Efficient Object Shape Recovery via Slicing Planes

Po-Lun Lai and Alper Yilmaz
Photogrammetric Computer Vision Lab, Ohio State University
233 Bolz Hall, 2036 Neil Ave., Columbus, OH 43210, USA
http://dpl.ceegs.ohio-state.edu/

Abstract

Recovering the three-dimensional (3D) object shape remains an unresolved area of research on the cross-section of computer vision, photogrammetry and bioinformatics. Although various techniques have been developed, the computational complexity and the constraints introduced to overcome the problems have limited their applicability in the real world scenarios. In this paper, we propose a method that is based on the projective geometry between the object space and the silhouette-images taken from multiple viewpoints. The approach eliminates the problems related to dense feature point matching and camera calibration that are generally adopted by many state of the art shape reconstruction methods. The object shape is reconstructed by establishing a set of hypothetical planes slicing the object volume and estimating the projective geometric relations between the images of these planes. The experimental results show that the 3D object shape can be recovered by applying minimal constraints.

1. Introduction

The growing demand for using 3D models on visualization, city planning and scene analysis makes the 3D object shape recovery a prevailing area of research in the field of Computer Vision. Numerous research efforts have been extended to recover the 3D object shape from images. The most common approach requires calibrating the camera prior to the 3D shape recovery. An intuitive approach to calibrate cameras is to use a set of matching features across different views of a calibrating box placed in the object space. Alternatively, the calibration can be accomplished by exploiting the projective geometry which relates different views of a scene to each other [4] [5] [7]. Once the cameras are calibrated, both approaches reconstruct the object shape by backprojecting the image points into the object space by triangulation [1] [11] and the 3D shape is recovered up to an unknown scale factor.

While calibrating cameras is a desirable procedure, it is usually not an intuitive one. Adding the limitation of observing limited number of features on the objects limits the applicability of these methods in scenarios when the images are of poor quality and perspectively distorted, and the objects are fully or partially occluded. These limitations suggest the development of methods which eliminate the requirement of establishing point correspondences and calibrating cameras.

In the recent years, another line of 3D shape recovery research, which exploits the silhouette images and visual-hull formation, has emerged. These methods exploit the fact that the shape information provided by the object silhouettes is sufficient for reconstructing the object shape without establishing point matches [3] [13]. However, these methods still require camera calibration. The camera calibration parameters are used for backprojecting each silhouette from the image space to the object space which forms a volume. The intersections of these volumes define the convex-hull; hence the object shape.

In this paper, we leverage the state of the art in shape reconstruction which adopts the use of silhouettes while eliminating the requirement of camera calibration. The proposed method utilizes the concept of slicing planes which is inspired by the 3D affine recovery technique discussed in [10]. The slicing planes are considered hypothetical planes which are parallel to each other and to a reference plane in the object space. The analytic relation between the images is derived from the homography of the reference planes.

The homography transform provides a strong geometric constraint and, in contrast to fundamental matrix, provides direct mapping across images [6]. The implied 3D scene information through the reference plane and its homography transform in the images have been used in various applications, including but not limited to the tracking of people [9], shadow removal [2], and detection of low-lying objects [8]. Most of these techniques have been conceptually proven to have robust performance in practical scenarios.

In the proposed approach, the homography transform of the reference plane across the images is combined with a minimal number of additional observations related to the

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

view geometry of the slicing planes for "metric recovery" of the object shape. The analytical relations of the slicing planes are derived from the use of several geometric properties. The intersections of the object silhouettes in different views and their mapping to a reference view provides the metric recovery of the object shape.

The merit of the proposed approach can be described in terms of efficiency, flexibility and practicability:

Efficiency: The proposed method does not require camera calibration and estimation of the fundamental matrix; hence, the computational complexity related to establishing abundant number of point correspondences is reduced. The object shape is reconstructed using the apparent contours of the projected silhouettes, which provide the surface of the 3D object without the necessity of estimating visual hulls.

Flexibility: The level of detail in the reconstructed object is dependent on the number of images used or the distance between the slicing planes, which provide a balance between the computation time and the smoothness of the recovered shape. Furthermore, since no dense point correspondences are required and the missing information can be recovered from other images from different viewpoints, the use of object silhouettes automatically eliminates the problems related to occlusion.

Practicability: The set of simple equations involved in our approach are computed in real time, providing real time metric recovery of the complete scene for use in different problem domains such as scene understanding and object tracking. By using techniques for finding points on the reference plane [12], and generating the silhouettes [14], the proposed approach achieves automated reconstruction. In addition, the 3D shapes of all the objects in the scene are recovered simultaneously as the slicing planes cuts all the object volumes in the object space.

The paper is organized as follows. In order to better manifest the core techniques adopted by the proposed method, we briefly review two important concepts of projective geometry in Section 2.1. The proposed approach for estimating equations of the slicing planes is described in Section 2.2. In Section 2.3 the techniques for recovering the 3D object shape using the slices are delineated. Four sets of experiments discussed in Section 3 are conducted to verify the applicability of the approach in close-range and aerial images. Finally, we conclude the paper in Section 4.

2. 3D Object Shape Recovery

2.1. Projective Geometry

The projective geometry describes the physical characteristics of the cameras and the relationships between the images. The projection of a point $\mathbf{X}_w$ in the object space to a point $\mathbf{x}_i$ in the image space using a projective camera is expressed in terms of a direct linear mapping in the homogeneous coordinates as:

$$\lambda\mathbf{x} = \mathrm{P}\mathbf{X}_w = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \qquad (1)$$

where λ is the scale factor due to the projective equivalency of $(kx,\ ky, k) = (x, y, 1)$, P is a 3×4 camera projection matrix and $\mathbf{p}_i$ the i^{th} column of P. Note that, throughout the paper, we use homogeneous coordinates with the last component set to be one for points both in the image and object spaces.

When the point in the object space lies on the ground plane such that $Z = 0$, the linear mapping given in (1) will reduce to the planar homography

$$s\mathbf{x} = \mathrm{H}\mathbf{X}_w^{'} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \qquad (2)$$

where H is the homography matrix, which is a direct mapping of the points lying on a plane in the object space across different images. This formulation introduces another scaling factor, s, to the mapping equation which stems from $Z = 0$. While equation (2) is defined for $Z = 0$, the same relation can be derived for any other plane in the object space. Hence, without the loss of generality, we will use $Z = 0$ in the remainder of the paper.

In the case when we have multiple images of a scene, an intuitive consequence of the homography transform from the ground plane to the image is the existence of a direct linear mapping between the two images:

$$\mathbf{x}_i = \mathrm{H}_{wi}\mathbf{X}_w^{'} = \mathrm{H}_{wi}(\mathrm{H}_{wj}^{-1}\mathbf{x}_j) = \mathrm{H}_{ji}\mathbf{x}_j, \qquad (3)$$

where H_{ji} is the homography matrix describing the projective transformation of the pixels lying on image planes i and j. The estimation of this transformation up to a scale factor requires a minimum of four points lying on the plane. When warping one image onto the other, only the pixels in the area that map the ground plane coincide, while other pixels will create discrepancies depending on their Euclidean distances to the plane π. We should state that it is these discrepancies, or the so called shadow projections, that let us estimate the 3D shape of the object.

Another important concept in the projective geometry is the vanishing point. Considering a pair of parallel lines $\mathbf{L}_1$ and $\mathbf{L}_2$ in the object space, their intersection is defined to lie at the infinity which is represented by $[X, Y, Z, 0]^T$. The intersection point of these parallel lines visible in the image plane is referred to as the vanishing point $\mathbf{v}$. The vanishing point can be computed from the cross product of the corresponding line pair $\mathbf{l}_1$, $\mathbf{l}_2$ in the image space (see Figure 1) as $\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2$. The projections of all parallel lines with the

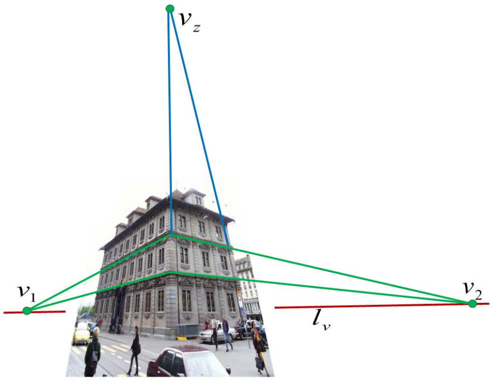


Figure 1. A parallel line pair which is parallel to the ground plane as well, when projected to the image plane, intersects at a single point referred to as the vanishing point. The cross product of two such points provides the vanishing line $\mathbf{l}_v$. The vanishing point of lines in the direction of ground plane normal is denoted as $\mathbf{v}_z$. The 3D shape can be recovered by exploiting the information gathered from $\mathbf{l}_v$ and $\mathbf{v}_z$.

same directions intersect at one single point in the image plane. A similar observation can also be made for parallel planes in the object space. Particularly, the parallel planes intersect at the pencil of planes which resides at infinity. This pencil can be represented by any two vanishing points creating a line in the image plane which is referred to as the vanishing line $\mathbf{l}_v$ (see Figure 1). Although the parallelism of lines or planes is not preserved after the projection, the information about the orientation of the lines or the planes, which is implied by the vanishing point and the vanishing line, plays a key role in the proposed approach.

2.2. Generating the Slicing Planes

In order to derive required equations to recover the 3D shape, we make extensive use of the concepts discussed in the previous section. In a nutshell, the basic idea behind the proposed approach is to create subspaces in the object space by means of a series of planes parallel to each other and to a reference plane which physically exists in the scene. This concept is illustrated in Figure 2. These planes and the homography transform of each of them onto the images generate silhouette coherency maps which provide the 3D shape information when projected onto a reference image.

Let there be a set of points $\mathbf{X}_i$, $i = 1, 2, \ldots, N$ located on the reference plane π in the object space, where $N \geq 4$. Imagine that these points are elevated by distance Z vertically in the direction of the plane normal generating a new set of points $\mathbf{X}_i^{'}$, $i = 1, 2, 3, \ldots, N$. The lines originating from the reference-plane points $\mathbf{X}_i$ passing through the new points $\mathbf{X}_i^{'}$ create a set of parallel lines intersecting at the infinity. Additionally, the new point set constitute a new plane $\pi^{'}$, which is parallel to π. These planes create a pencil at the infinity.

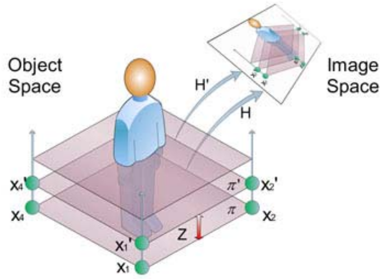


Figure 2. A set of hypothetical planes parallel to a reference plane intersect the object volume and create slices in the object space. By finding the successive image points $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$ along the reference direction, the homography across images is established for estimating the object slices.

The projections of two point sets $\mathbf{X}_i$ and $\mathbf{X}_i^{'}$ to an image plane result in two point sets $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$. Hence, the lines intersecting at the infinity in the object space become a set of lines passing through $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$ and intersecting at the vertical vanishing point $\mathbf{v}_z$. In similar vein, the pencil of planes π and $\pi^{'}$ becomes a vanishing line $\mathbf{l}_v$ in the image plane.

In the line of these observations, the creation of the image of a non-existing plane parallel to the reference plane requires estimation of a vanishing line $\mathbf{l}_v$ and the vanishing point $\mathbf{v}_z$. The vanishing line $\mathbf{l}_v$ can be obtained from the image of any two parallel line pairs that are also parallel to plane π. Similarly, the vanishing point $\mathbf{v}_z$ is estimated from the image of a line pair that is in the normal direction of π. The relationship between $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$ is established by rewriting equation (1) as:

$$\lambda_i \mathbf{x}_i^{'} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} + \mathbf{p}_3 Z. \qquad (4)$$

The column vector $\mathbf{p}_3$ corresponds to the vanishing point in the direction of the Z axis or the normal of the ground plane. By substituting $\mathbf{p}_3$ with $\mathbf{v}_z$ and combining with equation (2) provides:

$$\lambda_i \mathbf{x}_i^{'} = s_i \mathbf{x}_i + \mathbf{v}_z Z, \qquad (5)$$

where λ_i is simply the sum of s_i and Z due to the fact that the last components of the homogeneous coordinates being one. Once s_i is known, estimation of any image point along the line $\overrightarrow{\mathbf{x}_i \mathbf{v}_z}$ is achieved by setting Z to different set of values. The derivation of s_i for each feature point on π is defined by the following lemmas.

Lemma 1. Scale ratio of two image points. *When projecting two points* $\mathbf{X_1}$*,* $\mathbf{X_2}$ *lying on a plane in the object* *space onto the corresponding points $\mathbf{x_1}$, $\mathbf{x_2}$ in the image space using homography, the ratio of the scale factors s_1 and s_2 is the inverse proportion of the distances from the image points to the vanishing point of $\overrightarrow{\mathbf{X_1X_2}}$ direction.*

Proof. Given two points $\mathbf{X_1}$, $\mathbf{X_2}$ on the ground plane π and the corresponding points $\mathbf{x_1}$, $\mathbf{x_2}$ in the image space, equation (2) provides the following relation:

$$s_1\mathbf{x_1} - s_2\mathbf{x_2} = \mathrm{H}(\mathbf{X_1} - \mathbf{X_2}) = \mathrm{H}\begin{bmatrix} X_1 - X_2 \\ Y_1 - Y_2 \\ 0 \end{bmatrix}. \quad (6)$$

The right hand side is equivalent to computing the vanishing point $\mathbf{v}$ for the $\overrightarrow{\mathbf{X_1X_2}}$ direction, which can be expressed as

$$\mathrm{H}(\mathbf{X_1} - \mathbf{X_2}) = k\mathbf{v}, \quad (7)$$

where k is a scale factor. As stated in the previous section, the last component of the homogeneous coordinates is set to be one. By comparing the coefficients of equations (6) and (7) we observe that $k = s_1 - s_2$. Hence, the two equations can be combined and rearranged as

$$s_1(\mathbf{x_1} - \mathbf{v}) = s_2(\mathbf{x_2} - \mathbf{v}). \quad (8)$$

Taking norms to both sides of the equation above gives

$$\frac{s_1}{s_2} = \frac{|\mathbf{x}_2 - \mathbf{v}|}{|\mathbf{x_1} - \mathbf{v}|} \quad (9)$$

□

The relation can be further extended by applying the property of similar triangles as shown in Figure 3 which leads to:

Lemma 2. *The scale ratio in Lemma 1 equals to the inverse proportion of distances from the image points to the vanishing line of these parallel planes.*

The proof of this lemma is evident from Figure 3 and is not provided here. As long as the vanishing line $\mathbf{l}_{vi}$ is observed or computed from measuring two parallel line pairs in the image I_i, it can be transformed to any other image I_j by the homography as $\mathbf{l}_{vj} = \mathrm{H}_{ij}^{-T}\mathbf{l}_{vi}$, in which H_{ij}^{-T} is the transposed inverse of the homography from image I_i to image I_j.

The scale ratio provides a way to compute the scale factors for any number of feature points on a plane. Once the scale factor of an image point is determined, the others are computed from (9). The new point set $\mathbf{x}_i^{'}$ is then obtained from (4) and the homography between the images is estimated accordingly.

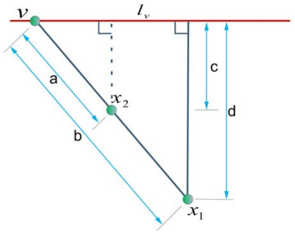


Figure 3. The ratio of the distances from $\mathbf{x}_1$ to $\mathbf{v}$ and from $\mathbf{x}_2$ to $\mathbf{v}$, which is b/a, equals to d/c according to similarity of triangles.

2.3. Recovering the Object Shape

Assume multiple images of a scene are provided and the first image is chosen as the reference image, such that the other images are warped onto the reference image by the mapping $I_{ij} = \mathrm{H}_{ij}I_j$, where the subscript ij indicates that the warping is from i^{th} to j^{th} image. In the sequel of a segmentation method, the object silhouettes extracted in these images create highlights of the slicing planes when they intersect with the object volume. To illustrate this, let's consider the object silhouette defined as a mask where the pixels inside the object are set to 1. The highlights of the slicing planes, when they intersect the object volume, is determined by warping all the silhouettes onto the reference image by:

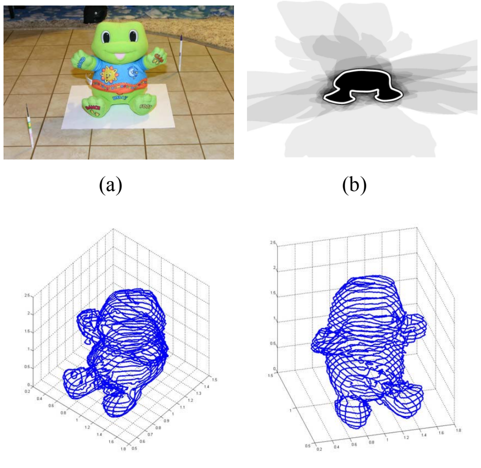
$$I_{intersection} = \frac{1}{n}\left(I_1 + \sum_{i=2}^{n} I_{i1}\right), \quad (10)$$

where n is the number of images. In equation (10), thresholding $I_{intersection}$ with the number of images n generates a mask image. This mask, also referred to as the slice-image in this paper, is conjectured to be the image of the intersections between the slicing planes and the object volume. Hence, using these masks, we generate the outlines of the object that corresponds to the surface of the object volume.

The mask generated from equation (10) can be back-projected to the object space in various ways. One approach is to use a set of feature points with known absolute object coordinates, such that the relation between the object space and the image space can be estimated. This can be realized by setting a specific feature such as a box, using the features on a building with known dimensions, or using the relative length ratio between linear features. Using one of these, one can generate a local Euclidean coordinate frame in the object space and the metric shape recovery can be achieved up to a scale. Selection of a local coordinate frame also reduces the effect of noisy silhouettes during the shape recovery. Theoretically, one can pick up any measurable feature on the reference plane even though the axes are not orthogonal, but ideally a square is preferred for achieving a robust metric reconstruction.

In this paper, we rather follow a different approach which eliminates the use of features that are known or extracted in the object space. We conjecture that the ground plane in the object space is identical to the reference image, such that the reference image is either affine rectified or acquired by an affine camera. In this setting, the shapes and coordinates of the objects on the ground plane appear exactly as those in the reference image; hence, all the slice-images, which are warped by respective homographies onto the affine rectified image, provides hypothetical planes at preset Z intervals. In the case of a perspective image of the ground plane which is conjectured to be acquired by an affine camera, the recovered shape will be projectively distorted. Nevertheless, the recovered shape provides the object shape. In our experiments, we have used ortho-rectified or affine rectified reference image where applicable.

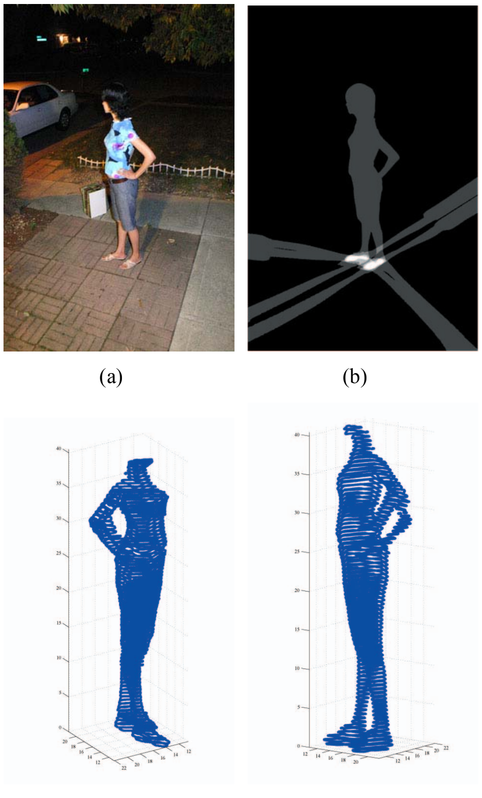
3. Experiments



(c) (d)

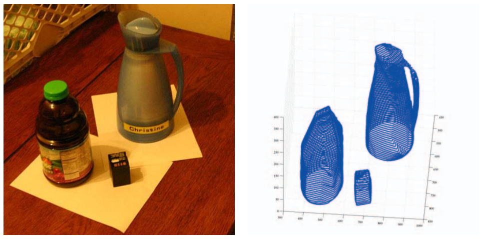
Figure 4. Recovering the 3D shape of a toy. (a) shows one of the eleven original images taken from different viewpoints. The contour of the darkest region in (b) provides the shape of the 3D volume sliced by the hypothetical plane at $Z = 1.5$. Two novel views of the reconstructed 3D shape are shown in (c) and (d).

In order to verify the proposed method, we have performed three sets of experiments. In the first experiment, as shown in Figure 4(a), we placed a toy, which contains irregular shape, on the ground plane. The ground plane contains squares which provides us with four measures required to estimate the homography from the images to the reference image and to the ground plane in the object space. Two pens oriented in the normal direction of the plane provides both the vanishing point in the direction of Z axis and the scale factor s given in equation (2). The dataset of the toy-experiment contains eleven images of the toy in which the objects parts are partially or fully occluded. We should note that, no length measurements are performed and the lengths of the vertical features are set to be a unit length. We do not consider availability of additional measurements in the object space except for the assumption that all the tiles have square shapes where the coordinates of four corners define a unit square. The 3D shape is reconstructed by setting the distance increments ΔZ in the vertical direction to 0.5 and computing corresponding Z values used to generate slicing planes. In order to generate fine 3D models, one can set ΔZ to lower values.



(c) (d)

Figure 5. Recovering the 3D shape of a human body. (a) shows one of the original images. The bright region in (b) reveals the shape of the feet which lie on the ground. Two novel views of the reconstructed 3D body are shown in (c) and (d).



(a) (b)

Figure 6. Recovering the 3D shapes of multiple objects. (a) is one of the original images. A novel view of the reconstructed 3D objects are shown in (b).

In this and the remaining three experiments, the scales and lengths of three axes are not measured, which suggests that the reconstruction is up to a scale factor which is the same in all dimensions. In order to expedite the reconstruction procedure, we use only the apparent contour of the intersection mask, such that only the contour generator of the object volume is recovered (Figure 4(b)). In this experiment, over 24,000 densely distributed 3D surface points are generated. The high details in the recovered shape suggests that the effect of occlusion in one view is compensated with information provided by other views. The shapes of the toes and hands are clearly visible from the rebuilt model shown in Figure 4(c) and 4(d).

The same approach is repeated for the second and the third experiments. In the second experiment five images of a person are taken. The tiles on the ground provide the features to estimate the vanishing line. The vanishing point in the Z direction is recovered from the features on the box in the images (Figure 5(a)). The result shows detailed shape recovery for the limbs and torso (Figure 5(c) and 5(d)). In order to demonstrate simultaneous recovery of multiple objects we conducted one last experiment where we placed three objects in the scene (Figure 6a). As shown without introducing additional complexity, the shapes of three different object are recovered with high precision.

4. Conclusions

The proposed approach in this paper reconstructs the 3D object shape by incorporating the silhouette images taken from uncalibrated cameras. The silhouette images are allowed to contain occlusions and distortions as long as some other views of the object reveal the occluded regions and do not observe the same distortion. The reconstruction achieved is a metric recovery up to a scale factor which can be determined if object space measurements are provided. The method provides an easy to implement algorithm while retaining strong geometric constraints. Compared to other algorithms, which require the generation of the visual hull or the estimation of the fundamental matrix, the proposed approach bears lower computational complexity and combines the merits of both approaches. The requirement of having abundant feature correspondences in other prevailing techniques is also removed to increase the computational efficiency. The adjustable balance between running time and accuracy is determined by the number of images and number of slicing planes. Additional post processing which has not been applied in this paper can be used to further improve the resulting 3D surfaces.

While eliminating most assumptions suggested by other algorithms, the proposed approach still requires the existence of specific features such as parallel lines for establishing geometric relations across images. However, these features are commonly observed in the real world scenes, as we have shown in the experiments. The experimental results also demonstrate the applicability of our method and the details of the objects are revealed using only few views. A variety of applications such as urban and rural surface modeling and real-time 3D object tracking can be realized by the proposed method.

References

[1] S. Blostein and T. Huang. Quantization error in stereo triangulation. *IEEE Int. Conf. on Computer Vision*, 1987. 1

[2] R. M. S. C. Jaynes, S. Webb. Camera-based detection and removal of shadows from interactive multiprojector displays. *IEEE Trans. on Visualization and Computer Graphics*, 10, 2004. 1

[3] K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *Int. Jrn. on Computer Vision*, 62, 2005. 1

[4] L. V. Gool, T. Tuytelaars, V. Ferrari, C. Strecha, J. Wyngaerd, , and M. Vergauwen. 3d modeling and registration under wide baseline conditions. *Proceedings of Photogrammetric Computer Vision*, pages 3–14, September 2002. 1

[5] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1992. 1

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in computer Vision-second edition*. Cambridge Un. Press, 2004. 1

[7] C. Hernandez, F. Schmitt, and R. Cipol. Silhouette coherence for camera calibration under circular motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007. 1

[8] P. Kelly, P. Beardsley, E. Cooke, N. O'Connor, and A. Smeaton. Detecting shadows and low-lying objects in indoor and outdoor scenes using homographies. *IEEE International Conference on Visual Information Engineering*, 2005. 1

[9] S. Khan and M. Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. *European Conf. on Computer Vision*, 2006. 1

[10] S. Khan, P. Yan, and M. Shah. A homographic framework for the fusion of multi-view silhouettes. *IEEE Int. Conf. on Computer Vision*, 2007. 1

[11] R. Koch, M. Pollefeys, and L. Gool. Realistic surface reconstruction of 3d scenes from uncalibrated image sequences. *Visualization and Computer Animation*, 2000. 1

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Jrn. on Computer Vision*, 60, 2004. 2

[13] W. Matusik, C. Bueler, and L. McMillan. Polyhedral visual hulls for real-time rendering. *Proceedings of 12th Eurographics Workshop on Rendering*, 2001. 1

[14] C. Stauffer and W. Grimson. Learning patterns of activity using real time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22, 2000. 2