3D RECONSTRUCTION USING PHOTO CONSISTENCY FROM UNCALIBRATED MULTIPLE VIEWS

Heewon Lee and Alper Yilmaz
Photommetry Computer Vision Lab, Ohio State University, Columbus, OH 43210, U.S.A.

Keywords: Photo consistency, Homography, 3D Recovery.

Abstract: This paper presents a new 3D object shape reconstruction approach, which exploits the homography transform and photo consistency between multiple images. The proposed method eliminates the requirement of dense feature correspondences, camera calibration, and pose estimation. Using planar homography, we generate a set of planes slicing the object to a set of parallel cross-sections in the 3D object space. For each object slice, we check photo consistency based on color observation. This approach in return provides us with the capability for expressing convex and concave parts of the object. We show that the application of our approach to a standard multiple view dataset achieves comparably better performance than competing silhouette based method.

1 INTRODUCTION

Three-dimensional (3D) reconstruction of the object shape from multiple images has an important role in many applications including: tracking (Yilmaz et al, 2006), action recognition (Yilmaz and Shah, 2008) and virtual reality applications (Bregler, Hertzmann, and Biermann, 2000). The recent availability of image collection on the Internet and publicly available applications such as Google Earth and Microsoft Photosynth has increased the popularity of research on 3D recovery.

Several approaches achieve 3D shape recovery by finding 3D locations of matching points in different view images using camera calibration and pose (Isidoro and Sclaroff, 2003). These methods require a triangulation step, which backprojects points from the image space to the object space. However, these methods explicitly require camera pose and calibration require establishing high number of point correspondences between the images. Due to perspective distortions, variations of color across different views, and different camera gains generate a high number of correspondences.

In contrast, using segmented objects in the images and their direct back-projections, which generate a visual hull (Slabaugh and Schafer, 2001) is more flexible and eliminates the requirement to establish point correspondences. However, back-projections from the image space to the object space still require camera pose and calibration for all the images. Another limitation of back-projection of the silhouette is that recovered 3D object will not contain details of objects such as convexities and concavities despite having a high number of images. To overcome the limitations of visual hull based methods, researchers have proposed voxel coloring technique, which measures the color consistency of projected 3D points to images (Seitz and Szeliski, 2006). These methods label each grid on the object space as opaque or transparent by projecting each voxel to input images. These projections after check the consistency of the projected colors. Each consistent voxel is then checked for visibility. These methods, however, need precise calibration and pose information and may result in wrong surfaces due to occlusion problems in the visibility test.

In this paper, we propose new technique that eliminates the requirements for known camera calibration and pose. Proposed object approach analyzes implicit scene and camera geometry through a minimal number of point correspondences across images, which are used to form virtual images of a series of hypothetical planes and their relations in the images space. This paper is an extension to our former paper using silhouette-based methods(Lai and Yilmaz, 2008), which improves the recovered 3D considerably by introducing the photo consistency check. Particularly, the photo consistency check introduced in this paper provides detailed 3D recovery of convexities and concavities in the object surface.

Lee H. and Yilmaz A. (2010).
3D RECONSTRUCTION USING PHOTO CONSISTENCY FROM UNCALIBRATED MULTIPLE VIEWS.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 484-487
DOI: 10.5220/0002850504840487

The paper is organized as follows. Section 2 introduces the main concepts of proposed method. Particularly, in subsections 2.1 and 2.2, we discuss homography transformations between the images with respect to hypothetical planes slicing the object space. In subsection 2.3, we introduce our approach, which utilizes the color information across images to generate the 3D object shape. Section 3 compares silhouette-based methods with those of the proposed method. We conclude in Section 4 and present directions for further research.

2 SHAPE RECONSTRUCTION

In this section we provide a discussion on the projective geometry for estimating the relation between multiple views. Using these relations, we generate image of object slice along the direction normal to the reference plane. We should note that, in our experiments we select one of the image planes as the reference plane such that the recovered shape is correct up to a projective scale. Finally, we check color consistency within the all position of 3D object and generate 3D object having volumetric information.

2.1 Relations between Images

Mapping from the object space to the image space is governed by the camera matrix. Let the points in the image space, x, and the object space be respectively represented in the homogeneous coordinates by $x = (x, y, 1)^T$, and $X = (X, Y, Z, 1)^T$,(Hartley and Zisserman, 2000). The projection from X to x,is governed by the projective camera matrix P which introduces a scale factor λ due to projective equivalency of points in the homogenous coordinates:

$$\lambda x = PX \tag{1}$$

Considering that X lies on plane at Z=0, which we refer to as the ground plane, its projection to x simplifies P to homography transform, H. Homography transform provides a direct mapping between 3D plane and the image plane. Conversely, H also can be written to linearly map points between images:

$$x = HX' \tag{2}$$

Note that this mapping introduces another scale factor, s, which is different from the scale in equation (1). Let two image points x_1 , x_2 and 3D point X_1 there are corresponding points. The relations described as x_1=H_1 X_1 and x_2=H_2 X_2 give rise to a mapping between x_1, x_2 is as:

$$x_1 = H_1(H_2)^{-1}x_2 = H_{12}x_{12} \tag{3}$$

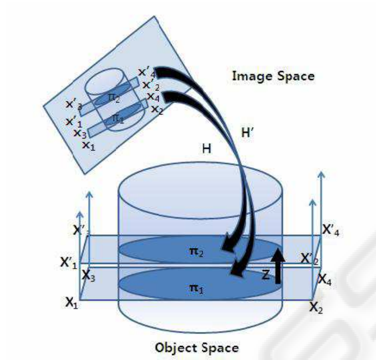

Figure 1: The relation between 3D planes in the object space and their images. A series of planes i parallel to the reference plane intersect the object volume and generate slices of the object.

2.2 Image of Object Slices

A stack of object slices when stacked together results in the object shape recovered up to a projective scale. We illustrate an instance of the conjecture in Figure 1, where the relations between image with respect to the object slices generated by hypothetical planes are denoted by H and H', planar homography matrix. These planes and homography transform of each of them on to the images generate coherency maps, which in turn provides the 3D object shape. Let each hypothetical plane has its normal direction aligned with the Z-axis, X_i be the set of 3D points X_1, X_2, X_3, and X_4 and X_i' be the set of 3D points X_1', X_2', X_3', and X_4', where X_i' andX_i are corresponding points in hypothetical planes. The image of the intersection of the lines connecting where X_i' andX_i provides us with the vertical vanishing point as shown in Figure 1. Using equations (1) and (2), the relation between points in two images can be de-fined based on the height of the hypothetical plane and the vertical vanishing point as:

$$\lambda_i x_i' = \begin{bmatrix} P_{11} & P_{21} & P_{41} \\ P_{12} & P_{22} & P_{42} \\ P_{13} & P_{23} & P_{43} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} + \begin{bmatrix} P_{31} \\ P_{32} \\ P_{33} \end{bmatrix} Z \tag{4}$$

$$= s_i x_i + V_z Z$$

In this equation, λ_i and s_i are the scale factors which are computed as elaborated in (Lai and Yilmaz, 2008). In these equation, parameters λ_i, s_i, V_z, and Z are known and provides direct relations between x_i and x_i'.

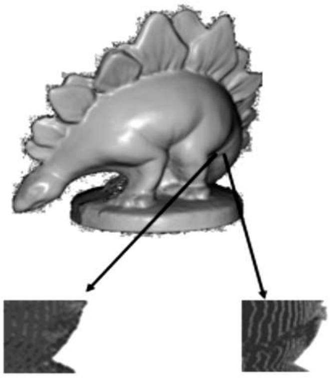

Figure 2: Concavities recovered using (left) silhouette based and (right) proposed approaches on leg of dinosaur.

2.3 Reconstructing the Object Shape

In the recent silhouette based approach by Lai and Yilmaz (2008), the indicator functions depicting the objects have resulted in removal of the object shape information implicitly encoded in the surface color. Therefore, resulting 3D object shape excluded convexities and concavities in the object surface. In Figure 2, we present convexities on the leg part of the dinosaur. A careful observation suggests that the silhouette-based model (left) couldnt detect concavities. In contrast, our approach (right) estimates the concavities correctly.

In our approach, we utilize the distribution functions of color observed in images in the form of kernel density estimate. We assume that a 3D point projecting to images consistently has same color at corresponding image pixels. By exploiting the relation in equation (4), the kernel density estimate is generated for corresponding image pixels using RGB bands as shown Figure 3. In Figure 4, we show the kernel density estimate generated from 48 views of the dinosaur object. According to the conclusion for color consistency presented in (Slabaugh and Schafer, 2001), the same point on object surface have similar color, hence provides a peak in the density estimate. In contrast, nonsurface points will have different colors and their histograms will have many low valued local maxima.

3 EXPERIMENTAL RESULTS

We experimented with dinosaur and temple images from the Middlebury multiview stereo dataset. Particularly, we compare our results with silhouette-based method. Figure 5 shows the resulting 3D shape for the dinosaur image set. In figure 5, (b) and (c) show two different types shape recovery using plot and iso-

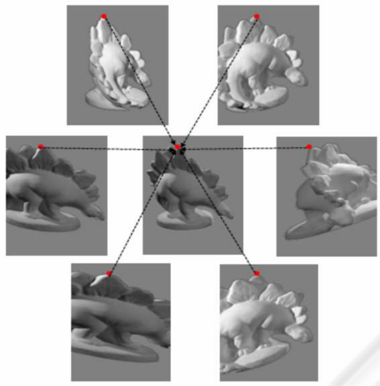

Figure 3: 3D point from multiple images warped to reference plane shown in the center.

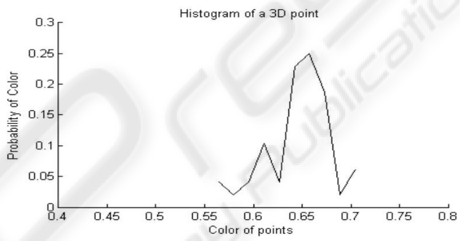

Figure 4: KDE generated from corresponding image pixels marked as dots in Figure 3 using 48 images.

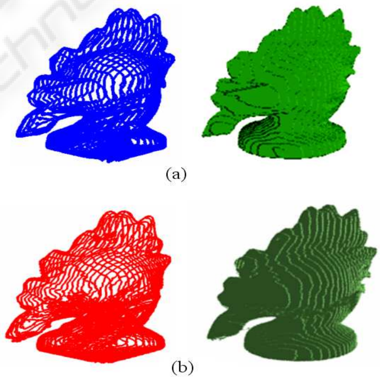

Figure 5: Recovering the 3D shape of dinosaurs. The 3D reconstructed object of silhouette-based method (a) and of our approach (b).

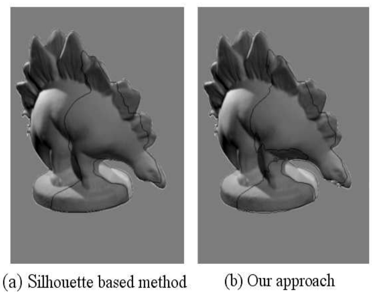

Figure 6: Reference image with sliced results using different methods of Z=-0.005.

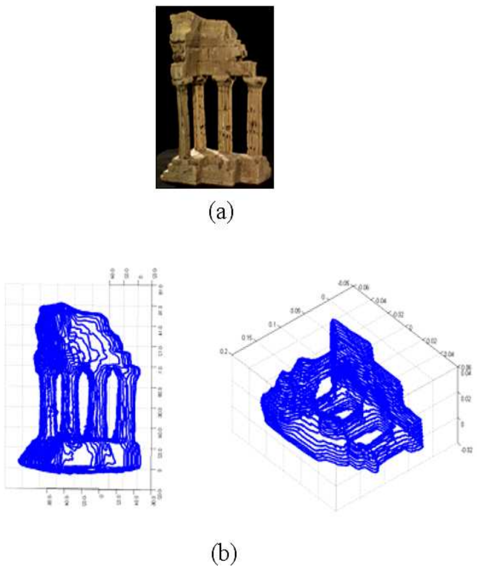

Figure 7: 3D shape recovered using 33 images. (a) Example image (b) two views of reconstructed shape.

surface function. In our results we used 0.005cm distance between the planes.

Comparing the silhouette based and proposed approaches, we observe that our approach can express details convex and concave parts on the leg and body part of the dinosaur effectively. This observation can be seen in Figure 6, where we show one particular slice superimposed on the dinosaur. As shown, the silhouette based method fails to generate precise object shape. Particularly, in part (a), the boundary from the slice incorrectly involves the dinosaurs neck and body parts. On the other hand, part (b), which is generated by our approach, estimates distinct concavity parts from neck to body. To demonstrate 3D recovery of a complex object, we conducted one last experiment as shown in Figure 7. It shows that a object having complex shapes are recovered with high performance.

4 CONCLUSIONS

In this paper we propose to generate sliced images based on photo consistency for 3D object shape recovery from uncalibrated cameras. Differ from silhouette-based methods, each image of object slices utilizes the color observed in 2D images. The probability value of histogram shows 3D object surface conditions such as concavity and convex. Without the need for camera calibration and estimation and unnecessary visibility checks, we recover 3D using simple linear mappings.

REFERENCES

Bregler, C., Hertzmann, A., and Biermann, H. (2000). Recovering non-rigid 3d shape from image streams. In *International Conference on Computer Vision*. IEEE.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge Univ Pr.

Isidoro, J. and Sclaroff, S. (2003). Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *IEEE International Conference on Computer Vision*. IEEE.

Khan, S. M., Yan, P., and Shah, M. (2007). A homographic framework for the fusion of multi-view silhouettes. In *ICCV 07, 11th International Conference on Computer Vision*. IEEE.

Lai, P. and Yilmaz, A. (2008). Efficient object shape recovery via slicing planes. In *CVPR 08, Conference on Computer Vision and Pattern Recognition*. IEEE.

Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR06, Conference on Computer Vision and Pattern Recognition*. IEEE.

Slabaugh, G., Culbertson, B., Malzbender, T., and Schafer, R. (2006). A survey of methods for volumetric scene reconstruction from photographs. In *n CVPR06, International Workshop on Volume Graphics.* IEEE.

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *ACM Journal of Computing Surveys*.

Yilmaz, A. and Shah, M. (2008). A differential geometric approach to representing the human actions. *Computer Vision and Image Understanding Journal*.