

Reconstruction of Linearly Parameterized Models from a Single Image Using the Vanishing Points

Yong-In Yoon¹, Jang-Whan Im¹, Dae-Hyun Kim¹, Jong-Soo Choi¹,
and Jeong-Su Oh²

¹Department of Image Engineering Graduate School of Advanced Imaging Science,
Multimedia, and Film Chung-Ang University, 221 Huksuk-Dong, DongJae-Ku, Seoul,
South Korea

{yoonyi, jhim, vante77, jschoi}@imagelab.cau.ac.kr

²Department of Image Engineering, Pukyong National University, 599-1 Daeyondong,
Namgu, Pusan, South Korea

{ojs}@imagelab.cau.ac.kr

Abstract. We present a method using only three vanishing points to recover the dimensions of object and its pose from single image of perspective projection with a camera of unknown focal length. Our approach is to compute the dimensions of objects represented by the unit vector of objects from the image. The dimension vector \mathbf{v} of objects can be solved by the standard nonlinear optimization techniques with a multistart method which generates multiple starting points for the optimizer by sampling the parameter space uniformly. This method allows model-based vision to be computed the dimensions of object for a 3D model from matches to a single 2D image. Experimental results demonstrate the dimension vector \mathbf{v} of the proposed method from a single image using three vanishing points and show a performance of the proposed method compared to the conventional. Then, the actual dimensions of object from the image agree well with the calculated results.

1 Introduction

An essential role of model-based vision is the ability to solve for the values of all viewpoint and model parameter that will best fit a model to some matching image. Most works of model-based vision have been based on models of simple polyhedral 3D objects that allow the model to be expressed with far fewer parameters. These models are composed of primitives of the Constructive Solid Geometry (CSG) [1]. The CSG primitives are simple objects such as cubes, boxes and tetrahedrons. The method for the reconstruction of CSG primitives allows model-based vision to be computed for the dimensions of object for a 3D model from matches to a single 2D image. There are thus many objects such as buildings, boxes, façade and architectures. Recently, the

techniques of model-based vision have applied in the fields of 3-dimensional reconstruction, 3D models and computer graphics [2][3].

As shown in Fig. 1, there is the object of primitives to be modeled as a polyhedral where the coordinates of the vertices can be represented as a linear function of a dimension vector \mathbf{v} . For this model, the technique of reconstruction made it possible to recover the model dimensions from measurement in a single image [3]. The input information of projection from 3D to 2D image is also used to a set of correspondences between features in the model, lines and points, and features in the image. Therefore, we can determine the perspective projection model for the dimensions of the object, its pose relative to the camera and the focal length of the camera from single image acquired with a camera of unknown focal length. From this model, it is most important problem to recover the dimensions of objects for a 3D model from matches to a single 2D image in order to recover the 3D reconstruction. It is thus to recover the dimensions of objects that are to calculate the dimension vector \mathbf{v} of objects represented by the unit vector of objects from single image of perspective projection.

A way to reconstruct models from one or more image acquired with calibrated cameras improved on those results by proposing efficient techniques to deal with situation where the imagery was acquired with an incompletely calibrated camera [2].

David et al. [3] described an effective method for recovering the dimensions of object under orthographic and perspective projection models with unknown focal length. For the case of perspective projection, the technique was used to no vanishing point, one vanishing point and two vanishing points to recover the dimensions of the object, its pose from the image.

In this paper, we present the method to use three vanishing points for recovering the dimensions of object and its pose from single image of perspective projection acquired with a camera of unknown focal length. If three vanishing points are located in the image, it is possible to compute the dimension vector, its pose and the focal length from the image acquired with the camera of unknown focal length. The dimension vector \mathbf{v} of objects can be solved by the standard nonlinear optimization techniques with a multistart method which generates multiple starting points for the optimizer by sampling the parameter space uniformly.

This paper can be represented as follows: Section 2 describes the solution of the proposed reconstruction algorithm in the case of perspective projection and accurately finding vanishing points. Section 3 presents results obtained with the proposed algorithm on actual images and on simulated the result. Section 4 describes conclusions and future works of this paper.

2 Reconstruction Algorithm

2.1 Perspective Projection Model

In the case of perspective projection model, the homogeneous camera projection matrix \mathbf{P} , which relates coordinates of points in the model to their projections on the image plane, can be given by:

$$\mathbf{K} = \begin{bmatrix} \mathbf{f} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{r}_{13} \\ \mathbf{r}_{21} & \mathbf{r}_{22} & \mathbf{r}_{23} \\ \mathbf{r}_{31} & \mathbf{r}_{32} & \mathbf{r}_{33} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} \mathbf{T}_x \\ \mathbf{T}_y \\ \mathbf{T}_z \end{bmatrix}$$

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \quad \mathbf{T}]$$

where \mathbf{f} denotes the focal length of a camera, \mathbf{R} and \mathbf{T} represent the rotation and translation of the camera with respect to the model frame. Also, it is assumed that the aspect ratio of the camera is one and the coordinates of the principal point in the image are (0, 0). In most cases, before the camera operates, the aspect ratio of the imaging device is, of course, known and the principal point is located at the image center to compute the reconstruction problem for all practical purposes from the image of perspective projection.

2.2 The relationship between vanishing points and world coordinate system

If $\hat{\mathbf{X}}=[1,0,0,0]$, $\hat{\mathbf{Y}}=[0,1,0,0]$ and $\hat{\mathbf{Z}}=[0,0,1,0]$ are a direction of lines parallel to \mathbf{X} , \mathbf{Y} and \mathbf{Z} axes respectively, the vanishing points for \mathbf{X} , \mathbf{Y} and \mathbf{Z} axes respectively are the points that these directions of $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ are mapped the image by camera projection $\mathbf{P}=\mathbf{K}[\mathbf{R} \mid \mathbf{T}]$. Suppose $\mathbf{V}_x, \mathbf{V}_y$ and \mathbf{V}_z are represented the directions of $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ respectively. Then, we can express the following equations below.

$$\begin{aligned} \mathbf{V}_x &= \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\hat{\mathbf{X}} \\ \mathbf{V}_y &= \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\hat{\mathbf{Y}} \\ \mathbf{V}_z &= \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\hat{\mathbf{Z}} \end{aligned} \quad (1)$$

In the equation (1), from the rotation matrix \mathbf{R} , we can get the equation as follows:

$$\mathbf{V}_x = \mathbf{K}\mathbf{R}_1\hat{\mathbf{X}}, \mathbf{V}_y = \mathbf{K}\mathbf{R}_2\hat{\mathbf{Y}}, \mathbf{V}_z = \mathbf{K}\mathbf{R}_3\hat{\mathbf{Z}} \quad (2)$$

where $\hat{\mathbf{X}}, \hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ are the directions along the \mathbf{X} , \mathbf{Y} and \mathbf{Z} axes respectively. $\mathbf{R}_1, \mathbf{R}_2$ and \mathbf{R}_3 denote the first, the second and the third column of rotation matrix \mathbf{R} .

2.3 The proposed algorithm for recovering scene dimensions

As shown in Fig. 1, consider the object of the image be modeled as a polyhedral where the coordinates of the vertices can be represented as a linear function of a dimension vector \mathbf{v} . Each coordinates of vertices is labeled from $\mathbf{P1}$ through $\mathbf{P7}$ and can be expressed as linear functions of the dimension vector $\mathbf{v}=(\mathbf{k} \quad \mathbf{w} \quad \mathbf{h})^T$ and $\mathbf{G}_j(3 \times \mathbf{N}$ matrices), where \mathbf{v} is $\mathbf{N} \times \mathbf{1}$ vector, \mathbf{N} is the number of vertices in the model and the position of the j th vertex is given by $\mathbf{G}_j \mathbf{v}$. Therefore, the coordinates of vertices can be represented as linear functions of parameterized vector as follows.

$$\mathbf{P}_1 = \begin{pmatrix} \mathbf{k}/2 \\ -\mathbf{w}/2 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0.5} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{k} \\ \mathbf{w} \\ \mathbf{h} \end{pmatrix} \mathbf{P}_2 = \begin{pmatrix} \mathbf{k}/2 \\ -\mathbf{w}/2 \\ \mathbf{h} \end{pmatrix} = \begin{pmatrix} \mathbf{0.5} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{k} \\ \mathbf{w} \\ \mathbf{h} \end{pmatrix}, \dots$$

In Fig. 1, if the positions of vertices are eight numbers, only three parameters can be characterized for the object. From this model, we present the method to calculate the

dimensions vector \mathbf{v} and the camera translation \mathbf{T} from a single image only using three vanishing points. Therefore, the vanishing points must be very accurately found by the object of model from the image of perspective projection. According to the model, the coordinates of the j th vertex in the world frame are given by $\mathbf{G}_j \mathbf{v}$. Let $\mathbf{L}_{jk} = (\mathbf{L}_{jk}^x \ \mathbf{L}_{jk}^y \ \mathbf{L}_{jk}^z)^T$ show the homogenous coordinates of the line in the image plane connecting points j and k . From this model, if a point j of vertices from the object lies on a line \mathbf{L} of the object, the line equation can be represented the following $\mathbf{L}^T \mathbf{j} = 0$. Let $\mathbf{J} = [\mathbf{X}, \mathbf{Y}, \mathbf{Z}, 1]^T$ be expressed the world coordinate of the point j . Then, we obtain the equation as follows:

$$\mathbf{L}^T \cdot \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\mathbf{J} = \mathbf{L}^T \cdot \mathbf{P}\mathbf{J} = 0$$

From the equation (2), we can represent the equation using the vanishing points as follows:

$$\begin{aligned} \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\mathbf{J} \\ = \mathbf{V}_x \mathbf{X} + \mathbf{V}_y \mathbf{Y} + \mathbf{V}_z \mathbf{Z} + \mathbf{K}\mathbf{T} \end{aligned} \quad (3)$$

Then, we get the following equation from the equation(3).

$$\begin{aligned} \mathbf{L}^T \cdot \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\mathbf{J} &= \mathbf{L}^T \cdot \mathbf{P}\mathbf{J} \\ &= \mathbf{L}^T [\mathbf{V}_x \mathbf{X} + \mathbf{V}_y \mathbf{Y} + \mathbf{V}_z \mathbf{Z} + \mathbf{K}\mathbf{T}] \\ \Rightarrow \mathbf{L}^T \begin{bmatrix} \mathbf{V}_x & \mathbf{V}_y & \mathbf{V}_z & \mathbf{K} \end{bmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \\ \mathbf{T} \end{pmatrix} &= 0 \end{aligned} \quad (4)$$

Therefore, the constraint that the projection of the j th vertex in the image using two vanishing points should lie along this line can be represented as follows :

$$\begin{aligned} \mathbf{L}_{jk}^T \mathbf{P}_j &= \mathbf{L}_{jk}^T \mathbf{P} \begin{pmatrix} \mathbf{G}_j \mathbf{v} \\ 1 \end{pmatrix} = \mathbf{L}_{jk}^T \mathbf{K}[\mathbf{R} \mid \mathbf{T}] \begin{pmatrix} \mathbf{G}_j \mathbf{v} \\ 1 \end{pmatrix} = 0 \\ \mathbf{L}_{jk}^T \mathbf{K}(\mathbf{R}\mathbf{G}_j \mathbf{v} + \mathbf{T}) &= 0 \\ \Rightarrow \mathbf{L}_{jk}^T [\mathbf{K}\mathbf{R}\mathbf{G}_j & \mathbf{K}] (\mathbf{v} \ \mathbf{T})^T \end{aligned} \quad (5)$$

Then, we can express from the equation (3) as follows:

$$\begin{aligned} \mathbf{L}_{jk}^T \mathbf{P}_j &= \mathbf{L}_{jk}^T \mathbf{K}[\mathbf{R} \mid \mathbf{T}] \begin{pmatrix} \mathbf{G}_j \mathbf{v} \\ 1 \end{pmatrix} = 0 \\ \mathbf{L}_{jk}^T \mathbf{K}(\mathbf{R}\mathbf{G}_j \mathbf{v} + \mathbf{T}) &= 0 \\ &= \mathbf{L}_{jk}^T (\mathbf{K}\mathbf{R}_1 \mathbf{G}_j \mathbf{v}_k + \mathbf{K}\mathbf{R}_2 \mathbf{G}_j \mathbf{v}_w + \mathbf{K}\mathbf{R}_3 \mathbf{G}_j \mathbf{v}_h + \mathbf{K}\mathbf{T}) \\ &= \mathbf{L}_{jk}^T (\mathbf{V}_x \mathbf{G}_j \mathbf{v}_k + \mathbf{V}_y \mathbf{G}_j \mathbf{v}_w + \mathbf{V}_z \mathbf{G}_j \mathbf{v}_h + \mathbf{K}\mathbf{T}) = 0 \\ \Rightarrow \mathbf{L}_{jk}^T [\mathbf{V}_x \mathbf{G}_j & \mathbf{V}_y \mathbf{G}_j \ \mathbf{V}_z \mathbf{G}_j \ \mathbf{K}] \begin{pmatrix} \mathbf{v} \\ \mathbf{T} \end{pmatrix} = 0 \end{aligned} \quad (6)$$

where $\mathbf{v}_k, \mathbf{v}_w$ and \mathbf{v}_h denote the parameter of the dimension vector $\mathbf{v} = (\mathbf{k} \ \mathbf{w} \ \mathbf{h})^T$.

From the equation (6), the focal length \mathbf{f} of \mathbf{K} matrix is computed by a method using three vanishing points from single image of perspective projection [9]. If the focal

length \mathbf{f} is obtained by the image, rotation matrix \mathbf{R} can be calculated by the equation (2). Then, the equation (6) can be obtained to compute the dimension vector and its pose of the model. Let \mathbf{A} be a matrix formed by stacking the rows of the form $\mathbf{L}_{jk}^T [\mathbf{v}_x \mathbf{G}_i \quad \mathbf{v}_y \mathbf{G}_i \quad \mathbf{v}_z \mathbf{G}_i \quad \mathbf{K}]$. Then, an estimate for $(\mathbf{v} \quad \mathbf{T})^T$, up to a scale factor, can be computed by finding the unit vector that minimizes

$$\|\mathbf{A}(\mathbf{v} \quad \mathbf{T})^T\|^2$$

This can be obtained by the constrained optimization method using standard techniques from linear algebra [6]. The dimension vector \mathbf{v} of objects can be therefore solved by the standard nonlinear optimization techniques with a multistart method which generates multiple starting points for the optimizer by sampling the parameter space uniformly.

2.4 Finding vanishing points

An input image used the object of primitives from single image of perspective projection model is constrained information that can be obtained with the image because of considering only the image of perspective projection using three vanishing points with a camera of unknown focal length. Therefore, we should accurately extract the information of the vanishing point to have orthogonal property from the image. From the theory, many systems using vanishing points from image have been proposed [4]. As shown in Fig.2, the vanishing point corresponds to the projection of the intersection of parallel lines at infinity in the \mathbf{X} , \mathbf{Y} and \mathbf{Z} directions from the model of an image. To very accurately explore the vanishing point, the user defines interactively lines of the model and finds correspondences between edges in the model and edges in the image by selecting a line in the model. Also, to compute the vanishing points of \mathbf{X} , \mathbf{Y} and \mathbf{Z} directions respectively, the user selects n lines in the model that are parallel to the \mathbf{X} , \mathbf{Y} and \mathbf{Z} axis of the object [5]. Let $\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_n$ be a set of image lines corresponding to parallel to the \mathbf{X} axis and let \mathbf{v}_x be the vanishing points defined by these lines of the \mathbf{X} direction. Then, the homogeneous coordinates of the vanishing points in the \mathbf{X} , \mathbf{Y} and \mathbf{Z} directions are the vectors $\mathbf{V}_x, \mathbf{V}_y, \mathbf{V}_z$ that minimize $\sum \|\mathbf{L}_i^T \mathbf{v}_x\|^2, \sum \|\mathbf{L}_i^T \mathbf{v}_y\|^2, \sum \|\mathbf{L}_i^T \mathbf{v}_z\|^2$. These vectors can be computed by eigenvalue decomposition of $\mathbf{D}^T \mathbf{D}$, where \mathbf{D} is the matrix whose rows consist of the \mathbf{L}_i^T 's, using a constrained optimization from linear algebra [6]. This technique is adapted to compute the vanishing points of the \mathbf{X} , \mathbf{Y} and \mathbf{Z} directions.

3 Experimental Results

According to the constraint to use the vanishing points in the perspective projection model, the experimental images are selected by single images of objects being three vanishing points from photographs taken with Canon EOS D30 digital camera. All of the images are 815 x 644 images and objects with weak perspective.

To evaluate the performance in this experiment, we use the well-known RMS (Root Mean Square) error method. We also compare the proposed algorithm with the conventional method [3] used two vanishing points. This is due to be derived our algorithm from the method exploiting two vanishing points.

For sample images M1-4 shown in Fig. 3, we have computed the dimension vector \mathbf{v} for the proposed algorithm and the conventional method. The dimension vector of M1-2 is therefore $N = 3$, an image of simple object and M3-4 are $N = 6$, the image of two blocks. As shown in Table 1, the dimension vector \mathbf{v} , which gives the dimensions of the boxes, were measured by hand and found to be $(\text{in cm})(20 \ 20 \ 20)^T(\text{M1})$, $(30 \ 20 \ 10)^T(\text{M2})$, $(20 \ 30 \ 10 \ 30 \ 20 \ 7)^T(\text{M3})$ and $(20 \ 30 \ 7 \ 20 \ 20 \ 20)^T(\text{M4})$. After choosing an appropriate scaling factor, the conventional method and the proposed algorithm are represented in Table 1. Note that we cannot check the accuracy the pose estimation because of no having a truth model of these parameters. For evaluating the performance of sample images M1-4, the conventional method shows RMS errors of 0.98cm (M1), 0.98cm (M2), 1.2cm (M3) and 1.2cm (M4). The proposed algorithm denotes RMS errors of 0.83cm (M1), 0.91cm (M2), 1.9 cm (M3) and 1.9cm (M4). According to using the vanishing points, the proposed algorithm for sample images of M1 and M2 represents better results than the conventional. The conventional method for sample images of M3 and M4 shows better results than the proposed. Therefore, if the objects of single image are situated to a center of the image, the results will show the performance of the proposed method is well better than the conventional. To recover 3D reconstruction of a photograph using the proposed algorithm based on the conventional system, we use the Art Center in Chung-Ang University (Fig.4a). Fig.4b shows texture-mapped 3D reconstruction of the Art Center in Chung-Ang University.

4 Conclusions

In this paper, we have presented the method using only three vanishing points to calculate the dimension vector \mathbf{v} of object from single image of perspective projection with a camera of unknown focal length. To accurately compute the dimensions of object, the user directly selects the vanishing points from the image in the paper. Experimental results show that our method have the accuracy and efficacy of the dimension vector in comparison with the actual images.

Future work has been described the use of automated edge extraction in order to calculate the vanishing points.

Acknowledgment

This research was supported by the Ministry of Education, Seoul, Korea, under the BK21 project, and by the Ministry of Science and Technology, Seoul, Korea, under the NRL project.

References

1. W. G. Kropatsch and H. Bischof : Digital Image Analysis, Springer, 2001.
2. R P.E DEBEVEC, C.J TAYLOR, AND J. MALIK: Modeling and Rendering Architecture from Photographs: A Hybrid Geometry and Image-Based Approach, In Proceedings of ACM SIGGRAPH 1996, ACM Press / ACM SIGGRAPH, pp.11-21, Aug. 1996.
3. D. JELINEK AND C.J. TAYLOR: Reconstruction of linearly parameterized models from single images with a camera of unknown focal length, IEEE PAMI vol.23, no.7 July 2001.
4. B. Caprile and V. Torre: Using vanishing points for camera calibration, IJCV, vol. 4, pp. 127-140, 1990.
5. R.T Collins and R.S Weiss: Vanishing point calculation as a statistical inference on the unit sphere, In Proc.Third Int. Conference on Computer Vision, pp. 400-403, Dec.1990.
6. O. Faugeras.: Three-dimensional computer vision, MIT Press, 1993.
7. D.Liebowitz A. Criminisi and A.Zisserman: Creating architectural models from image, In Proc. Eurographics, Sep 1999.

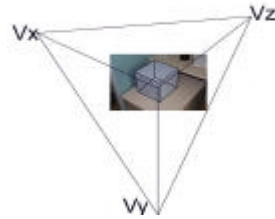
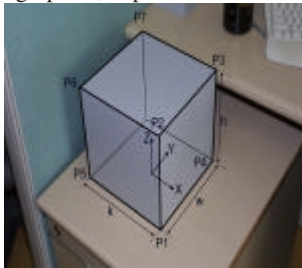
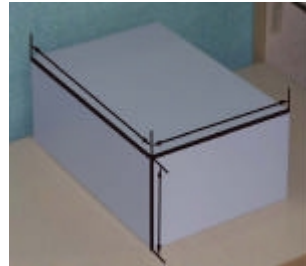


Fig.1.An example of linearly parameterized box model.**Fig.2.** Each vanishing points for the model of X, Y, and Z directions.



M1



M2



M3



M4

Fig.3. Sample images (M1-4)



Fig.4a. The Art Center in Chung-Ang University



Fig.4b. Texture mapped 3D reconstruction of the Art Center.

Sample images	Measurement by hand	Conventional method		Proposal method	
		The dimension vector \mathbf{v}	RMS error	The dimension vector \mathbf{v}	RMS error
M1	$(20\ 20\ 20)^T$	$(19.2\ 20.9\ 18.8)^T$	0.98	$(20.3\ 20.9\ 18.9)^T$	0.83
M2	$(30\ 20\ 10)^T$	$(31.2\ 21.1\ 9.5)^T$	0.98	$(28.5\ 20.4\ 10.3)^T$	0.91
M3	$(20\ 30\ 10\ 30\ 20\ 7)^T$	$(22.1\ 31.2\ 9.5\ 30.6\ 19.5\ 8.1)^T$	1.2	$(22.3\ 28.1\ 12.3\ 28.3\ 18.1\ 9.3)^T$	1.9
M4	$(20\ 30\ 7\ 20\ 20)^T$	$(19.5\ 30.6\ 8.1\ 18.6\ 21.8\ 18.8)^T$	1.2	$(21\ 28.3\ 9.3\ 22.3\ 22.2\ 18.5)^T$	1.9

Table1. The results of the dimensions vector \mathbf{v} for the images.
(A measure of length: cm)