Shape from Regularities for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images

Zhenguo Li[1], Jianzhuang Liu[1], and Xiaoou Tang[1,2]
[1]Dept. of Information Engineering, The Chinese University of Hong Kong
[2]Visual Computing Group, Microsoft Research Asia, Beijing, China
{zgli5, jzliu}@ie.cuhk.edu.hk, xitang@microsoft.com

ABSTRACT

3D object reconstruction from single 2D images has many applications in multimedia. This paper proposes an approach based on image regularities such as connectivity, parallelism, and orthogonality possessed by the objects with simple user interactions. It is assumed that the objects are piecewise planar. By representing the 3D objects as a shape vector consisting of the normals of the faces of the objects, we impose geometric constraints on this shape vector using the regularities of the objects. We derive a system of equations in terms of the shape vector and the focal length, which we can solve for the shape vector optimally. Experimental results on real images are shown to demonstrate the effectiveness of this method.

Categories and Subject Descriptors: I.4.5 [Image Processing and Computer Vision]: Reconstruction–Transform methods

General Terms: Algorithms, Theory

Keywords: Interactive 3D reconstruction, image regularities, shape vector

1. INTRODUCTION

3D object reconstruction from images is an important topic in computer vision, and has many applications such as user-friendly query interface for 3D object retrieval [5] and 3D photo-realistic scene creation for game, movie, and webpage design. The tremendous effort devoted to this area has shown that inference of 3D shapes from 2D images is very challenging and generally underconstrained, even from multiple views of a scene or an object. Extensive research has been done on 3D reconstruction from multiple views [1].

In most cases, however, we have only one view of an object or a scene. 3D object reconstruction from one single view is obviously a harder problem because less information is kept in one view than in multiple views. Although this problem is usually ill-posed, the human visual system can easily perceive the 3D shapes of the objects in an image based on the knowledge learnt. In this paper, we focus on 3D reconstruction of piecewise planar objects from single images. These objects are often seen in daily life. In our system, the user draws the edges of the objects in an image and provides some basic image regularity information (line parallelism, line orthogonality, face parallelism, and face orthogonality) to the system; then the algorithm recovers the 3D shapes of the objects by combining the regularities and solving an optimization problem.

Many methods have been proposed for 3D reconstruction from single images, such as those in [9], [6], [3], [4], [8], [2] and [10]. Zhang et al. [9] tried to reconstruct free-form 3D model from a single image. This method needs a lot of user interaction and may take more than one hour for the user to specify constraints from one image. Prasad et al. [6] recovered a curved 3D model from its silhouettes in an image. This algorithm is a development of that in [9] and does not need so such interaction, but the reconstructed objects are restricted to those whose silhouettes are contour generators where the surface normals are known. The Facade system [3] models a 3D building using parametric primitives (blocks) from a single view or multiple views of the scene. In [4], Liebowitz et al. created architectural models using geometric relationships from the architectural scenes. Their method requires to rectify two or more planes and to compute the vanishing lines of all the planar faces. Besides, the reconstruction errors may accumulate. Sturm and Maybank's method [8] first does camera calibration, then recovers part of points and planes by assuming the availability of some vanishing points and lines, and finally obtains the 3D positions of other faces by propagation. The drawbacks of this method are that the parts of the objects have to be sufficiently interconnected and the reconstruction errors may be accumulated. Jelinek and Taylor [2] proposed a method of polyhedron reconstruction from single images using several camera models. The main restriction of this method is that the polyhedra have to be linearly parameterized, which limits the application of the method. Shimodaira [10] used the information of shading, one horizontal or vertical face, and convex and concave edges to recover the shape of polyhedra in single images. This method can only handle very simple polyhedra.

The common point in 3D reconstruction from single images is that the user's interaction is a necessary step because current computer vision algorithms still cannot extract only the required edges from general images. Compared with the related methods for planar object reconstruction, ours has several contributions: (i) the whole shape of the objects in an image is formulated by a shape vector which consists of the normals of all the planar faces, and the shape (up to a scale) is determined simultaneously by solving a system of equations optimally. Our method does not accumulate reconstruction errors. (ii) The focal length does not need to be found before the reconstruction. It is determined together

MM'06, October 23–27, 2006, Santa Barbara, California, USA.
ACM 1-59593-447-2/06/0010.

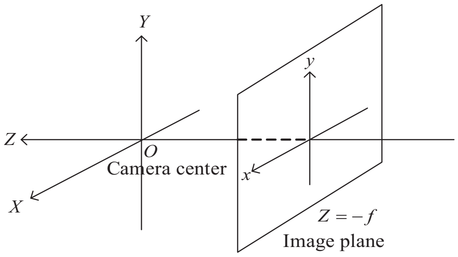

Figure 1: Camera model and coordinate system.

with the shape vector by the optimization. This avoids the problem that a focal length that is not estimated precisely can cause large reconstruction errors. (iii) If the user also provides the hidden edges during the interaction, our algorithm can recover both the visible and invisible shapes of the objects.

2. THE IMAGING MODEL AND SHAPE VECTOR

Homogeneous coordinates are used for the analysis of the problem in this paper unless Euclidean coordinates are specified somewhere. Besides, due to usually limited cues presented in a single image, we assume a simplified camera model with a calibration matrix $\mathbf{K}$ given by

$$\mathbf{K} = \begin{pmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad (1)$$

where f is the focal length. In this model, the only parameter needs to be found is f. Furthermore, without loss of generality we align the world frame with the camera frame as shown in Fig. 1, where the image plane is $Z = -f$, and the projection matrix takes a simple form $\mathbf{P} = [\mathbf{K}|\mathbf{0}]$. The relation between a point $\mathbf{M} = (X, Y, Z, 1)^T$ in the world frame and its projection $\mathbf{m} = (x, y, 1)^T$ in the image is $\lambda\mathbf{m} = \mathbf{PM}$, from which we have $\lambda = Z$, $X = -Zx/f$, and $Y = -Zy/f$.

A scene is said to be projected from a *generic view* if perceptual properties in the image are preserved under slight variations of the viewpoint. We suppose that the objects in an image are the projection in a generic view where no face is projected to a line. Let $\boldsymbol{\pi} = (a, b, c, d)^T$ denote a plane $ax+by+cz+d = 0$ in 3D space. Then we have the following property.

PROPERTY 1. *A plane $\boldsymbol{\pi} = (a, b, c, d)^T$ in a generic view satisfies that $d \neq 0$.*

PROOF. Assume, to the contrary, that $d = 0$. Then the camera center, $\mathbf{v} = (0, 0, 0, 1)^T$, satisfies $\boldsymbol{\pi}^T\mathbf{v} = d = 0$, meaning that $\boldsymbol{\pi}$ is passing through $\mathbf{v}$ and projected as a line, which contradicts the assumption that $\boldsymbol{\pi}$ is in a generic view. □

According to Property 1, a plane in a generic view can be written as $\boldsymbol{\pi} = (a, b, c, 1)^T = (\mathbf{R}^T, 1)^T$, where $\mathbf{R} = (a, b, c)^T$ is the normal of the plane. Since we are dealing with planar objects consisting of faces, $\mathbf{R}$ is also called the normal of a face that is passed through by $\boldsymbol{\pi}$. We represent the objects of interest in a scene with a vector consisting of all the normals of the faces of the objects, i.e., $\mathbf{q} = (\mathbf{R}_1^T, \mathbf{R}_2^T, ..., \mathbf{R}_{N_f}^T)^T$, where N_f is the number of faces. We call $\mathbf{q}$ the *shape vector* of the objects. In the following, we will impose geometric constraints on this shape vector which are extracted from the regularities in the image.

3. FORMULATING IMAGE REGULARITIES

Mathematically, 3D objects are characterized by topological and geometric properties. In this paper, the topological property we use is connectivity, and the geometric properties are parallelism and orthogonality. These properties are also called image regularities.

3.1 Connectivity

In a planar object, a vertex is often shared by more than one face. This connectivity leads to constraints that relate the normals of these faces through the vertex.

Let $\mathbf{x} = (x, y, 1)^T$ be the imaged vertex of $\mathbf{X} = (X, Y, Z, 1)^T$ which lies on both the ith face (plane) $\boldsymbol{\pi}_i = (a_i, b_i, c_i, 1)^T = (\mathbf{R}_i^T, 1)^T$ and the jth face (plane) $\boldsymbol{\pi}_j = (a_j, b_j, c_j, 1)^T = (\mathbf{R}_j^T, 1)^T$. Then, $\lambda\mathbf{x} = \mathbf{PX}$, $\boldsymbol{\pi}_i^T\mathbf{X} = 0$, and $\boldsymbol{\pi}_j^T\mathbf{X} = 0$, where λ is some non-zero scalar. From these relations, we have

$$\mathbf{R}_i^T\mathbf{x}' = \mathbf{R}_j^T\mathbf{x}' = f/Z, \qquad (2)$$

where $\mathbf{x}' = (x, y, -f)^T$ is the 3D Euclidean coordinate of $\mathbf{x}$. Furthermore, we have

$$(\mathbf{x}'^T, -\mathbf{x}'^T) \cdot (\mathbf{R}_i^T, \mathbf{R}_j^T)^T = 0. \qquad (3)$$

The constraint in (3) is termed the shared vertex constraint. Similarly, if the vertex $\mathbf{x}'$ is shared by n faces, $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, ..., \boldsymbol{\pi}_n$, then we have $n - 1$ independent constraints:

$$(\mathbf{x}'^T, -\mathbf{x}'^T) \cdot (\mathbf{R}_k^T, \mathbf{R}_{k+1}^T)^T = 0, \quad k = 1, 2, ..., n-1. \qquad (4)$$

In terms of the shape vector $\mathbf{q}$, (4) can be written in matrix form:

$$\mathbf{A}_1\mathbf{q} = \mathbf{0}, \qquad (5)$$

where $\mathbf{A}_1 =$

$$\begin{pmatrix} \mathbf{x}'^T & -\mathbf{x}'^T & \mathbf{0}'^T & \mathbf{0}'^T & \cdots & & & & \\ \mathbf{0}'^T & \mathbf{x}'^T & -\mathbf{x}'^T & \mathbf{0}'^T & \cdots & & & & \\ & & & & \cdots & & & & \\ \mathbf{0}'^T & \mathbf{0}'^T & \mathbf{0}'^T & \mathbf{0}'^T & \cdots & \mathbf{x}'^T & -\mathbf{x}'^T & \mathbf{0}'^T & \cdots \end{pmatrix},$$

$\mathbf{0}' = (0, 0, 0)^T$, and $\mathbf{q} = (\mathbf{R}_1^T, \mathbf{R}_2^T, ..., \mathbf{R}_n^T, ..., \mathbf{R}_{N_f}^T)^T$. $\mathbf{A}_1$ is a $(n-1) \times (3N_f)$ matrix where in the last row, from the first $\mathbf{0}'^T$ to $-\mathbf{x}'^T$, there are totally $3n$ elements. This is the shared vertex constraint contributed by one image vertex $\mathbf{x}'$ shared by the n faces. Putting the constraints from all the shared vertices together, we have

$$\mathbf{Aq} = \mathbf{0}, \qquad (6)$$

where $\mathbf{A} = (\mathbf{A}_1^T, \mathbf{A}_2^T, ..., \mathbf{A}_{N_v}^T)^T$, N_v is the number of all the vertices of the objects. With (6) only, we cannot recover the 3D objects because usually there are infinite 3D objects that project to the same image. To obtain desired 3D reconstruction, we need to use more regularities.

3.2 Line Parallelism

Suppose that $\mathbf{L}_1$ and $\mathbf{L}_2$ are two 3D parallel lines whose projections are $\mathbf{l}_1$ and $\mathbf{l}_2$ in the image, respectively. Then we can infer the direction of $\mathbf{L}_1$ and $\mathbf{L}_2$, with which a constraint on the normal of the face parallel to $\mathbf{L}_1$ and $\mathbf{L}_2$

can be imposed. In what follows, $\mathbf{l}_i$ representing a line $\alpha_i x + \beta_i y + \gamma_i = 0$ in the image is also defined as such a vector $\mathbf{l}_i = (\alpha_i, \beta_i, \gamma_i)^T$.

PROPERTY 2. *If two lines, $\mathbf{L}_1$ and $\mathbf{L}_2$, and a plane are parallel in 3D space, where $\mathbf{l}_1$ and $\mathbf{l}_2$ are the projections of $\mathbf{L}_1$ and $\mathbf{L}_2$ in the image, respectively, then the normal $\mathbf{R}$ of this plane satisfies*

$$\mathbf{R}^T(\mathbf{K}^{-1}(\mathbf{l}_1 \times \mathbf{l}_2)) = 0. \quad (7)$$

PROOF. The direction $\mathbf{D}$ of $\mathbf{L}_1$ satisfies a well-known relation $\mathbf{D} = \mathbf{K}^{-1}\mathbf{v}$, where $\mathbf{v}$ is the vanishing point of $\mathbf{L}_1$ [1]. Since $\mathbf{L}_1$ and $\mathbf{L}_2$ are parallel, we have $\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2$ [1]. Since $\mathbf{L}_1$ and the plane are also parallel, we further have $\mathbf{R} \perp \mathbf{D}$, and thus $0 = \mathbf{R}^T\mathbf{D} = \mathbf{R}^T(\mathbf{K}^{-1}(\mathbf{l}_1 \times \mathbf{l}_2))$. □

3.3 Line Orthogonality

The following property gives a constraint relating two perpendicular lines and the two planes passing through the two lines.

PROPERTY 3. *If line $\mathbf{L}_1$ lies on plane $\boldsymbol{\pi}_1$ and $\mathbf{L}_2$ lies on plane $\boldsymbol{\pi}_2$ such that $\mathbf{L}_1$ is perpendicular to $\mathbf{L}_2$ in 3D space, then we have*

$$(\mathbf{R}_1 \times (\mathbf{K}^T\mathbf{l}_1))^T(\mathbf{R}_2 \times (\mathbf{K}^T\mathbf{l}_2)) = 0, \quad (8)$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are the normals of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$, respectively, and $\mathbf{l}_1$ and $\mathbf{l}_2$ are the projections of $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively.

PROOF. Let $\mathbf{N}_1$ ($\mathbf{N}_2$) be the normal of the plane $\boldsymbol{\pi}'_1$ ($\boldsymbol{\pi}'_2$) that passes through $\mathbf{l}_1$ ($\mathbf{l}_2$) and the camera center, and let $\mathbf{D}_1$ ($\mathbf{D}_2$) be the direction of $\mathbf{L}_1$ ($\mathbf{L}_2$). Then $\mathbf{N}_1 = \mathbf{K}^T\mathbf{l}_1$ and $\mathbf{N}_2 = \mathbf{K}^T\mathbf{l}_2$ [1]. Since $\mathbf{L}_1$ lies on both $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}'_1$, we have $\mathbf{D}_1 \perp \mathbf{R}_1$ and $\mathbf{D}_1 \perp \mathbf{N}_1$, thus $\mathbf{D}_1 = \mathbf{R}_1 \times \mathbf{N}_1 = \mathbf{R}_1 \times (\mathbf{K}^T\mathbf{l}_1)$. Similarly, $\mathbf{D}_2 = \mathbf{R}_2 \times (\mathbf{K}^T\mathbf{l}_2)$. From $\mathbf{D}_1 \perp \mathbf{D}_2$, we have (8) immediately. □

3.4 Face Parallelism

If face $\boldsymbol{\pi}_1$ and face $\boldsymbol{\pi}_2$ are parallel, with $\mathbf{R}_1 = (a_1, b_1, c_1)^T$ and $\mathbf{R}_2 = (a_2, b_2, c_2)^T$ being the corresponding normals, then $\mathbf{R}_1 \times \mathbf{R}_2 = \mathbf{0}$, which leads to two independent quadratic constraints:

$$b_1c_2 - b_2c_1 = 0, \quad a_1c_2 - a_2c_1 = 0. \quad (9)$$

3.5 Face Orthogonality

If face $\boldsymbol{\pi}_1$ is perpendicular to face $\boldsymbol{\pi}_2$, with $\mathbf{R}_1 = (a_1, b_1, c_1)^T$ and $\mathbf{R}_2 = (a_2, b_2, c_2)^T$ being the corresponding normals, then $\mathbf{R}_1^T\mathbf{R}_2 = 0$, which leads to one quadratic constraint:

$$a_1a_2 + b_1b_2 + c_1c_2 = 0. \quad (10)$$

4. IMPLEMENTATION

The following steps show how to implement the 3D reconstruction where how to find the shape vector is the key.

1. Draw manually the edges of the desired objects in an image.

2. Use the algorithm presented in [11] to find the faces of the objects.

3. Provide by the user the information of image regularities exhibiting in the objects, i.e., the lines and/or faces that are parallel and/or orthogonal.

4. Form a set of equations based on these regularities:

$$\mathbf{Aq} = \mathbf{0}, \quad (11)$$
$$g_i(f, \mathbf{q}) = 0, \quad i = 1, 2, ..., m, \quad (12)$$

   where (12) contains all the constraints stated in (7)–(10).

5. Set Z_0 to be the depth (Z-coordinate) of one vertex $\mathbf{x}_0 = (x_0, y_0)$ which is on one face with normal $\mathbf{R}_0$, then by (2) we have

$$\mathbf{R}_0\mathbf{x}'_0 - f/Z_0 = 0, \quad (13)$$

   where $\mathbf{x}'_0 = (x_0, y_0, -f)^T$ is the 3D Euclidean coordinate of $\mathbf{x}_0$ in the image plane.

6. Solve the following optimization problem for $\mathbf{q}$ and f:

$$\min : |\mathbf{Aq}|^2 + |\mathbf{R}_0\mathbf{x}'_0 - f/Z_0|^2 + \sum_{i=1}^{n} |g_i(f, \mathbf{q})|^2, \quad (14)$$

$$\text{s.t.} : \; -1 < \mathbf{R}_j^T\mathbf{x}'_{jk} < 0, \quad j = 1, 2, ..., N_f, \quad k = 1, 2, ..., n_j, \quad (15)$$

   where N_f is the number of all the faces, and $\mathbf{x}'_{jk} = (x_{jk}, y_{jk}, -f)^T, k = 1, 2, ..., n_j$, are the 3D Euclidean coordinates of the vertices (in the image plane) of face j that has n_j vertices.

7. For any vertex (X, Y, Z) (with $\mathbf{x}' = (x, y, -f)^T$ as its image in the image plane), compute all the Z-coordinate Z_i by $Z_i = f/(\mathbf{R}_i^T\mathbf{x}')$ (see (2)) if it is on ith face (with normal $\mathbf{R}_i$). Take the average as the final Z-coordinate Z if there are multiple Z_i. Compute the X-coordinate and Y-coordinate by $X = -Zx/f$, $Y = -Zy/f$.

It is necessary to explain why the constraints in (15) are imposed. From Fig. 1, we see that all the vertices of the 3D objects should be on the right-hand side of the image plane, which means $0 < f < -Z$, where Z denotes the Z-coordinate of any vertex in the world frame. Thus from (2), we have (15). The optimization problem in (14) and (15) can be solved efficiently using the function *lsqnonlin* in MATLAB 7.1.

5. EXPERIMENTAL RESULTS

We have conducted a number of experiments on real images to verify the effectiveness of our approach. Due to the space limitation, only part of them are given here. In Fig. 2, (a1)–(f1) are the original images with the drawn red lines superimposed on the edges of the objects, and (a2)–(f2) and (a3)–(f3) are the reconstructed 3D objects with texture mapped, with each result shown in two different views. In Figs. 2(a1)–(d1), no hidden edges are drawn. In Fig. 2(e1), all the hidden edges are drawn which are guessed by the user. Fig. 2(e3) shows the backs of buildings. In Fig. 2(f1), part of the hidden edges of the buildings are drawn. These results clearly demonstrate that our method successfully creates the desired 3D objects from the images.

6. CONCLUSIONS

We have proposed an approach to reconstructing 3D piecewise planar objects from single images based on the drawn edges and image regularities provided by the user. We first represent the objects in an image as a shape vector, then

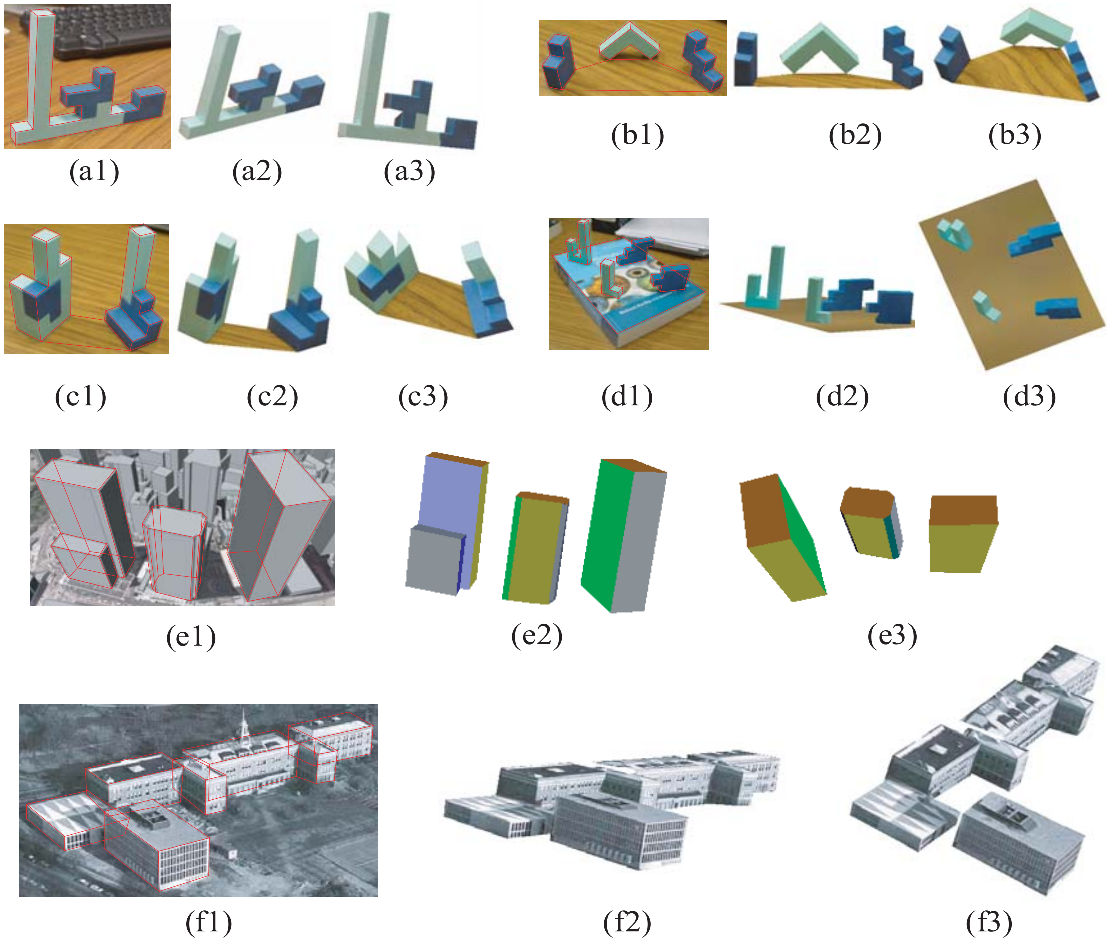

Figure 2: (a1)–(f1) Original images with drawn edges (red lines) superimposed. (a2)–(f2) One view for each reconstructed result. (a3)–(f3) Another view for each reconstructed result.

formulate the regularities as geometric constraints, and finally solve an optimization problem to obtain the optimal shape vector. Compared with the related methods, ours has several advantages: (i) the whole shape of the objects (up to a scale) in an image is determined simultaneously, and thus our method does not accumulate reconstruction errors. (ii) The focal length is determined together with the shape vector by the optimization. This avoids the problem that a focal length that is not estimated precisely can cause large reconstruction errors. (iii) If the user also provides the hidden edges during the interaction, our algorithm can recover both the visible and invisible shapes of the objects.

7. ACKNOWLEDGMENTS

This work was supported by the Research Grants Council of the Hong Kong SAR (Project No. CUHK 414306) and the CUHK Direct Grant.

8. REFERENCES

[1] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, second edition, 2004.

[2] D. Jelinek and C. J. Taylor. Reconstruction of linearly parameterized models from single images with a camera of unknown focal length. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 767–773, 2001.

[3] P. E. Debevec, C.J. Taylor and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. *Proc. SIGGRAPH*, pp. 11-20, 1996.

[4] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. *Proc. EuroGraphics*, vol. 18, pp. 39–50, Sept. 1999.

[5] L. L. Cao, J. Z. Liu, and X. Tang. 3D object retrieval using 2D line drawing and graph based relevance feedback. *Proc. ACM Int. Conf. on Multimedia*, Oct. 2006.

[6] M. Prasad, A. Zisserman, and A. W. Fitzgibbon. Fast and controllable 3D modelling from silhouettes. *Proc. of the 26th Annual Conference of the European Association for Graphics, Dublin*, pp. 9–12, Sept. 2005.

[7] H. Y. Shum, M. Han, and R. Szeliski. Interactive construction of 3D models from panoramic mosaics. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 427-433, 1998.

[8] P. F. Sturm and S. J. Maybank. A method for interactive 3d reconstruction of piecewise planar objects from single images. *Proc. British Machine Vision Conference*, 1999.

[9] L. Zhang, G. Dugas. Phocion, J. S. Samson, and S. M. Seitz. Single view modeling of free-form scenes. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 990–997, 2001.

[10] H. Shimodaira. A shape-from-shading method of polyhedral objects using prior information. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 612–624, 2006.

[11] J. Z. Liu and Y. T. Lee Graph-based method for face identification from a single 2D line drawing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1106–1119, 2001.