Shape Recovery Using Rotated Slicing Planes

Po-Lun Lai and Alper Yilmaz
Photogrammetric Computer Vision Laboratory
The Ohio State University
Columbus, Ohio, USA
<http://dpl.ceegs.ohio-state.edu/>

Abstract—This paper discusses a novel approach for image-based three-dimensional (3D) object shape recovery. Let there be a set of hypothetical planes in the object space created by rotating a reference plane about an arbitrary axis. These planes slice the objects and create cross-sections at which the object shapes become the outlines on the plane, which we will refer to as the slice images. The object outlines are observed by projecting a series of object silhouettes at images taken by uncalibrated cameras at unknown viewpoints. The projections related to the slice images are computed from the plane projective geometry between images and these planes. Experimental results show that the proposed approach can recover 3D shape efficiently with minimum number of constraints.

I. INTRODUCTION

Shape recovery remains active area of research to researchers from different disciplines due to its wide applications such as visualization, city planning and scene analysis. Realistic 3D shapes are brought to life by powerful graphics processing hardware and advanced reconstruction algorithms. However, several obstacles reducing the efficiency and performance of the reconstruction process remain challenges to be addressed.

The conventional shape recovery requires that the cameras be calibrated and the object shape is reconstructed by triangulation of back-projected image points in the object space. In [8] camera self-calibration is obtained by robust tracking of salient feature points over the image sequence and dense depth maps are computed to create 3D surface mesh. Specific camera configurations, such as turn-table sequences, are commonly adopted in the reconstruction procedures [6] [4]. In [15] the authors exploit object silhouettes from an image sequence captured under circular motion to establish a set of 3D rim curves enclosing the object.

Voxel coloring for shape recovery has emerged recently and became popular. The shape is recovered from several images taken by calibrated cameras but no other assumptions are made about the scene or the relative orientation of the cameras [9] [3]. These methods do not require explicit pixel correspondences; instead, they rely on testing photo-consistency of the scene points to reconstruct the 3D scene. In the case where some concavities are not modeled or sufficient views are not acquired, the assumption has to be made that all points in the reconstruction space can be globally ordered in relation to the camera centers to avoid the visibility problems in the photo-consistency test [12].

Another line of 3D recovery methods rely on recovery the shape from silhouette (SFS). An advantages of these techniques over voxel coloring is its simplicity due to eliminating the photo-consistency test. These methods assume availability of silhouettes which can be obtained from object tracking in videos. Different articles on SFS use different constraints, such as marching intersections [13] requires large amount of structured grids and SFS across time [2] relies on temporal continuum. Literature on shape recovery also contains articles on combination of voxel coloring and SFS. In [1], a 3D shape recovery approach fuses silhouette, texture and shadow information to accurately generate the 3D shape. Another method exploits the duality principle governing surface points and their corresponding tangent planes for reconstructing complex 3D objects with unknown topology using silhouettes [11]. Both the SFS and voxel coloring methods, however, rely on high precision calibrated cameras and a nominal change in view points degrade their performance considerably.

In this paper, we propose a shape recovery technique incorporating silhouettes and the concept of slicing planes [14]. A similar line of work is reported in [7] and [10] respectively for 3D affine or metric recovery. The assumptions required in both these articles, such as requirement of affine cameras in [7] and the estimation of vanishing line in [10], are relaxed in our approach. Instead of having a set of parallel slicing planes, we assume that the planes are created by rotating the reference plane about an axis. A traditional approach to estimate linear mappings between images of our hypothetical planes can be estimated from the fundamental matrix and the epipole, along with additional pixel correspondences [5]. In this paper, we relax most of these requirements and define direct geometric relations which also eliminates the need for camera calibration or specific scene configuration. Proposed method is computationally efficient and easy to implement compared to other prevailing techniques. The formulation is elaborated in the following sections.

II. SLICING PLANES VIA ROTATION

A. Projection from rotation of a plane

Without loss of generality, let's assume a point in the object space lies on the ground plane π ($Z = 0$) and is represented in homogeneous coordinates by $\mathbf{X} = [X, Y, 0, 1]^T$. We should note that, throughout the paper we will use homogeneous coordinates for points both in the image and object spaces, and

Fig. 1. (a) Plane π is rotated about X axis by θ . (b) The new plane π_θ intersects the vertical line passing point $\mathbf{X}$ at $\mathbf{X}_\theta$.

the last element is always set to 1. Under projective geometry, $\mathbf{X}$ is mapped to an image point $\mathbf{x}$ in the image plane by:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \mathbf{P} \left(\begin{bmatrix} X \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ Y \\ 0 \\ 0 \end{bmatrix} \right) = s_1 \mathbf{m} + s_2 \mathbf{v}_y, \quad (1)$$

where $\mathbf{P}$ is 3×4 camera projection matrix which keeps the scale of $\mathbf{x}$ to remain at 1, $\mathbf{m}$ is any point on the image of X axis and $\mathbf{v}_y$ denotes the vanishing point of the Y axis. The introduced scale factors s_1 and s_2 reflect a projective equivalency class of $s[x, y, 1] = [x, y, 1]$.

As illustrated in Figure 1(a), rotation of plane π about X axis by an angle θ creates a new plane π_θ which intersect the vertical line passing through $\mathbf{X}$ at $\mathbf{X}_\theta = [X \ Y \ Y \tan \theta \ 1]^T$ (see Figure 1(b)). Let $a = \tan \theta$, then the image point corresponding to $\mathbf{X}_\theta$ in (1) becomes

$$\begin{aligned} s_\theta \mathbf{x}_\theta &= \mathbf{P}\mathbf{X}_\theta = \mathbf{P} \left(\begin{bmatrix} X \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ Y \\ 0 \\ 0 \end{bmatrix} + a \begin{bmatrix} 0 \\ 0 \\ Y \\ 0 \end{bmatrix} \right) \\ &= s_1 \mathbf{m} + s_2 \mathbf{v}_y + a s_3 \mathbf{v}_z \\ &= \mathbf{x} + a s_3 \mathbf{v}_z, \end{aligned} \quad (2)$$

where $\mathbf{v}_z$ denotes the vanishing point of Z axis and $s_\theta = 1 + a s_3$ due to the setting that the last elements of all point coordinates being one. For different points on the plane, the scale factors s_3 change with their orthogonal distances to the X axis, which correspond to their Y coordinates. However, on the absence of known projection matrix $\mathbf{P}$ and the absolute Y coordinates of $\mathbf{X}$, the location of $\mathbf{x}_\theta$ can not be determined explicitly without additional information.

Fig. 2. The image point $\mathbf{m}$, where the line connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ intersects the axis of rotation $\mathbf{l}$, remain unchanged after rotation. This property is used to find the points on the rotated plane.

B. Homography from rotation

A set of points $\mathbf{X}_i$ ($i = 1, 2, 3, \dots, N$; $N \geq 4$) lying on the plane π project to images $\mathbf{I}$ and $\mathbf{I}'$ by plane projective transform. These relations define a linear mapping between corresponding image points in $\mathbf{x}_i \in \mathbf{I}$ and $\mathbf{x}'_i \in \mathbf{I}'$ referred to as the homography transform $\mathbf{H}_\pi$:

$$\mathbf{x}' = \mathbf{H}_\pi \mathbf{x}, \quad (3)$$

induced by the plane π . After rotating plane π to π_θ , this mapping becomes $\mathbf{x}'_\theta = \mathbf{H}_{\pi_\theta} \mathbf{x}_\theta$. Combining equations (2) and (3) results in:

$$(\mathbf{x}' + a s_3 \mathbf{v}'_z) = \mathbf{H}_{\pi_\theta} (\mathbf{x} + a s_3 \mathbf{v}_z). \quad (4)$$

Equation (4) reveals the fact that under rotation of plane π , all points lying on the line passing through an image point and the vanishing point, when mapped onto another image by homography, remain on the corresponding line in another image.

Since the choice of axis in Euclidean coordinates is arbitrary, one can select any plane in the scene as the XY plane and choose any line as the X axis. Assume a vertical feature is identifiable in both images, and its height (Z) and distance to the axis of rotation (Y) are known, such that $a = \tan \theta = Z/Y$ is determined. The scale factor s_3 in equation (2) for all image point on the plane are solved by using the geometry shown in Figure 2. Assume the vanishing point $\mathbf{v}_z$ is estimated using vertical linear features, and $\mathbf{x}_1$ and $\mathbf{x}_2$ are two image points on plane π and $\mathbf{x}_{1\theta}$ is identified on the image. The line connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ intersects line $\mathbf{l}$, which is the image of the axis of rotation, at point $\mathbf{m}$. Since $\mathbf{m}$ is on the axis, its position remain unchanged after rotation. Hence, the image coordinates of $\mathbf{x}_{2\theta}$ can be computed by the cross product:

$$\mathbf{x}_{2\theta} = \mathbf{l}_{\mathbf{x}_{1\theta}, \mathbf{m}} \times \mathbf{l}_{\mathbf{x}_2, \mathbf{v}_z} \quad (5)$$

where $\mathbf{l}_{\mathbf{x}_{1\theta}, \mathbf{m}}$ and $\mathbf{l}_{\mathbf{x}_2, \mathbf{v}_z}$ are the lines connecting $(\mathbf{x}_{1\theta}, \mathbf{m})$ and $(\mathbf{x}_2, \mathbf{v}_z)$, respectively.

The above procedure is applied to all the selected image points on the plane π . Rearrange equation (2) and let $A =$

Fig. 3. Formation of the slicing planes. The points on different plane π_θ are shown with two points on the ground chosen as the axis of rotation. The angle θ goes from 5° to 70° with 5° increment.

$\mathbf{v}_z - \mathbf{x}_\theta$ and $B = \mathbf{x}_\theta - \mathbf{x}$, the scale factor s_3 for every point is then computed by

$$s_3 = \frac{1}{a} (A^T A)^{-1} A^T B. \quad (6)$$

Estimated scale factors give rise to new point sets in all images, which are computed using equation (2) by changing the value of θ . The new point set is then used for estimating the homography H_{π_θ} across images. In our setup, we use two of the N points to construct the axis of rotation, and the remaining $N-2$ points are used along with θ to estimate points related to rotated planes. Figure 3 demonstrates two views of a scene with the points on different hypothetical planes, π_θ .

We should note two degenerate configurations stemming from equation (5) and related geometry given in Figure 2. The first of these arise at $\theta = \pi/2$ where all the image points are either lie at infinity or coincide with the vanishing point $\mathbf{v}_z$. The second degeneracy occurs at a specific angle with which the projection of plane π_θ is line 1, indicating that π_θ is orthogonal to the image plane. Nevertheless, these conditions can be avoided by simple tests. Particular tests, we have included in our paper are as follows:

$$\begin{cases} |\theta| \leq \theta_{threshold} \\ \mathbf{1}^T \mathbf{x}_\theta \neq 0 \end{cases} \quad (7)$$

An alternative approach to avoid the first degenerate condition is to select the axis of rotation such that all scene objects reside on one side of the axis. In addition, since $\tan \theta$ increases rapidly beyond 45° , such that computational errors start to emerge, best recovery performance is achieved when the distance between the rotation-axis and the object is greater than the height of the object.

III. SHAPE RECOVERY

Given multiple images of a scene and the object silhouettes extracted from these images, the slice images corresponding

to the cross-sections of the objects and plane π_θ are generated by the homography transform. Let the first image be chosen as the reference image, such that the other images are warped onto the reference image by the mapping $I_{ij} = H_{ij} I_j$, where the subscript ij indicates that the warping is from i^{th} to j^{th} image. The object silhouette, which constitutes a mask, contains binary values set to 0 and 1 respectively for the non-object areas and object areas. The slice image is determined by mapping all silhouettes onto the reference image by:

$$I_{intersection} = \frac{1}{n} \left(I_1 + \sum_{i=2}^n I_{i1} \right), \quad (8)$$

where n is the number of images. Equation (8) creates a mask image by thresholding $I_{intersection}$ with the number of images. This mask is conjectured to be the image of the intersections between the slicing planes and the object volume due to the fact that transferred images coincide only at the locations that belong to the slicing-plane. Hence, by using this mask for every plane π_θ , we generate the outlines of the object shape which correspond to the surface of the 3D object volume.

With known absolute coordinates in the object space, the mask generated from equation (8) can be mapped to create metric 3D shape. For instance, by setting a specific feature such as a box, or using the relative length ratio between linear features, one can establish a local Euclidean coordinate frame in the object space and the metric shape recovery can be achieved up to some unknown scale. When ortho-rectified or affine-rectified reference image is applicable, we follow the assumptions in [10] to eliminate the use of features that are known or extracted in the object space by conjecturing that the ground plane in the object space is identical to the reference image. Every mask in π_θ is first mapped onto the ground by the homography between π_θ and π , then back-projected to the object space by the preset coordinates mapping, which is another homography. Dissimilar to the approaches using parallel slicing planes where Z is preset for each slice image, the Z coordinates for each point in the slice image varies with Y , and have to be reassigned as $Z = Y \tan \theta$. The absence of ground truth or ortho-rectified image may result in the recovered shape being projectively distorted. Nevertheless, the approach provides simple and efficient object shape recovery for visualization of the scene.

IV. EXPERIMENTS

Two image sets used in our previous paper [10] are adopted in the experiments for comparison reasons. The silhouettes are extracted manually for both sets. In the first experiment, a toy is placed in the scene and images of resolution 1536×1024 pixels are taken at eight different viewpoints of the toy. The corners of the tiles provide points for homography mapping and two of the points are selected to be lying on the axis of rotation (Figure 4(a)). Two pens are placed on the ground to provide linear features for the estimations of vanishing point and scale factors. Slice image is generated for each plane at angle θ which ranges from 0° to 45° with 1° increment. The outlines of the slice images are back-projected to the

Fig. 4. Shape recovery for a toy. The axis of rotation and points for constructing homography are shown in 4(a). The highlight in 4(b) corresponds to the slice image at $\theta = 15$. Two 3D views of the recovered shape are shown in 4(c) and 4(d). 4(e) and 4(f) show the results from [10] for comparison.

object space, creating 55,856 3D points. Two novel views of the shape are demonstrated in Figure 4(c) and 4(d). The detailed shape such as the fingers can be observed from the reconstructed shape. Compare to the visual result presented in [10] (Figures 4(e) and 4(f)), the proposed approach creates smoother shape using even fewer images. This can be attributed to the elimination of estimation for additional vanishing line. Extra post processing can be used to further improve the resulting 3D surfaces but is beyond the context of this paper.

In the second experiment, four images of a human body as shown in Figure 3(a), 3(b) and 5(a), 5(b) are used. A box on the ground plays the same role as the pens in the first experiment. The slice images of angles from 0° to 75° with 1° increment create 3D shape which consists of 15,006 points when back-projected. One may notice the sparseness of the points above the shoulders, which is caused by the high angles, but the points can be easily densified by changing the incremental angle between the slicing planes. However, the shapes of torso and legs are created smoothly, responding to

Fig. 5. Shape recovery for a human body. Figure 5(a) and 5(b) show two of the original images. Two 3D views of the recovered shape are shown in 5(c) and 5(d).

the statement that better performance is expected for the angles lower than 45° .

In both the experiments, the processing time for generating ten thousands of points using an coarse Matlab implementation is within an hour, and varies with the image size and the number of images used. The accuracy of reconstructed 3D models is not evaluated since no ground truth is obtained while taking the images.

V. CONCLUSION

We have developed a simple yet practicable method for the task of shape recovery. The proposed approach in this paper incorporates homography and silhouette images taken from uncalibrated cameras. The requirements for applying the approach are:

- a minimum of four points on a plane for constructing homography across images.
- vanishing point of the direction orthogonal to the plane.

The conditions are easily fulfilled since these features are commonly observed in urban or indoor environments. The silhouette images avoid problems caused by occlusion and distortions as long as some other views of the object compensate the occluded regions. The homography mapping provides strong geometric constraint while reducing the computational complexity compared to that in the estimation of fundamental matrix or the generation of visual hull. The experimental results reveal the practicability of the proposed approach for recovering object shape using fewer images.

REFERENCES

- [1] L. Ballan and G. M. Cortelazzo. Multimodal 3d shape recovery from texture, silhouette and shadow information. In *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 924–930, Washington, DC, USA, 2006.
- [2] K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *Int. J. on Computer Vision*, 62(3):221–247, 2005.
- [3] K. Chiang and K. Chan. Volumetric model reconstruction from unrestricted camera views based on the photo-consistency of 3d voxel mask. *J. of Machine Vision and Applications*, 17(4):229–250, September 2006.
- [4] A. W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3d model construction for turn-table sequences. In *SMILE'98: Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pages 155–170, London, UK, 1998. Springer-Verlag.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry in computer Vision - second edition*. Cambridge Un. Press, 2004.
- [6] G. Jiang, L. Quan, and H. Tsui. Circular motion geometry using minimal data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(6):721–731, June 2004.
- [7] S. Khan, P. Yan, and M. Shah. A homographic framework for the fusion of multi-view silhouettes. *IEEE Int. Conf. on Computer Vision*, pages 1–8, 2007.
- [8] R. Koch, M. Pollefeys, and L. Gool. Realistic surface reconstruction of 3d scenes from uncalibrated image sequences. *The Journal of Visualization and Computer Animation*, 11(3):115–127, 2000.
- [9] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38:307–314, 2000.
- [10] P. Lai and A. Yilmaz. Efficient object shape recovery via slicing planes. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. Unpaginated DVD-Rom.
- [11] C. Liang and K. Wong. Complex 3d shape recovery using a dual-space approach. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages II: 878–884, 2005.
- [12] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *Int. J. on Computer Vision*, 25(1), November 1999.
- [13] M. Tarini, M. Callieri, C. Montani, C. Rocchini, K. Olsson, and T. Persson. Marching intersections: An efficient approach to shape-from-silhouette. In *Proceedings of VMV 2002*, pages 255–262, 2002.
- [14] T. Wada, X. Wu, S. Tokai, and T. Matsuyama. Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras. *IEEE International Workshop on Computer Architectures for Machine Perception*, 2000.
- [15] H. Zhong, W. Lau, W. Sze, and Y. Hung. Shape recovery from turntable sequence using rim reconstruction. *Pattern Recognition J.*, 41(11):3295–3301, November 2008.