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

David Jelinek Camillo J. Taylor
General Robotics and Active Sensory Perception (GRASP) Laboratory
University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104-6228
E-mail: {davidj2, cjtaylor}@grip.cis.upenn.edu

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. We demonstrate that this reconstruction task can be framed as an unconstrained optimization problem involving a small number of variables, no more than four, regardless of the number of parameters in the dimension vector.

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 $K_i\lambda$. In practice, many man-made objects can be described by such a model. For example, most buildings can be readily described in terms of a collection of axis-aligned polyhedral primitives, blocks, wedges, frustums etc.

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. We demonstrate that this reconstruction problem can be framed as an unconstrained optimization problem over a small number of variables, no more than four, regardless of the number of parameters in the dimension vector.

In [3] 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 [6] and Pollefeys, Van Gool and Proesmans [4, 5] describe effective techniques for recovering the structure of a rigid scene from a sequence of images acquired under orthographic and perspective projection respectively. However, multiframe techniques are not applicable in situations where only one image is available.

Caprile and Torre [1] describe a method for calibrating a perspective camera from three vanishing points in the image. Our procedure takes advantage of vanishing points when they are available but does not require them. When sufficiently many vanishing points are found an initial solution for the unknown parameters can be computed directly and then refined using non-linear optimization techniques. If the vanishing point information is inconclusive or unavailable the system resorts to solving an optimization problem over four variables in the worst case. One of the contributions of this paper is a novel technique for recovering the structure of a scene under scaled orthographic projection when no vanishing points are available.

Section 2 of this paper presents an outline of the reconstruction procedure while sections 3 and 4 describe the solution to various subproblems of this reconstruction task. Section 5 presents results that were obtained with this algorithm on actual images. A discussion of our conclusions and future work are presented in Section 6.

2 Reconstruction Procedure

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 lines. 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.

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. The homogeneous coordinates of the vanishing points in the image are computed in the usual manner. Suppose 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 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.

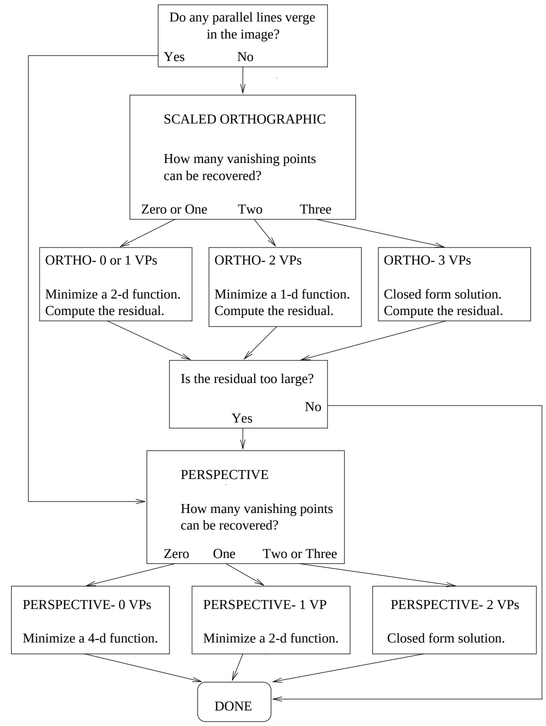

Figure 1: A flow chart describing the operation of the reconstruction procedure.

Under a scaled orthographic projection model there are three cases to consider.

- Three vanishing points recovered.
- Two vanishing point recovered.
- No 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 are recovered and not at infinity.
- One vanishing point recovered and not at infinity.
- No finite 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.

3 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:

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

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.

3.1 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:

$$v_x \propto P \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \propto \begin{pmatrix} R_{11} \\ R_{21} \\ 0 \end{pmatrix} \tag{2}$$

In an analogous manner, we can obtain expressions for v_y and v_z: $v_y \propto \begin{pmatrix} R_{12} \\ R_{22} \\ 0 \end{pmatrix}$, $v_z \propto \begin{pmatrix} R_{13} \\ R_{23} \\ 0 \end{pmatrix}$.

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:

$$aR_{11},\ bR_{12},\ cR_{13},$$
$$aR_{21},\ bR_{22},\ cR_{23},$$

We shall names these quantities as follows:

$$A,\ B,\ C$$
$$D,\ E,\ F$$

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

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

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

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 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 3.4 will be invoked to obtain a solution.

3.2 Recovering Scene Dimensions

Once an estimate for the rotation matrix becomes available all that remains is to calculate λ and t. According to the model, the coordinates of the jth 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:

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

Where $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. 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 in the usual manner. 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.

3.3 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 $[0, \pi]$ to $\Re^+$:

`function Res` (θ)

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

`Step 2) Using the procedure in the previous section, 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 technique 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, it can be solved quite quickly.

3.4 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 a product of a series of rotations about the x, y and z axes and the matrix C is defined in Section 3.2.

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

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 $L_i = CR_y(\beta)R_x(\alpha)K_i$, $c = \cos\gamma$, $s = \sin\gamma$, $\lambda' = f\lambda$ 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 $x = \begin{pmatrix} c & s & T'_x & T'_y & \lambda' \end{pmatrix}^t$ that minimizes $\|Ax\|^2$ 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 using standard techniques from linear algebra.

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` $L_i := CR_y(\beta)R_x(\alpha)K_i$ `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.

4 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.

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

If two vanishing points can be recovered under perspective projection where neither one is a point at infinity, then the rotation matrix, R, can be recovered in closed form. Suppose, for example, we are given v_x and v_y. Then we have the following proportions:

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

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.

4.2 Recovering Scene Dimensions

The parameters λ and t can be found in a manner similar to the method described in Section 3.2. If l_{jk} represents 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 $\|M\underline{x}\|^2$. This is a standard eigenvalue problem.

4.3 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 three degrees of the rotation matrix R. 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. The reconstruction problem can then be solved by finding the minimum of this residual function.

Consider the case where the vanishing point in the x-direction, v_x, is known (and is not at infinity for simplicity) [1].

We can choose to parameterize the problem in terms of an angle 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 resulting residual function is given below:

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) Let $C_1 := A^{-1}v_x$. This represents the first column of R (up to a scale).

Step 4) Let $C_2 := \begin{pmatrix} cos(\theta) \\ sin(\theta) \\ 0 \end{pmatrix} \times C_1$. This represents the second column of R (up to a scale). Note that C_2 is orthogonal to C_1.

Step 5) Let $C_3 := C_1 \times C_2$.

Step 6) Let $R := [\widehat{C_1} \ \widehat{C_2} \ \widehat{C_3}]$.

[1]This restriction can be removed with a slight increase in the complexity of the algorithm

Step 7) Compute estimates for λ and T
Step 8) Calculate the residue,
$\Sigma(l_{ij}^t A(RK_i\lambda + T))^2$, and return this value.

4.4 No Vanishing Points Recovered

When no vanishing point information is available finding a solution for the reconstruction problem involves finding values for R and f that result in the lowest residual values. This can be seen as an optimization problem involving four degrees of freedom as described below:

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

5 Experimental Results

We present the results of three experiments using different photographs taken with a Kodak DC210 digital camera. All of the images were acquired in high-resolution mode, which produces 864×1152 images.

Figure 2: Two boxes with slight perspective effects.

Figure 2 shows a Jell-O box adjacent to a block of wood, and Figure 3 shows a wireframe reconstruction of the scene viewed from a completely different vantage point. The reconstruction was done using the method of Section 4.1 (two or three vanishing points found under perspective) and then the estimates of the parameters were refined using the non-linear minimization of Section 4.4. The vector λ, which gives the dimensions of of the object were

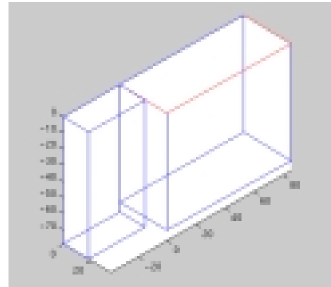

Figure 3: Wireframe reconstruction of Figure 2.

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.

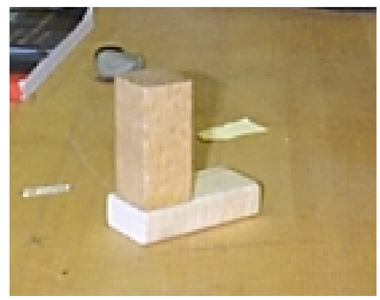

Figure 4: Two boxes under a near-orthographic projection.

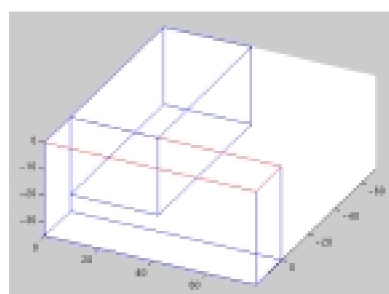

Figure 5: Wireframe reconstruction of Figure 4.

Figure 4 is an image of two blocks of wood under a near-orthographic projection. The wireframe reconstruction in Figure 5 was obtained using the algorithm of Section 3.4 (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.

Figure 6: A pyramid atop three boxes under a near-orthographic projection.

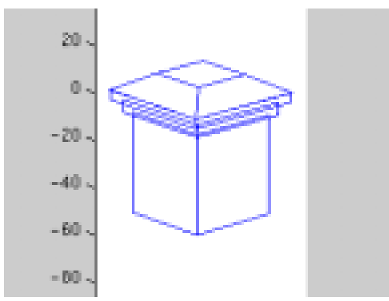

Figure 7: Wireframe reconstruction of Figure 6.

The image in Figure 6 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 7. 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 partially because much of the stone was chipped away from the structure and this made our 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.

6 Conclusion

This paper presents a practical scheme for recovering models of polyhedral objects from single images taken with a camera of unknown focal length. Experimental results have been presented which demonstrate the accuracy and efficacy of these techniques on actual image data.

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 by examining the image gradients.

Acknowledgments

This work was supported by an NSF Training Grant GER93-55018, by a DoE GAANN fellowship, and DoD MURI DAAH04-96-1-0007.

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] David G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5), May 1991.

[4] 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.

[5] 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.

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