sequence of images acquired under orthographic and perspective projection models respectively. However, multiframe techniques are not applicable in situations where only one image is available.

Section 2 of this paper presents an outline of the reconstruction procedure while sections 2.1 and 2.2 describe the solution to various subproblems of this reconstruction task. Section 3 presents results that were obtained with this algorithm on actual images and on simulated data. A discussion of our conclusions and future work is presented in Section 4.

2 Reconstruction Procedure

As described in the previous section, the reconstruction procedure hinges on the observation that the primary difficulties in the reconstruction problem center around the non-linearities introduced by the rotation between the camera frame and the objects frame of reference. Given an estimate for this rotation and the focal length of the camera, the other unknowns can be determined by finding the minima of a positive definite quadratic form - a well understood and well conditioned optimization problem which can be solved efficiently using standard techniques from linear algebra (see sections 2.1.2 and 2.2.2). This being the case, the proposed reconstruction method proceeds by conducting a search over the set of camera orientations and focal lengths for values that are in best agreement with the observed image measurements.

In the sequel we will discuss a variety of subcases for both perspective and orthographic projection, ranging from situations where all of the vanishing points can be observed to situations where none can be found. For each case we detail how the resulting reconstruction problem can either be solved directly or reformulated as an optimization over a compact set with low dimension. Once the problem has been reduced to this form it can be solved by applying standard non-linear optimization techniques and a multistart method which chooses starting points for the optimization procedure by sampling the parameter space uniformly. Such an approach is made feasible because of the fact that the parameter space can be bounded and can, therefore, be sampled effectively.

The first stage of the reconstruction procedure involves finding feature correspondences in the image data. A software system has been implemented that allows the user to specify correspondences between edges in the model and edges in the image by selecting a line in the model and then tracing the corresponding line in the image. Since the lines that the user draws are superimposed with the image, this method allows for very accurate recovery of

the image edges. Through this procedure we are able to associate vertices in the model with lines in the image. These point-to-line correspondences will be used in most calculations; however, in some cases we will require correspondences between model vertices and image points. These image points can be found by computing the intersections of the lines drawn by the user.

Once these correspondences have been established, the reconstruction procedure attempts to determine whether a scaled orthographic or perspective camera model should be employed. One simple way to distinguish between the two imaging situations is by analyzing lines in the image that correspond to parallel lines in the scene. If a set of lines in the image corresponding to parallel lines in the scene appear to verge then the system employs a perspective projection model.

In situations where no verging lines are found, the reconstruction procedure assumes a scaled orthographic projection model, recovers a solution for the unknown parameters, and then computes the residual disparity between the reprojected model vertices and the lines in the image. If this residual is above a certain threshold value, the system switches to a perspective model. Thus, the simpler projection model (ie. scaled orthographic) is favored if it explains the data sufficiently well.

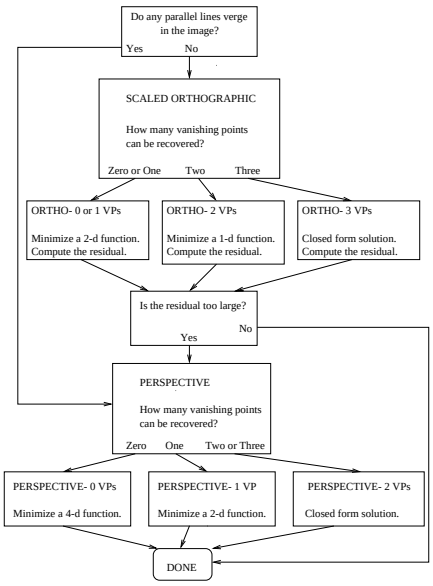


Figure 2: A flow chart describing the operation of the reconstruction procedure.

The next step in the reconstruction procedure is the computation of vanishing points in the image of the x-, y-, and z-directions of the model if possible. Suppose that the user specifies n lines in the model that are each parallel to the x-axis of the object. Let $l_1, l_2, ..., l_n$ be 3-vectors representing the projective coordinates of the corresponding lines in the image. Then the homogeneous coordinates of the vanishing point in the x-direction is the vector v_x that minimizes $\Sigma(l_i^t v_x)^2$. This vector can be found by eigenvalue decomposition of $A^t A$, where A is the matrix whose rows consist of the l_i^t's. The "best estimate" for the vanishing point is the eigenvector that corresponds to the eigenvalue of $A^t A$ with smallest magnitude.

Under a scaled orthographic projection model there are three cases to consider: three vanishing points recovered, two vanishing point recovered to vanishing points recovered. In the first case, the unknowns can be found in closed form. If only two vanishing points are recovered, the unknowns can be found by solving a one-dimensional minimization problem. In the last case, a two-dimensional optimization problem must be solved.

If the projection model is perspective, there are three possible cases: two or three vanishing points recovered, one vanishing point recovered, no vanishing points recovered. In the first case, the system can be solved in closed form. In the second case, the problem reduces to minimizing a function of two variables. In the last case, the problem reduces to minimizing a function of four variables.

In the sequel it is assumed that, after a suitable change of image coordinates, the aspect ratio of the camera is one and the coordinates of the principal point in the image are $(0, 0)$. In most situations the aspect ratio of the imaging device is known a'priori and the principal point is, for all practical purposes, coincident with the image center. In the case of scaled orthographic projection, the exact location of the principal point is, of course, immaterial to the reconstruction computation.

2.1 Scaled Orthographic Cases

Under the scaled orthographic projection model the projection matrix, P, which relates coordinates of points in the model to their projections on the image plane can be written as follows:

$$P = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \tag{2}$$

where f denotes the scale factor associated with this camera and $R \in SO(3)$ and $T \in \Re^3$ represent the rotation and translation of the camera with respect to the model frame.

2.1.1 Recovering Rotation from Vanishing Points

The homogeneous coordinates of the vanishing point in the image, v_x, corresponding to the x-direction in the model frame can be computed as follows:

$$v_x \propto P \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \propto \begin{pmatrix} R_{11} \\ R_{21} \\ 0 \end{pmatrix} \tag{3}$$

In an analogous manner, we can obtain expressions for v_y and v_z: $v_y \propto (R_{12}R_{22}0)^t$, $v_z \propto (R_{13}R_{23}0)^t$

When all three vanishing points can be recovered, we are effectively given three pieces of information about the rotation matrix R. That is, for some a, b, and c, the vanishing points give us:

$$\begin{pmatrix} A \\ D \end{pmatrix} = a \begin{pmatrix} R_{11} \\ R_{21} \end{pmatrix}, \begin{pmatrix} B \\ E \end{pmatrix} = b \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix}, \begin{pmatrix} C \\ F \end{pmatrix} = c \begin{pmatrix} R_{13} \\ R_{23} \end{pmatrix}, \tag{4}$$

Since the first two rows of R are each of unit length, we have the equations:

$$\left(\frac{A}{a}\right)^2 + \left(\frac{B}{b}\right)^2 + \left(\frac{C}{c}\right)^2 = 1 \tag{5}$$

$$\left(\frac{D}{a}\right)^2 + \left(\frac{E}{b}\right)^2 + \left(\frac{F}{c}\right)^2 = 1 \tag{6}$$

Because the first two rows of R are orthogonal to each other, we have the equation:

$$\frac{AD}{a^2} + \frac{BE}{b^2} + \frac{CF}{c^2} = 0$$

This can be summarized as a system of three linear equations in three unknowns:

$$\begin{bmatrix} A^2 & B^2 & C^2 \\ D^2 & E^2 & F^2 \\ AD & BE & CF \end{bmatrix} \begin{pmatrix} \frac{1}{a^2} \\ \frac{1}{b^2} \\ \frac{1}{c^2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

which can easily be solved to yield a,b, c, and ultimately R by utilizing the fact that the third row of R is simply the cross product of the first two rows. There is actually a four-way ambiguity in recovering R because the signs of a, b, and c are unknown. The rotation matrix

is chosen in such a way that the corresponding optimal solution for the dimension vector λ consists entirely of positive entries.

There are situations where the system of linear equations described above will become singular. This will occur when two of the vanishing points are coincident. In this case the more general reconstruction procedure described in Section 2.1.4 will be invoked to obtain a solution.

2.1.2 Recovering Scene Dimensions

Once an estimate for the rotation matrix becomes available all that remains is to calculate the dimension vector λ and t. According to the model, the coordinates of the jth vertex in the world frame are given by $K_j\lambda$. Let $l_{jk} = (l^x_{jk} l^y_{jk} l^z_{jk})^t$ represent the homogeneous coordinates of the line in the image plane connecting points j and k. Then the constraint that the projection of the jth vertex in the image should lie along this line can be expressed as follows:

$$\begin{aligned} l^t_{jk} P \begin{pmatrix} K_j\lambda \\ 1 \end{pmatrix} &= 0 \\ \Rightarrow (l^x_{jk} l^y_{jk})[fG(RK_j\lambda + T)] + l^z_{jk} &= 0 \\ \Rightarrow (l^x_{jk} l^y_{jk}) \begin{bmatrix} (GRK_j) & I \end{bmatrix} \begin{pmatrix} f\lambda \\ fT_x \\ fT_y \end{pmatrix} + l^z_{jk} &= 0 \end{aligned}$$

Where $G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. So for each point to line correspondence we can construct an affine equation in the parameter vector $\begin{pmatrix} f\lambda \\ fT_x \\ fT_y \end{pmatrix}$. If a sufficient number of correspondences are available one can obtain a solution for this parameter vector by solving the resulting linear system. Note that this procedure yields no information about the z component of the translation vector T. It is also important to keep in mind that the solution only yields the dimensions of the scene up to a scale factor since it is impossible to separate the scale parameter f from the other variables in the vector.

2.1.3 Two Vanishing Points Recovered

In situations where only two of the three vanishing points are available it is possible to obtain a solution for the reconstruction problem using the procedures given above by optimizing over all possible values for the missing vanishing point.

Suppose, for example, we are given v_x and v_y then we can obtain estimates for the scene structure by minimizing the following function from the interval $[0, \pi]$ to $\Re^+$:

`function Res (`θ`)`

`Step 1) Let` $v_z = \begin{pmatrix} cos(\theta) \\ sin(\theta) \\ 0 \end{pmatrix}$.

`Step 2) Using the procedures in sections 2.1.1 and 2.1.2, compute` $R, f\lambda, fT_x$ `and` fT_y.

`Step 3) Calculate the residue,` $\Sigma(l_{ij}^t P \begin{pmatrix} K_i\lambda \\ 1 \end{pmatrix})^2$`, and return this value.`

One can use standard minimization techniques such as Golden Section Search to minimize the value of $Res(\theta)$ and thus find the appropriate values for the unknown parameters. Since this is an optimization problem with only one degree of freedom on a bounded interval it can be solved quite quickly.

2.1.4 No Vanishing Points Recovered

In the case where no vanishing point information is available the reconstruction system makes use of correspondences between model vertices and image points. If (u_i, v_i) represents the measured location of the projection of the ith model vertex in the image then the system chooses values of the unknown parameters to minimize the discrepancy between the observed image locations and the predicted values. That is, the goal of the reconstruction system is to minimize the following objective function, O, where the rotation matrix R has been rewritten as the product of a series of rotations about the x, y and z axes and the matrix G is defined in Section 2.1.2.

$$O = \Sigma \left\| \begin{pmatrix} u_i \\ v_i \end{pmatrix} - fG(R_z(\gamma)R_y(\beta)R_x(\alpha)K_i\lambda + T) \right\|^2$$

This expression can be simplified by utilizing the fact that rotation about the optical axis, z, corresponds to a planar rotation of the image features. So if the angles α and β were known, O could be rewritten as:

$$\begin{aligned} O &= \Sigma \left\| \begin{pmatrix} u_i \\ v_i \end{pmatrix} - \begin{pmatrix} c & -s \\ s & c \end{pmatrix} (L_i \lambda' + \begin{pmatrix} T'_x \\ T'_y \end{pmatrix}) \right\|^2 \\ &= \Sigma \left\| \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} - (L_i \lambda' + \begin{pmatrix} T'_x \\ T'_y \end{pmatrix}) \right\|^2 \end{aligned}$$

Where $L_i = GR_y(\beta)R_x(\alpha)K_i$, $c = \cos\gamma$, $s = \sin\gamma$, $\lambda' = f\lambda$ and $\begin{pmatrix} T'_x \\ T'_y \end{pmatrix} = f \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} T_x \\ T_y \end{pmatrix}$

In this situation it is possible to compute optimal estimates for γ, λ', T'_x and T'_y by rewriting the objective function as follows:

$$\begin{aligned} O &= \Sigma \left\| \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} - (L_i \lambda' + \begin{pmatrix} T'_x \\ T'_y \end{pmatrix}) \right\|^2 \\ &= \Sigma \left\| \begin{pmatrix} u_i & v_i \\ v_i & -u_i \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} - I \begin{pmatrix} T'_x \\ T'_y \end{pmatrix} - L_i \lambda' \right\|^2 \\ &= \Sigma \left\| \begin{bmatrix} u_i & v_i & 1 & 0 & -L_i \\ v_i & -u_i & 0 & 1 & \end{bmatrix} \begin{pmatrix} c \\ s \\ T'_x \\ T'_y \\ \lambda' \end{pmatrix} \right\|^2 \end{aligned}$$

This can be recognized as the standard problem of finding a vector $x = \begin{pmatrix} c & s & T'_x & T'_y & \lambda' \end{pmatrix}^t$ to minimize $\|Ax\|^2$ subject to the constraint $\|Bx\|^2 = 1$ where the matrix B is chosen to reflect the constraint that $c^2 + s^2 = 1$. This generalized eigenvalue problem can be solved efficiently using standard techniques from linear algebra [3].

The ability to compute optimal estimates for $f\lambda$, γ, fT_x and fT_y in this manner suggests that a solution for the reconstruction problem can be obtained by finding values of α and β that minimize the following residual function:

`function Res2(`α, β`)`

`Step 1) Let` $L_i := GR_y(\beta)R_x(\alpha)K_i$ `for all` i.

Step 2) Solve the generalized eigenvalue problem to recover γ, λ', T'_x and T'_y and return the residual value, O, for these values.

Once again, the problem has been reduced to an optimization over a small number of bounded parameters, in this case α and β.

2.2 Perspective Cases

In the case of perspective projection the matrix of intrinsic parameters is given by:

$$A = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where f is the focal length of the camera.

2.2.1 Recovering Rotation from Two Vanishing Points (not at Infinity)

If two of the vanishing points corresponding to the axes of the objects frame of reference can be recovered where neither one is a point at infinity, then the rotation matrix, R, can be recovered in closed form [1]. Suppose, for example, we are given v_x and v_y. Then we have the following proportions:

$$v_x \; \alpha \; AR\hat{x}, \quad v_y \; \alpha \; AR\hat{y}$$

where $\hat{x}$ and $\hat{y}$ are simply the unit vectors along the x and y axes respectively. Since $R\hat{x}$ is orthogonal to $R\hat{y}$ we have the equation:

$$(A^{-1}v_x)^t(A^{-1}v_y) = 0$$

which can be rewritten as follows:

$$\frac{v_{x1}v_{y1}}{f^2} + \frac{v_{x2}v_{y2}}{f^2} + v_{x3}v_{y3} = 0$$

$$\Rightarrow f = \sqrt{\frac{v_{x1}v_{y1} + v_{x2}v_{y2}}{-v_{x3}v_{y3}}}$$

The first column of R can then be found by normalizing the vector $A^{-1}v_x$ The second column can be found in a similar manner, and the third column is simply the cross product

of the first two columns. Again there will be a four-way ambiguity in the solution for R which can be resolved by choosing the a solution which results in a dimension vector with positive entries.

This method will succeed so long as neither v_{x3} or v_{y3} are equal to zero. If one or both of the vanishing points are at infinity then the method described in section 2.2.3 can be employed to produce a reconstruction.

2.2.2 Recovering Scene Dimensions

The dimension vector λ and the camera translation T can be found in a manner similar to the method described in Section 2.1.2. Let l_{jk} represent the homogeneous coordinates of the line in the image plane connecting points j and k. Then the constraint that the projection of this vertex in the image should lie along this line can be expressed as follows:

$$\begin{aligned} l_{jk}^t A(RK_j\lambda + T) &= 0 \\ \Rightarrow l_{jk}^t [ARK_j \ A] \begin{pmatrix} \lambda \\ T \end{pmatrix} &= 0 \end{aligned}$$

Let M be a matrix formed by stacking the rows of the form $l_{jk}^t[ARK_j \ A]$. Then an estimate for $\begin{pmatrix} \lambda \\ T \end{pmatrix}$, up to a scale factor, can be obtained by finding the unit vector that minimizes $\|M \begin{pmatrix} \lambda \\ T \end{pmatrix}\|^2$. This is a standard eigenvalue problem.

2.2.3 One Vanishing Point Recovered

The previous section describes how estimates for λ and T can be computed once estimates for R and f are available. Knowledge of any vanishing points in the image essentially constrains two of the four degrees of freedom associated with the rotation matrix R and the focal length parameter f. We can exploit this constraint by constructing an objective function which computes the residual of the reconstruction as a function of the remaining two degrees of freedom.

Consider the case where the vanishing point in the x-direction, v_x, is known. In this case the problem can be parameterized in terms of a variable θ which captures the remaining

degree of freedom of the rotation matrix and an angle ρ which denotes the field of view of the camera in the x-direction. If the x dimension of the image is m pixels then the focal length, f, is given by $(m/2)\cot(\rho/2)$. The advantage of parameterizing the system in terms of the field of view, ρ, instead of the focal length, f, is that the parameter ρ can be restricted to the interval $[0, \pi]$ while the parameter, f is unbounded.

For a given value of ρ one can compute f and, hence, the matrix A. Once a value of A has been generated it is quite easy to generate a rotation matrix that would generate the observed vanishing point, that is a matrix $R' \in SO(3)$ such that $R' \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \propto A^{-1} v_x$. One way to accomplish this is by a Gram-Schmidt orthonormalization process. The entire set of rotation matrices which preserve the vanishing point in the x direction can then be parameterized as follows: $R = R' R_x(\theta)$. Once again there will be a four way ambiguity in the rotation matrix that must be accounted for.

Based on this analysis, the reconstruction problem can be solved by finding the minimum of the following residual function.

`function Res3` (ρ, θ)

`Step 1) Let` $f = (m/2)cot(\rho/2)$.

`Step 2) Let` $A = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$

`Step 3) Generate a matrix` R' `such that` $R' \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \propto A^{-1} v_x$

`Step 4) Let` $R := R' R_x(\theta)$

`Step 5) Compute estimates for` λ `and` T

`Step 6) Calculate the residue,`

$\Sigma(l_{ij}^t A(RK_i\lambda + T))^2$`, and return this value.`

If a second vanishing point is available it can be used to resolve the ambiguity associated with the rotation matrix. For example if the vanishing point in the y direction, v_y, is available then once an A matrix has been chosen one can determine the camera orienta-

tion immediately by selecting a rotation matrix, R, where the first and second columns are proportional to $A^{-1}v_x$ and $A^{-1}v_y$ respectively. Effectively this reduces the reconstruction problem to an optimization over a single parameter, ρ. As mentioned previously, this approach should be preferred to the one described in section 2.2.1 in situations where one or both of the vanishing points are at infinity.

2.2.4 No Vanishing Points Recovered

When no vanishing point information is available finding a solution for the reconstruction problem can be recast as finding values for R and f that minimize the residual function described below. This optimization is carried out over four bounded parameters: α, β, and γ which represent an Euler angle parameterization of R and ρ which denotes the field of view of the camera.

`function Res4` $(\alpha, \beta, \gamma, \rho)$

`Step 1) Let` $R = R_z(\gamma)R_y(\beta)R_x(\alpha)$ `and let` $f = (m/2)cot(\rho/2)$.

`Step 2) Let` $A = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$

`Step 3) Using the procedure described in section 2.2.2, compute` λ, `and` T.

`Step 4) Calculate the residue,` $\Sigma(l_{ij}^t AR(K_i\lambda + T))^2$, `and return this value.`

3 Experimental Results

3.1 Simulation Results

In order to investigate the efficacy of the proposed reconstruction system a series of trials were carried out on simulated data sets. In these experiments the most general versions of the perspective and orthographic reconstruction techniques were used; namely those described in sections 2.2.4 and 2.1.4 respectively. These methods do not make any use of vanishing point information and are formulated as optimizations over four and two parameters respectively.

For each of these cases we generated image measurements corresponding to a polyhedron with 64 vertices parameterized by 19 dimensions viewed from 20 different vantage points. The

simulated image measurements were corrupted with noise equivalent to 1 pixel in a 400 by 300 image. The multistart optimization procedure invoked standard numerical minimization procedures from randomly chosen starting points until a minima with an acceptable residual value is found. The number of trials required to find an acceptable minima along with the errors in the estimated parameters at convergence was recorded.

Note that since the reconstruction procedure can only recover the dimension and camera translation parameters up to a scale factor, the error was calculated by first scaling the recovered parameters until the mean squared disparity between the recovered parameters and the true parameter values was minimized. The percentage error between the recovered parameter vector, λ, and the actual parameter values, λ_t, was then computed from the following ratio, $\|\lambda - \lambda_t\|/\|\lambda_t\|$.

For the perspective case the average number of trials needed to find the minimum was 4.9. At convergence the average error in the rotation parameter was 0.30 degrees the average error in the recovered field of view was 0.42 degrees and the average error in the dimension parameters was 0.66%. For the orthographic case the multistart method required 2 trials on average to find an appropriate minimum. At convergence the average error in the rotation parameter was 0.25 degrees while the average error in the dimension parameters was 2%.

As with any image-based reconstruction technique where disparity in the image is used as a proxy for metric error, it is possible to construct degenerate configurations where one or more of the object dimensions cannot be recovered from the image data. The canonical example would be a box viewed under orthographic projection along one of its axes, in this case the dimension of the object along the viewing direction cannot be recovered. In these situations the proposed technique may return results with a large error in the unobservable parameters.

3.2 Results on real images

The following results were obtained using photographs taken with a Kodak DC210 digital camera. All of the images were acquired in high-resolution mode, which produces 864×1152 images.

Figure 3a shows a Jell-O box adjacent to a block of wood, and Figure 3bc show texture-

a.  b.  c. 

Figure 3: a. Two boxes with slight perspective effects. b. c. Texture mapped reconstructions of the scene.

mapped reconstructions of the scene viewed from novel vantage points. The reconstruction was done using the method of Section 2.2.1 (two or three vanishing points found under perspective) and then the estimates of the parameters were refined using the non-linear minimization of Section 2.2.4. The vector λ, which gives the dimensions of of the object were measured by hand and found to be (in millimeters) $(35\ 86\ 72\ 19\ 39\ 78)^t$. After choosing an appropriate scaling factor, the reconstruction gave an estimate (in millimeters) of $(33.7\ 85.7\ 72.4\ 18.0\ 39.2\ 78.6)^t$. This represents an RMS error of 0.75 mm. Notice that we cannot check the accuracy of the pose estimation because we do not have a truth model of these parameters.

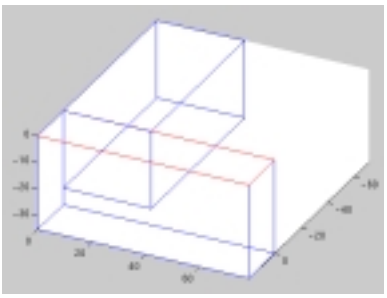
a.  b. 

Figure 4: a. Two boxes under a near-orthographic projection. b. Wireframe reconstruction

Figure 4a is an image of two blocks of wood under a near-orthographic projection. The wireframe reconstruction in Figure 4b was obtained using the algorithm of Section 2.1.4 (no vanishing points under orthography) though we could have obtained a starting point for this minimization using the available vanishing points. The dimension vector was given in millimeters by $(78\ 19\ 39\ 31\ 69.5\ 31)^t$ and the algorithm gave an estimate in millimeters of $(78.2\ 19.6\ 35.3\ 32.5\ 71.0\ 29.1)^t$, which yields an RMS error of 1.9 mm.

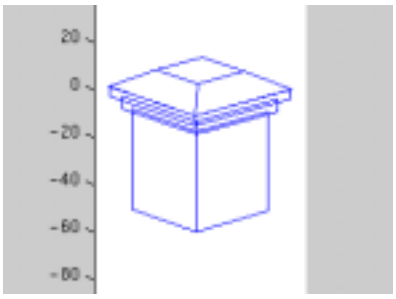
a.  b. 

Figure 5: a. A pyramid atop three boxes under a near-orthographic projection. b. Wireframe reconstruction

The image in Figure 5a is a stone structure on the University of Pennsylvania campus. We modeled it as a frustum atop a stack of three boxes. (We ignored the pyramid that is above the frustum.) Using a scaled orhtographic projection model, we obtained the wireframe in Figure 5b. The dimension of the object are given (in inches) by $(25\ 6.5\ 13\ 24\ 2\ 22\ 2.5\ 18\ 45)^t$ and the algorithm estimated the dimensions as $(26\ 7.5\ 13\ 26\ 4\ 22\ 5\ 19\ 42)^t$. The RMS error in this case was 1.7 inches. This reconstruction was not as accurate as the others partly because much of the stone was chipped away from the structure and this made edge identification difficult. Additionally, the structure does not have precise right angles and only somewhat approximates our model of a frustum above a stack of boxes. It should be noted, however, that the only inaccurate measures corresponded to the height of each box. These heights are small compared to the other measurements and difficult to discern in the photograph.

a.  b.  c. 

Figure 6: a. The Penn Center in Philadelphia. b. c. Texture mapped reconstructions

Figure 6a shows the Penn Center in Center City, Philadelphia. The reconstructions are shown in Figure 6b and 6c. Philadelphia's Art Museum is shown in figure 7a; its reconstruc-

tions are shown in figure 7b and 7c.

a.  b.  c. 

Figure 7: a. The Art Museum in Philadelphia. b. c. Texture mapped reconstructions

4 Conclusion

This paper presents a practical scheme for recovering models of polyhedral objects from single images taken with a camera of unknown focal length. The resulting algorithm can be used to recover accurate three dimensional models of polyhedral objects from commonly available imagery including images obtained from websites or scanned from newspapers. Experimental results have been presented which demonstrate the accuracy and efficacy of the proposed techniques on simulated data and on actual images.

Future work will address the use of multiple views of objects to better recover parameters and the use of automated edge extraction. We believe that most of the error in our estimates of λ were due to human error in drawing the edges. A better system would allow the user to specify the approximate location of an edge and then have the software refine this estimate automatically based on image gradients.

Acknowledgements This research was supported by the National Science Foundation under a CAREER grant (IIS98-7687) and under an NSF Graduate Research Training Grant (GER93-55018).

References

[1] Bruno Caprile and Vincent Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–140, March 1990.

[2] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of SIGGRAPH 96. In Computer Graphics Proceedings, Annual Conference Series*, pages 11–21, New Orleans, LA, August 4-9 1996. ACM SIGGRAPH.

[3] Olivier Faugeras. *Three-Dimensional Computer Vision.* MIT Press, 1993.

[4] David G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(5):441–450, May 1991.

[5] M. Pollefeys, L Van Gool, and M. Proesmans. Euclidean 3d reconstruction from image sequences with variable focal lengths. In *European Conference on Computer Vision*, pages 31–42, 1996.

[6] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *International Conference on Computer Vision*, pages 90–95, 1998.

[7] Jeffery A. Shufelt. *Projective Geometry and Photometry for Object Detection and Delineation.* PhD thesis, Carnegie Mellon University, July 1996. CMU-CS-96-164.

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