3D Object Reconstruction from Image Sequences with a One Line Search Method

J. Zhang, G. Chesi, and Y. S. Hung

Abstract

This paper addresses the problem of estimating the 3D model of an object from a sequence of images where the object is visible from different point of views. In particular, the paper considers the case of turntable image sequences, i.e. images captured under circular motion. A new method is hence proposed for this problem, which consists of determining the image point correspondences of each 3D point through a one line search where a photo-consistency index is maximized. Results with both synthetic and real data validate and illustrate the proposed method.

1 Introduction

In computer vision, 3D modeling of real objects is a key problem and has been widely applied in augmented reality, pose estimation, and digital entertainment. Generally, the methods take advantages of the points and silhouettes. They could be used exclusively. In the previous methods, feature points are used to compute the dense depth map [1, 2] of a 3D object by stereo matching techniques [3, 4] and fuse all points on the 3D object [5]. Silhouettes are used to compute visual hull [6, 7] of the 3D object. Recently, more and more methods combine these two information together to achieve high-quality reconstruction result with low computation complexity. They [8, 9, 10] reconstruct the visual hull as an initial solution and then refine the model using the feature point information. For example, in [11], object silhouettes in two views are exploited to establish a 3D rim curve, which is defined with respect to the two frontier points arising from two views. To reconstruct this 3D rim curve, its images in the two views are matched using traditional cross-correlation technique, and then the 3D rim curve is reconstructed using triangulation method over the two images, for instance with the approach in [12]. The method can reconstruct concave object surface with fast surface extraction. And in [15], a method with multi-view has been proposed to improve the depth precision of rim reconstruction in [11], which could obtain the point matching over wide baseline.

However, lots of aspects should be considered in the previous methods when determining the 2D matching points and the 3D reconstructed points. This paper presents a new method for dealing with these aspects. The idea is to reduce the problem of estimating the position of each 3D point to a one line search in spite of the number of available views. This is done by parametrizing the line where the candidate 3D point has to lie according to one view, and by determining its position on such a line through maximization of a photo-consistency index obtained from the other views. Such a maximization can be simply and quickly performed as there is only one scalar variable, moreover the maximum is typically very well distinct which is important in order to provide accurate results. This is shown in the paper with some examples with both synthetic and real data which validate and illustrate the proposed method.

The paper is organized as follows. Section 2 provides some preliminaries. Section 3 describes the proposed method. Section 4 presents some examples. Lastly, Section 5 concludes the paper with some final remarks.

2 Preliminaries

Let us introduce the notation used in this paper:

- $\mathbb{R}$: set of real numbers;
- $\mathbb{P}^n$: set of vectors of $\mathbb{R}^n$ with the last entry equal to 1;
- A^T: transpose of matrix A;
- I_n: $n \times n$ identity matrix;
- e_i: i-th column of I_3.

We denote with $\mathcal{I}_i$ the i-th image of the sequence, with $P_i \in \mathbb{R}^{3\times 4}$ the projection matrix of the i-th image, and with $C_i \in \mathbb{P}^4$ the camera center of the i-th image (in homogeneous coordinates). Moreover, we denote with $x_{i,j} \in \mathbb{P}^3$ the image projection on the i-th image of the j-th 3D point (in homogeneous coordinates). This image projection is given by the perspective law

$$x_{i,j} = \frac{P_i X_j}{e_3^T P_i X_j} \tag{1}$$

where $X_j \in \mathbb{P}^4$ is the j-th 3D point (in homogeneous coordinates).

The 3D object reconstruction problem is as follows: given a sequence of N images $\mathcal{I}_1, \ldots, \mathcal{I}_N$ where an object is visible from different point of views, estimate the 3D model of the object, i.e. a set of 3D points $X_1, \ldots, X_M$ lying on the surface of the object, where M defines the accuracy of the reconstruction. In the sequel we will consider the case of turntable image sequences, i.e. images captured under circular motion as shown in Figure 1. Also, we will consider the case where the projection matrices $P_1, \ldots, P_N$ have been already estimated.

3 Proposed Method

Given a pair of images in the sequence, we can define a 3D rim curve as explained in [11]. Specifically, denote

Figure 1. 3D object reconstruction from turntable image sequence.

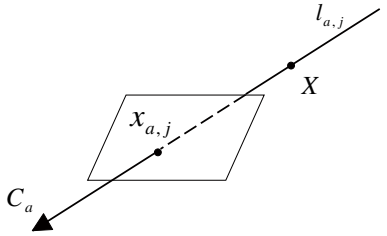


Figure 2. Line $l_{a,j}$. This line is defined by C_a (camera center for view a) and $x_{a,j}$ (considered point in view a).

with $\mathcal{I}_a$ and $\mathcal{I}_f$ the two images with camera centers C_a and C_f respectively. The two planes passing by C_a and C_f that are tangent to the object are called frontier planes. Their tangent points with the object are called frontier points [13]. Let us define the frontier points as X_1 and X_2 (generally, each frontier plane touches the object at one point only, hence there are two frontier points). Consider the plane passing by X_1, X_2 and C_a. This plane cuts the object, and the intersection between the plane and the object is a planar curve divided into two segments by the points X_1 and X_2. The 3D rim curve is the segment closest to C_a. This 3D rim curve projects onto $\mathcal{I}_a$ as a straight line, and onto $\mathcal{I}_f$ as a curve called 2D rim curve. The straight line in $\mathcal{I}_a$ is sampled, obtaining a sequence of image points called source points.

Consider a generic source point previously obtained, and denote it as $x_{a,j}$. The first step of the proposed method consists of identifying the ray corresponding to this point, which is the line passing by C_a and $x_{a,j}$, see Figure 2. This line can be expressed as

$$l_{a,j} = \left\{ X \in \mathbb{P}^4 : \; x_{a,j} = \frac{P_a X}{e_3^T P_a X}, \; e_3^T P_a X > 0 \right\}. \quad (2)$$

Let us parametrize this line. A simple possibility is according to

$$l_{a,j} = \left\{ X_{a,j}(\alpha) : \; \alpha \in \mathbb{R}, \; e_3^T P_a X_{a,j}(\alpha) > 0 \right\} \quad (3)$$

where

$$X_{a,j}(\alpha) = U_{a,j} + \alpha V_{a,j}. \quad (4)$$

In the above expression, $U_{a,j} \in \mathbb{P}^4$ is a generic point on the line, $V_{a,j} \in \mathbb{R}^4$ is the direction of the line with the last entry of $V_{a,j}$ equal to zero, and $\alpha \in \mathbb{R}$ is a free parameter. These quantities can be found as follows. Define the matrix

$$M = N(I_3 - x_{a,j} e_3^T) P_a \quad (5)$$

where

$$N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

and let W be a matrix whose columns form an orthnormal basis for the null space of M. Since the size of M is 2×4 and P_a has full rank, it follows that the size of W is 4×2. Decompose W as

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \quad (6)$$

where $W_1, W_2 \in \mathbb{R}^{2\times 2}$. In general, W_2 is invertible, and we can define

$$Z = W W_2^{-1} = \begin{pmatrix} W_1 W_2^{-1} \\ I_2 \end{pmatrix} \quad (7)$$

and the sought vectors $U_{a,j}$ and $V_{a,j}$ are given by

$$\begin{array}{rcl} U_{a,j} & = & Z_2 \\ V_{a,j} & = & Z_1 \end{array} \quad (8)$$

where Z_i is the i-th column of Z. However, if W_2 is not invertible, let us observe that it must have the form

$$W_2 = \begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix} \quad (9)$$

where $c_3^2 + c_4^2 \neq 0$ since the line must exist, and where $c_1 c_4 - c_2 c_3 = 0$. Suppose without loss of generality that $c_3 \neq 0$ and define Z as above by replacing W_2^{-1} with W_3 where

$$W_3 = \begin{pmatrix} 1/c_3 & -c_4/c_3 \\ 0 & 1 \end{pmatrix}. \quad (10)$$

It follows that the last row of Z is again $(1, 0)$, and hence $U_{a,j}$ and $V_{a,j}$ can be defined as in the previous case.

The next step is to reproject the generic point $X_{a,j}(\alpha)$ on the line $l_{a,j}$ onto images that are close to $\mathcal{I}_a$. Hence, denote with $\mathcal{I}_{b_i}$ a generic image of these ones, where $i = 1, \ldots, R$ and R is the number of close images that are considered. We have that $X_{a,j}(\alpha)$ projects onto $\mathcal{I}_{b_i}$ on the image point given by

$$x_{b_i,j}(\alpha) = \frac{P_{b_i} X_{a,j}(\alpha)}{e_3^T P_{b_i} X_{a,j}(\alpha)} \quad (11)$$

as illustrated by Figure 3.

From the image points $x_{b_i,j}(\alpha)$, $i = 1, \ldots, R$, we can define a photo-consistency index with the source point $x_{a,j}$ as

$$\mu_{a,j}(\alpha) = \sum_{j=1}^{R} ncc(\mathcal{A}_a(x_{a,j}), \mathcal{A}_{b_i}(x_{b_i,j}(\alpha))) \quad (12)$$

where $\mathcal{A}_k(x)$ denotes an image area of chosen size in $\mathcal{I}_k$ centered on x, and $ncc(\mathcal{A}_1, \mathcal{A}_2)$ denotes the normalized cross-correlation (NCC) score between the image areas $\mathcal{A}_1$ and $\mathcal{A}_2$.

Therefore, we estimate the 3D point corresponding to $x_{a,j}$ as the point on the line $l_{a,j}$ that maximizes the photo-consistency index $\mu_{a,j}(\alpha)$. This point is hence given by $X_{a,j}(\alpha^*)$ where α^* is the solution of

$$\begin{array}{rcll} \alpha^* & = & \arg \sup_{\alpha} & \mu_{a,j}(\alpha) \\ & & \text{s.t.} & e_3^T P_a (U_{a,j} + \alpha V_{a,j}) > 0. \end{array} \quad (13)$$

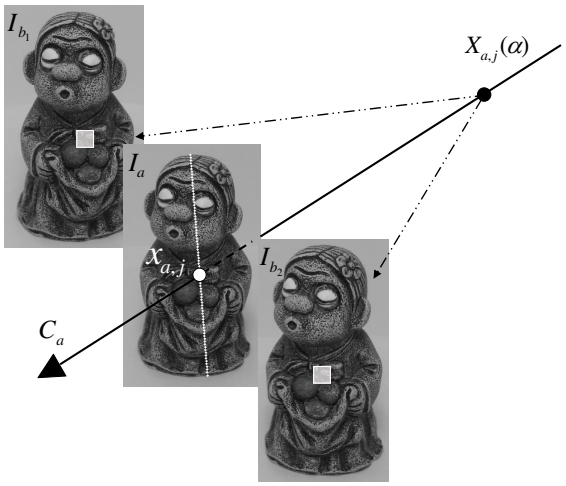


Figure 3. A generic point on the line $l_{a,j}$ is reprojected onto views $\mathcal{I}_{b_1}, \mathcal{I}_{b_2}, \ldots$.

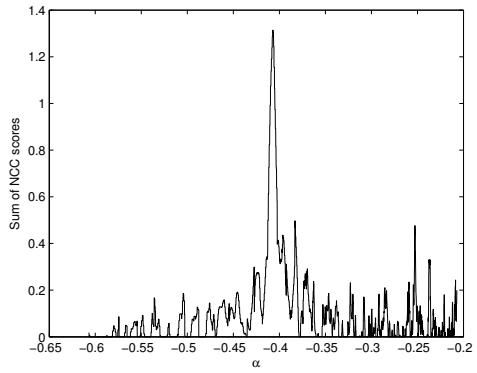


Figure 4. A typical case for the function $\mu_{a,j}(\alpha)$.

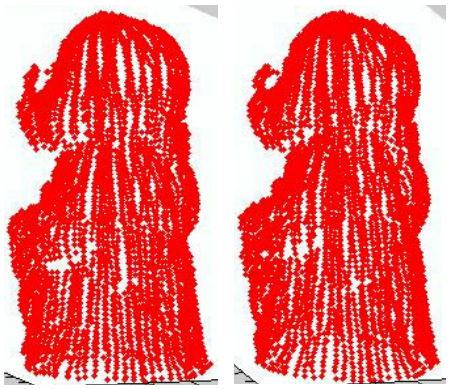
The above problem can be simply solved since only a scalar variable is involved. Indeed, a simple strategy consists of defining a grid of points $\alpha_1, \alpha_2, \ldots$ in an interval of interest, and evaluate $\mu_{a,j}(\alpha)$ at each of these points, hence taking the one that maximizes $\mu_{a,j}(\alpha)$. This procedure is simple, fast and reliable as it will be shown in the results section. Figure 4 shows a typical shape for $\mu_{a,j}(\alpha)$, where it can be seen that the peak is clearly defined.

4 Experimental Results

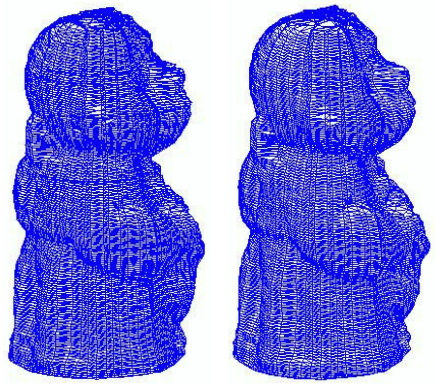
The proposed method has been applied to a real turntable sequence (the claylady) and a synthetic sequence (the ball-cone model). We have chosen sequences with $N = 36$ images, i.e. a rotation angle equal to 10 degrees between any two consecutive images of the turntable sequence. The matching images located right next and left next to $\mathcal{I}_a$ if three views are used for one rim reconstruction. More than three views could also be implemented. We always select the views most close to $\mathcal{I}_a$.

Figure 5 shows the 3D rim curves estimated with the proposed approach and the triangulated mesh model based on the 3D rim curves. Figure 6 gives the mesh model with the synthetic ball-cone model.

To measure the reconstructed error, we computed



(a) Reconstructed rims with 3 views. (b) Reconstructed rims with 5 views.



(c) Reconstructed mesh model with 3 views. (d) Reconstructed mesh model with 5 views.

Figure 5. Reconstructed results with real object - claylady.

Table 1. 3D Error for the ball-cone object.

number of views	average	max	std dev
3	0.8858	13.7359	0.7234
5	0.8604	5.8658	0.6984

the corresponding ground truth 3D points in ball-cone model with the known geometry constraints. The 3D errors are show in Table 1. The 3D average error and standard deviation of the same model in [11] are 1.0085 and 0.9078 respectively.

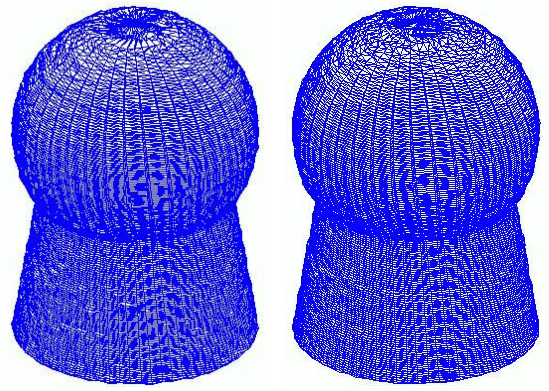
The reconstructed 3D rim curves have also been projected back to the 2D images to check the reconstruction. As shown in Figure 7, the 2D rim curves are consecutive and the points have the similar location as in the source image.

5 Conclusion

In this paper, the one line search method has been proposed to obtain the 3D model of an object from a turntable image sequence, which could predict the 3D object surface points along the ray defined by the camera center and the corresponding 2D points efficiently and satisfactorily.

(a) Source 2D rim points in left view. (b) Matched 2D rim points in source view. (c) Matched 2D rim points in right view.

Figure 7. Projected rim curves for the claylady model with 3 views. The 3D rim curve corresponding to the 2D rim curve in source view (straight line) has been projected back to the left and right images. As the 2D curves match with each other, the 3D reconstructed points should be correct.



(a) Reconstructed mesh model with 3 views. (b) Reconstructed mesh model with 5 views.

Figure 6. Reconstructed results with synthetic object - ball-cone.

References

[1] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7-55, 1987.

[2] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 353-363, 1993.

[3] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 787-800, 2003.

[4] G. Van Meerbergen, M. Vergauwen, M. Pollefeys, and L. Van Gool, "A Hierarchical Symmetric Stereo Algorithm Using Dynamic Programming," *International Journal of Computer Vision*, vol. 47, pp. 275-285, 2002.

[5] R. Koch, M. Pollefeys, and L. J. V. Gool, "Multi Viewpoint Stereo from Uncalibrated Video Sequences," the 5th European Conference on Computer Vision, Freiburg, Germany, 1998.

[6] B. G. Baumgart, "Geometric modeling for computer vision," Ph.D Thesis, Stanford University, 1974.

[7] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 150-162, 1994.

[8] Y. Furukawa and J. Ponce, "Carved Visual Hulls for Image-Based Modeling," European Conference on Computer Vision, Graz, Austria, 2006.

[9] C. H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," *Computer Vision and Image Understanding*, vol. 96, pp. 367-392, 2004.

[10] J. Isidoro and S. Sclaroff, "Stochastic Refinement of the Visual Hull to Satisfy Photometric and Silhouette Consistency Constraints," the Ninth IEEE International Conference on Computer Vision, Beijing, China, 2003.

[11] H. Zhong, W. S. Lau, W. F. Sze, and Y.S.Hung, "Shape recovery from turntable sequence using rim reconstruction," *Pattern Recognition*, vol. 41, pp. 295-301, 2008.

[12] G. Chesi and Y.S. Hung, "Fast Multiple-View L2 Triangulation with Occlusion Handling," *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 211-223, 2011.

[13] R. Cipolla and P. J. Giblin, *Visual Motion of Curves and Surfaces*: Cambridge University Press, 1999.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*: Cambridge University Press, 2000.

[15] J. Zhang, F. Mai, Y.S. Hung and G. Chesi, "3D Model Reconstruction from Turntable Sequence with Multiple-View Triangulation," *International Symposium on Advances in Visual Computing*, vol. 5876, pp. 470-479, 2009.