Multiview stereo object reconstruction with a one-line search method

Jian Zhang
Graziano Chesi

Multiview stereo object reconstruction with a one-line search method

Jian Zhang
Graziano Chesi
University of Hong Kong
Department of Electrical and Electronic Engineering
Pokfulam Road, Hong Kong
E-mail: jzhang@eee.hku.hk

Abstract. *We propse the one-line search (OLS) method for estimating the three-dimensional (3-D) model of an object from a sequence of images where the object is visible from different points of view. The OLS method estimates each 3-D point through a OLS along the ray defined by the camera center and its two-dimensional (2-D) image projection in one view by maximizing a photo-consistency index built with the other views. The key advantage of the proposed method is that only one variable is introduced for estimating the sought 3-D point, whereas approaches for multiview stereo typically exploit multiple epipolar lines and hence require multiple variables. Results with both synthetic and real data illustrate and validate the proposed method. © 2013 SPIE and IS&T [DOI: 10.1117/1.JEI.22.2.023019]*

1 Introduction

Three-dimensional (3-D) modeling of real objects is a key problem in computer vision, in particular for augmented reality,[1] pose estimation,[2] and digital entertainment. Generally, the existing methods can be classified into pixel-based methods, which solve the problem mainly in image space, and voxel-based methods, which solve the problem mainly in scene space.

In pixel-based methods, feature points are used to compute the dense depth map[3–5] of an object by stereo matching techniques[6–8] and to fuse all points on the object.[9] Silhouettes are used to compute visual hull[10–12,13] of the object and iteratively get an approximation polyhedron.[14,15] Also, silhouettes are exploited in Refs. 16 and 17 to recover the shape of the object through rim reconstruction.

In voxel-based methods, space carving,[18–21] level-sets,[22–24] and surface mesh grid optimization[25–27] are typically exploited. These methods start with an initial surface guess larger than the object to reconstruct, which could be the visual hull or just a bounding box. In particular, space carving is used to identify scene surface and carve voxels that are not consistent with photographs. To this end, photo-consistency[28,29,30] and volumetric min-cut like graph cuts[18,31] and geo-cuts[32] are exploited. Level-sets are used to carve the initial space and shrink to the sought object by minimizing an energy function. In general, the object is implicitly represented as a zero-level set of a volumetric function. Last, surface mesh grid optimization is used to refine the model mesh by using silhouette- and texture-driven force[25] and iteratively minimizing the reprojection errors at randomly sampled points.[26]

This paper presents a novel solution for addressing multiview stereo object reconstruction. The main idea is to reduce the problem of estimating the position of each 3-D point to a one-line search (OLS) in scene space. This is done by parametrizing the line where the candidate 3-D point has to lie according to the information provided by one view and by determining its position on such a line through the maximization of a photo-consistency index obtained from the other views. The proposed OLS method has two main advantages. The first advantage is that each 3-D point is estimated through the maximization of a photo-consistency index that depends on one variable only, whereas approaches for multiview stereo typically exploit multiple epipolar lines and hence require multiple variables. The second advantage is that such a maximization can be simply performed since the peak of the photo-consistency index with respect to the variable introduced can be identified very clearly in general. The proposed method is illustrated and validated by using synthetic and real data.

The paper is organized as follows. Section 2 provides some preliminary definitions and the problem formulation. Section 3 describes the proposed method. Section 4 presents the experimental results. Last, Sec. 5 concludes the paper with some final remarks.

2 Preliminaries

Let us introduce the notation used in the paper. $\mathbb{R}$ is the set of real number. $\mathbb{P}^n$ is the set of vectors of $\mathbb{R}^n$ with the last entry equal to 1. A^T is the transpose of matrix A. I_n is the $n \times n$ identity matrix. E_i is the I'th column of I_3. $\|a\|$ is the Euclidean norm of vector a. $\text{mod}(a, b)$ means the modulus after division of a with b.

Denote with $\mathcal{I}_a$ the a'th image of the sequence, with $P_a \in \mathbb{R}^{3 \times 4}$ the projection matrix of the a'th image, and with $C_a \in \mathbb{P}^4$ the camera center of the a'th image (in homogeneous coordinates). Moreover, $x \in \mathbb{P}^3$ denotes the image projection (in homogeneous coordinates) on the a'th image of the 3-D point $X \in \mathbb{P}^4$ (in homogeneous coordinates) according to the perspective law

$$x = \frac{P_a X}{e_3^T P_a X}. \qquad (1)$$

The multiview stereo 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 points of view, estimate the 3-D model of the object, i.e., a set of 3-D points $X_1, \ldots, X_M$ lying on the surface of the object. In the sequel, consider the case of turntable image sequences, i.e., images captured under circular motion as shown in Fig. 1, though the method can be used also in other cases. Moreover, we will assume that an estimate of the projection matrices $P_1, \ldots, P_N$ is available.

3 Proposed Method

The first step of the proposed method consists of extracting the silhouette of the object in all images of the sequence. This can be done in various ways, for instance using a level-set evolution method.[33]

The second step consists of defining the rim curves as described in Ref. 16. Specifically, let $\mathcal{I}_a$ and $\mathcal{I}_f$ be a pair of images in the sequence, and let C_a and C_f be their camera centers. The two planes that pass by C_a and C_f and are tangent to the silhouette of the object are called frontier planes, and their tangent points with the silhouette of the object are called frontier points.[34] Define the frontier points as X_1 and X_2 (generally, each frontier plane touches the silhouette of the object at one point only, and thus two frontier points exist). Consider the plane that passes 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, which is divided into two segments by the points X_1 and X_2. The 3-D rim curve is the segment between these two that is closer to C_a. This 3-D rim curve projects onto $\mathcal{I}_a$ as a straight line, and onto $\mathcal{I}_f$ as a curve called 2-D rim curve.

The third step consists of estimating the 3-D rim curve. To this end, the straight line in $\mathcal{I}_a$ is sampled to obtain a sequence of image points called source points. Let $x_{a,j}$ be the j'th source point in image $\mathcal{I}_a$. Denote the ray defined by $x_{a,j}$ and the camera center C_a in image $\mathcal{I}_a$ as $l_{a,j}$, whose expression is given by

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

Fig. 1 Multiview stereo object reconstruction from a turntable image sequence. A sequence of images are taken from different points of view.

Parametrize the ray $l_{a,j}$. A simple possibility is according to

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

where

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

In the above expression, $U_{a,j} \in \mathbb{P}^4$ represents 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 calculated as follows. Define the matrix

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

where

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

and let W be a matrix whose columns form an orthonormal 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

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

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

$$Z = WW_2^{-1} = \begin{pmatrix} W_1 W_2^{-1} \\ I_2 \end{pmatrix}, \qquad (7)$$

which provides the sought vectors $U_{a,j}$ and $V_{a,j}$ as

$$U_{a,j} = Z_2 \quad V_{a,j} = Z_1, \qquad (8)$$

where Z_i is the i'th column of Z. If W_2 is not invertible, it must have the form

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

where $c_1 c_4 - c_2 c_3 = 0$, moreover $c_3^2 + c_4^2 \neq 0$ since the line is not at infinity; hence supposing without loss of generality that $c_3 \neq 0$, the matrix Z from which the sought vectors $U_{a,j}$ and $V_{a,j}$ are extracted can be defined as above by replacing W_2^{-1} with W_3 given by

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

The computed ray $l_{a,j}$ contains the sought 3-D point corresponding with the image point $x_{a,j}$. In particular, the position on the ray $l_{a,j}$ of such a 3-D point is defined by the parameter α.

In order to estimate α for recovering the sought 3-D point corresponding to the image point $x_{a,j}$ on the ray $l_{a,j}$, project the generic point $X_{a,j}(\alpha)$ onto images close to $\mathcal{I}_a$. Hence denote a generic image among the selected ones by $\mathcal{I}_{b_i}$, $i = 1, \ldots, R$, where $\mathcal{I}_{b_1}, \ldots, \mathcal{I}_{b_R}$ are the close

images that are considered. The projection of $X_{a,j}(\alpha)$ onto $\mathcal{I}_{b_i}$ is 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 in Fig. 2.

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

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

where $\mathcal{A}_a(x_{a,j})$ denotes an image area of chosen size in image $\mathcal{I}_a$ centered at $x_{a,j}$, and ncc denotes the normalized cross-correlation (NCC) score which can be calculated using Eq. (13). Specifically, let $\mathcal{A}_1(x_1)$ and $\mathcal{A}_2(x_2)$ be image areas with centers $x_1 = (u_1, v_1, 1)^T$ and $x_2 = (u_2, v_2, 1)^T$ in images $\mathcal{I}_1$ and $\mathcal{I}_2$. Let $(2n+1) \times (2m+1)$ be the size of these image areas (in pixels). Then, the NCC score between $\mathcal{A}_1(x_1)$ and $\mathcal{A}_2(x_2)$ is given by

$$\mathrm{ncc}[\mathcal{A}_1(x_1), \mathcal{A}_2(x_2)]$$
$$= \frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} \prod_{k=1}^{2} [f_k(u_k+i, v_k+j) - \gamma_k]}{(2n+1)(2m+1)\delta_1\delta_2}, \quad (13)$$

where $f_k(u_k+i, v_k+j)$ is the intensity of image $\mathcal{I}_k$ at pixel $(u_k+i, v_k+j, 1)^T$,

$$\gamma_k = \frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} f_k(u_k+i, v_k+j)}{(2n+1)(2m+1)}, \quad (14)$$

and
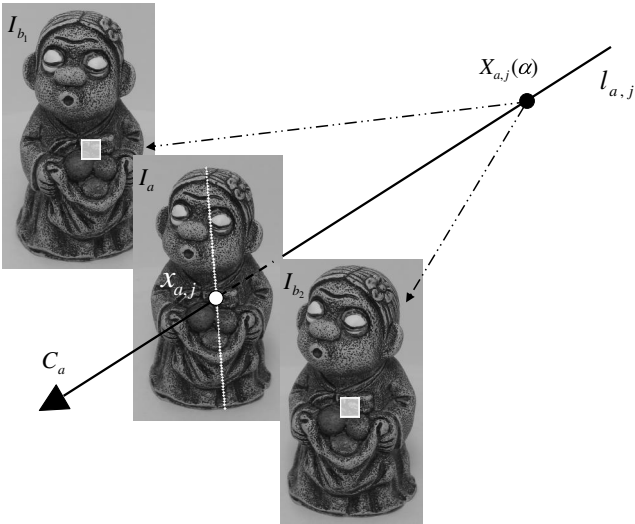
Fig. 2 The generic point $X_{a,j}(\alpha)$ on the ray $l_{a,j}$ is projected onto views $\mathcal{I}_{b_1}, \mathcal{I}_{b_2}, \ldots$. This ray is defined by the image point $x_{a,j}$ in image $\mathcal{I}_a$ and the camera center C_a of image $\mathcal{I}_a$. The matching scores are computed among patches in close views. The real 3-D point is the one whose projected patches have the highest matching scores. The white point sequence is the sampled source points. $x_{a,j}$ is the j'th one among them on image I_a.

$$\delta_k = \sqrt{\frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} [f_k(u_k+i, v_k+j) - \gamma_k]^2}{(2n+1)(2m+1)}}. \quad (15)$$

Then estimate the 3-D point corresponding to $x_{a,j}$ as the point on the ray $l_{a,j}$ that maximizes the photo-consistency index $\mu_{a,j}(\alpha)$. Hence this point is given by

$$X_{a,j}^* = X_{a,j}(\alpha^*), \quad (16)$$

where α^* is given by

$$\alpha^* = \arg \sup_{\alpha} \mu_{a,j}(\alpha) \quad \text{s.t.} \quad e_3^T P_a(U_{a,j} + \alpha V_{a,j}) > 0. \quad (17)$$

This optimization problem can be easily solved since only one variable is involved, for instance via grid techniques. Figure 3 shows a typical shape for $\mu_{a,j}(\alpha)$, where the peak is clearly defined. Observe that an interval $[\alpha_{\min}, \alpha_{\max}]$ where confining α in the search for α^* can be easily obtained by imposing that the projection of $X_{a,j}(\alpha)$ lies inside the silhouette of the object in all views.

Table 1 summarizes the proposed OLS method. Due to the object shape, some of the reconstructed 3-D rim curves may be scattered. Hence, if two successive 3-D rim curves are too far from each other, it is necessary to insert one or more new 3-D rim curves between them. To this end, one can compute the distance of two successive 3-D rim curves (one in image $\mathcal{I}_a$ and another in image $\mathcal{I}_{a+1}$) as

$$\max_{j,k} \|X_{a,j}^* - X_{a+1,k}^*\|. \quad (18)$$

4 Experimental Results

In this section, we present some results obtained with the proposed OLS method. In particular, three examples are reported in this section. For all examples, the camera parameters were estimated beforehand.

The first example considers a real turntable sequence (the claylady). The resolution of each image is 2160×1440 pixels. Choose sequences with $N = 36$ images, i.e., a rotation
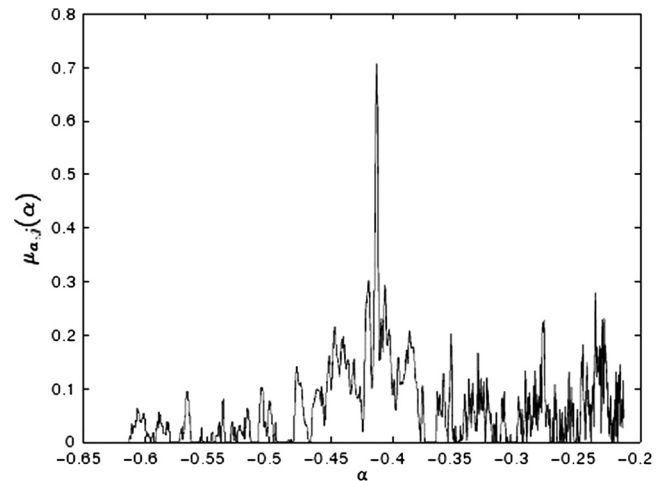
Fig. 3 A typical case for the photo-consistency index $\mu_{a,j}(\alpha)$. The maximizer of this function is the quantity α^*.

Table 1 Proposed OLS method.

Input: sample points of 2-D rim curves in each image

Output: 3-D rim curve points

For each 2-D rim curve r_a in image $\mathscr{I}_a$, $a = 1, \ldots, N$

    For each sample point $x_{a,j}$ in the 2-D rim curve r_a, $j = 1, \ldots, M_a$

        Compute $U_{a,j}$ and $V_{a,j}$ which define $l_{a,j}$

        Compute $\alpha_{\min}$ and $\alpha_{\max}$ which define the search interval for α^*

        Compute $\alpha^* = \arg \sup_\alpha \mu_{a,j}(\alpha)$ s.t. $e_3^T P_a(U_{a,j} + \alpha V_{a,j}) > 0$

        Define the estimated 3-D point as $X_{a,j}^* = X_{a,j}(\alpha^*)$

    Loop

Loop

---

angle equal to 10 deg between any two consecutive images of the turntable sequence.

The second example considers a synthetic turntable sequence (the ball-cone model). The ball and the cone are generated according to

$$X = X_B \cup X_C$$
$$X_B = \{X: x^2 + y^2 + (z - 90)^2 = 50^2 \quad \text{and} \quad z \geq 60\}$$
$$X_C = \{X: z = 6(x^2 + y^2)^{(1/2)} + 300 \quad \text{and} \quad z < 60\}. \quad (19)$$

The 2-D images are obtained with persistence of vision ray-tracer, which could put artificial textures on the object surface and project to 2-D figures with given camera settings. There are a total of 21,900 points on the surface and the image resolution is 1280×1024 pixels. Also in this example, $N = 36$ images, i.e., a rotation angle equal to 10 deg between any two consecutive images of the turntable sequence.

The third example is a dinosaur image sequence downloaded from the Middlebury website.[35,36] This sequence consists of 48 photos, which are taken by the cameras located as a ring around the object, which is not exactly centered in the ring. The rotation angles between successive images are roughly equal. A total of 46 images are used. The model surface is smoother and does not have any texture which could be adjusted to by human naked eyes. The image resolution is 640×480 pixels.

In the case of three views used for reconstructing a rim curve, i.e., $R = 2$, the images located right next and left next to the image $\mathcal{I}_a$, namely $\mathcal{I}_{b_1}$ and $\mathcal{I}_{b_2}$, are selected according to the rule

$$R = 2 \Rightarrow b_1 = \mathrm{mod}(i - 2, N) + 1, \quad a = i,$$
$$b_2 = \mathrm{mod}(i, N) + 1.$$

Similarly, in the case of five views, i.e., $R = 4$, denote with $\mathcal{I}_{b_3}$ and $\mathcal{I}_{b_4}$ the images located right next and left next to the

images $\mathcal{I}_{b_1}$ and $\mathcal{I}_{b_2}$, respectively, which are selected according to the rule

$$R = 4 \Rightarrow b_3 = \mathrm{mod}(i - 3, N) + 1,$$
$$b_1 = \mathrm{mod}(i - 2, N) + 1, \quad a = i$$
$$b_2 = \mathrm{mod}(i, N) + 1, \quad b_4 = \mathrm{mod}(i + 1, N) + 1.$$

Table 2 provides the average, maximum, and minimum values of the quantity

$$\mu_{a,j}^* = \mu_{a,j}(\alpha^*) \quad (20)$$

over all possible a and j for the three examples considered by using either three or five views for reconstructing a rim curve. The matching half-window size is 15 pixels for the dinosaur model and 10 pixels for the other models.

Figure 4 shows the estimated 3-D rim curves for the claylady example using either three or five views to reconstruct each rim curve. Figure 5 shows the triangulated mesh

**Table 2** Average, maximum, and minimum values of $\mu_{a,j}^*$ over all possible a and j.

Example	Number of views	Average	Maximum	Minimum
claylady	3	0.8640	0.9911	0.3244
claylady	5	0.4957	0.8217	0.1925
ball-cone	3	0.7460	0.9910	0.1931
ball-cone	5	0.3730	0.4955	0.0966
dinosaur	3	0.9277	0.9990	0.2616
dinosaur	5	0.8898	0.9986	0.2832

(a)   (b)
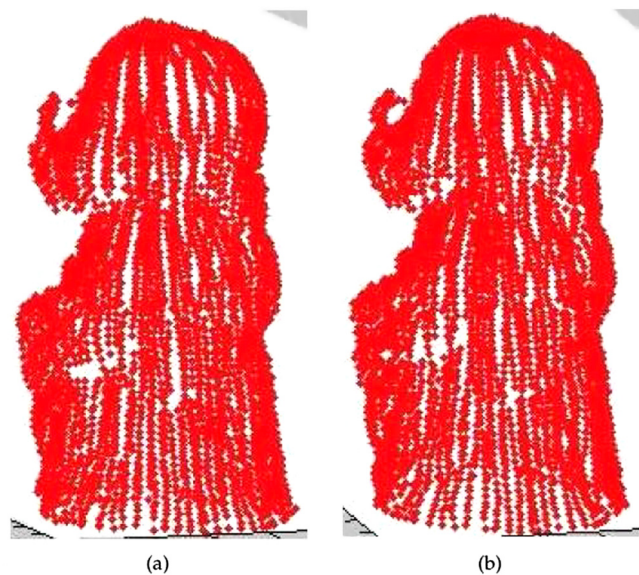
Fig. 4 Reconstructed 3-D rim curves for the claylady example with (a) three views and (b) five views. Each 3-D rim curve was reconstructed by corresponding three or five views.
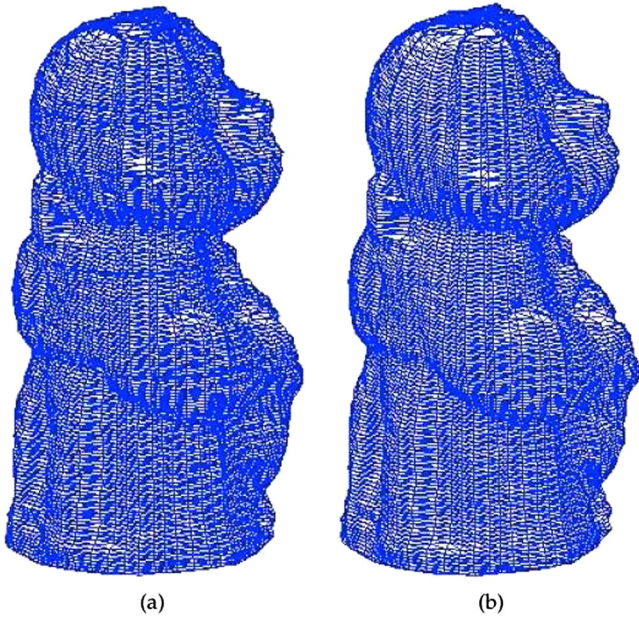
Fig. 5 Reconstructed triangulated mesh model for the claylady example with (a) three views and (b) five views. The mesh model was computed according to the relative reconstructed 3-D rim curves in Fig. 4.
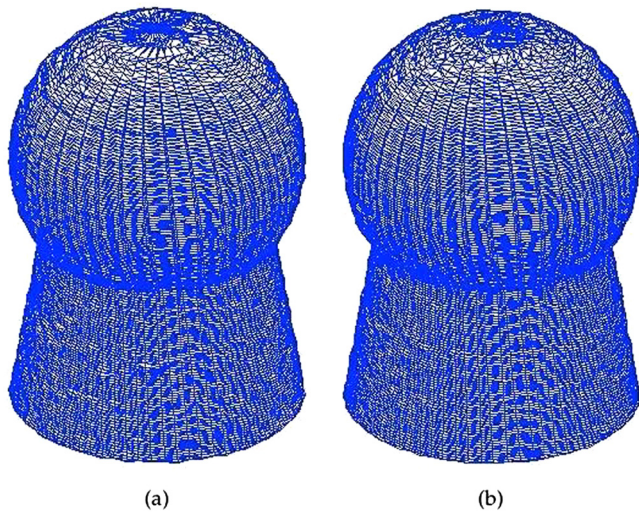

Fig. 6 Reconstructed triangulated mesh model for the ball-cone with (a) three views and (b) five views. The mesh model was computed according to the relative reconstructed 3-D rim curves for the ball-cone.

**Table 3** 3-D error for the ball-cone example.

Error type	Number of views	Average	Maximum	Maximum (90%)	Std. Dev.
Type A	3	0.5113	13.7359	1.0248	0.7234
Type A	5	0.5027	5.8658	1.0044	0.6984
Type B	3	0.7100	1.8555	1.0499	0.2586
Type B	5	0.7100	2.0178	1.0497	0.2547

model[37] obtained for the claylady example with 68 3-D rim curves and 150 slicing planes.

Similarly, Fig. 6 shows the triangulated mesh model obtained for the ball-cone example with 72 rim curves and 150 slicing planes. Since the data in this example are synthetic, the 3-D error of the proposed method can be computed. In particular, Table 3 shows two types of 3-D
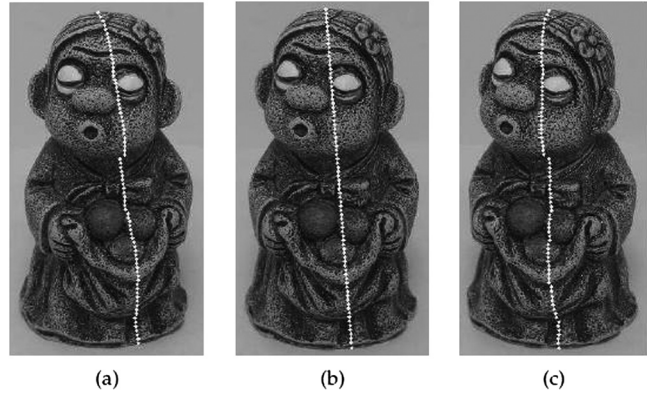

Fig. 7 Projected rim curve with three relative views for the claylady. 3-D rim points projected to (a) left view, (b) source view, and (c) right view. One of the 3-D rim curves in Fig. 4(a) was projected back to the relative images (left view, source view, and right view). Since the 2-D curves match with one another, the 3-D reconstructed points should be correct.
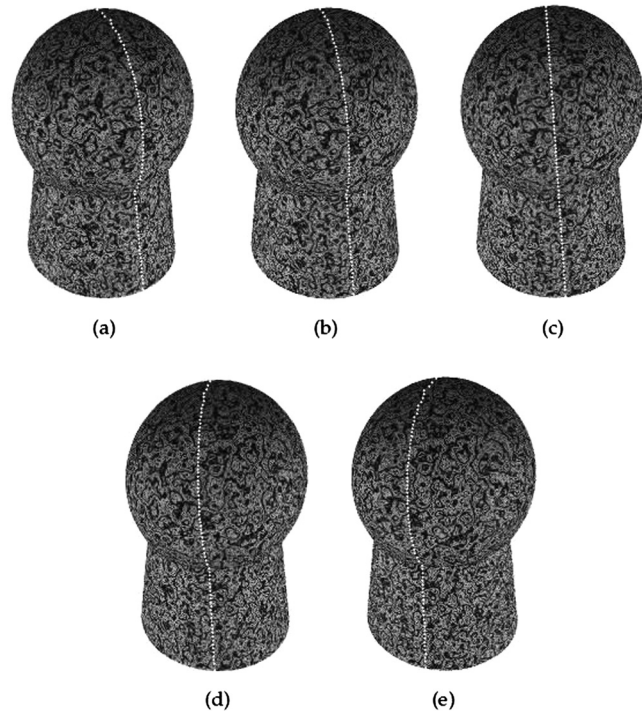
Fig. 8 Projected rim curves with five relative views for the ball-cone. 3-D rim points projected to (a) second left view, (b) left view, (c) source view, (d) right view, and (e) second right view. One of the 3-D rim curves of the ball-cone model was projected back to the relative images (two left views, source view, and two right views). Since the 2-D curves match with one another, the 3-D reconstructed points should be correct.
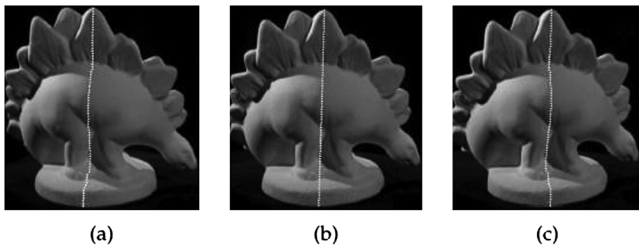
Fig. 9 Projected rim curve on a smooth surface for the dinosaur. 3-D rim points projected to (a) left view, (b) source view, and (c) right view. This rim curve lies on a smooth surface. It has been projected back to the relative three views. The 2-D curves match with one another.
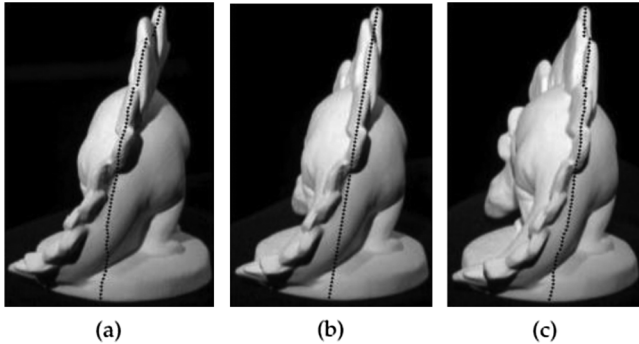
Fig. 10 Projected rim curve on a complicated surface for the dinosaur. 3-D rim points projected to (a) left view, (b) source view, and (c) right view. This rim curve lies on a complicated surface. It has been projected back to the relative three views. The 2-D curves match with one another.

errors:[38,39] the distance between a reconstructed 3-D point and its true value (type A), and the distance between a reconstructed 3-D point and its nearest point in the cloud of true points (type B). The column "maximum (90%)" shows the maximum evaluated over the best 90% reconstructed points, while the column "std. dev." shows the standard deviation.

Figure 7 shows the projection of a reconstructed 3-D rim curve onto different 2-D images of the sequence in order to evaluate the quality of the reconstruction for the claylady example.

Similarly, Fig. 8 shows the projection of a reconstructed 3-D rim onto different 2-D images of the sequence for the ball-cone example.

Last, Figs. 9 and 10 show the projections of two reconstructed 3-D rim curves onto different 2-D images of the sequence for the dinosaur example.

5 Conclusion

This paper has proposed the OLS method for estimating the 3-D model of an object from a sequence of images. This method estimates each 3-D point through a OLS along the ray defined by the camera center and its 2-D image projection in one view by maximizing a photo-consistency index built with the other views. The key advantage of the proposed method is that only one variable is introduced for estimating the sought 3-D point, whereas approaches for multiview stereo typically exploit multiple epipolar lines and hence require multiple variables. Future work

will consider the introduction of more complex photo-consistency indexes in the proposed method.

References

1. R. T. Azuma, "A survey of augmented reality," *Teleoperators Virt. Environ.* **6**(4), 355–385 (1997).
2. E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: a survey," *IEEE Trans. Patt. Anal. Mach. Intell.* **31**(4), 607–626 (2009).
3. R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: an approach to determining structure from motion," *Int. J. Comput. Vis.* **1**(1), 7–55 (1987).
4. M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. Patt. Anal. Mach. Intell.* **15**(4), 353–363 (1993).
5. R. Szeliski and S. B. Kang, "Direct methods for visual scene reconstruction," in *Int. Conf. Image Proc.*, New York, pp. 26–33 (2002).
6. I. J. Cox, S. L. Hingorani, and S. B. Rao, "A maximum likelihood stereo algorithm," *Comput. Vis. Image Under.* **63**(3), 542–567 (1996).
7. J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Patt. Anal. Mach. Intell.* **25**(7), 787–800 (2003).
8. G. Van Meerbergen et al., "A hierarchical symmetric stereo algorithm using dynamic programming," *Int. J. Comput. Vis.* **47**(1–3), 275–285 (2002).
9. R. Koch, M. Pollefeys, and L. J. V. Gool, "Multi-viewpoint stereo from uncalibrated video sequences," in *5th European Conf. Computer Vision*, Freiburg, Germany, Vol. 1, pp. 55–71 (1998).
10. N. Ahuja and J. Veenstra, "Generating octrees from object silhouettes in orthographic views," *IEEE Trans. Patt. Anal. Mach. Intell.* **11**(2), 137–149 (1989).
11. B. G. Baumgart, "Geometric modeling for computer vision," Ph.D. Thesis, Stanford University (1974).
12. A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(2), 150–162 (1994).
13. E. Boyer and J. S. Franco, "A hybrid approach for computing visual hulls of complex objects," in *Proc. Comp. Vis. Patt. Recong.*, Madison, Wisconsin, Vol. 1, pp. 695–701 (2003).
14. P. Srinivasan et al., "Computational geometric mehtods in volumetric intersection for 3-D reconstruction," *Pattern Recogn.* **23**(8), 843–857 (1990).
15. A. Laurentini, "The visual hull: a new tool for contour-based image understanding," in *Proc. 7th Scandinavian Conf. Image Analysis*, Aalborg, Denmark, pp. 993–1002 (1991).
16. H. Zhong et al., "Shape recovery from turntable sequence using rim reconstruction," *Pattern Recogn.* **41**(11), 295–301 (2008).
17. J. Zhang et al., "3-D model reconstruction from turntable sequence with multiple-view triangulation," in *International Symposium on Advances in Visual Computing*, Las Vegas, Vol. 5876, pp. 470–479 (2009).
18. Y. Furukawa and J. Ponce, "Carved visual hulls for image-based modeling," in *European Conf. Computer Vision*, Graz, Austria, Vol. 1, pp. 564–577 (2006).
19. K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *Int. J. Comput. Vis.* **38**(3), 199–218 (2000).
20. G. Slabaugh, R. Schafer, and M. Hans, "Multi-resolution space carving using level set methods," in *Int. Conf. Image Proc.*, New York, Vol. 2, pp. 545–548 (2002).
21. G. Zeng et al., "Progressive surface reconstruction from images using a local prior," in *IEEE Inter. Conf. Comput. Vis.*, Beijing, China, Vol. 2, pp. 1230–1237 (2005).
22. S. Soatto, A. Yezzi, and H. Jin, "Tales of shape and radiance in multi-view stereo," *IEEE Inter. Conf. Comput. Vis.*, Nice, France, Vol. 2, pp. 974–981 (2003).
23. J.-P. Pons, R. Keriven, and Q. Faugeras, "Modelling dynamic scenes by registering multi-view image sequences," *IEEE Conf. Comput. Vis. Pattern Recogn.*, San Diego, Vol. 2, pp. 822–827 (2005).
24. H. Jin, S. Soatto, and A. Yezzi, "Multi-view stereo reconstruction of dense shape and complex appearance," *Int. J. Comput. Vis.* **63**(3), 175–189 (2005).
25. C. H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3-D object modeling," *Comp. Vis. Image Underst.* **96**(3), 367–392 (2004).
26. J. Isidoro and S. Sclaroff, "Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints," in *Ninth IEEE Intern. Conf. on Computer Vision*, Beijing, China, Vol. 1, pp. 1335–1342 (2003).
27. H. Kuck, W. Heidrich, and C. Vogelgsang, "Shape from contours and multiple stereo: a hierarchical mesh-based approach," *IEEE Canadian Conf. Comp. Robot Vision*, London, Canada, pp. 76–83 (2004).
28. G. Vogiatzis, P. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts," *IEEE Conf. Comput. Vis. Pattern Recogn.*, San Diego, Vol. 2, pp. 391–398 (2005).
29. K. N. Kutulakos and C. R. Dyer, "Global surface reconstruction by purposive control of observer motion," *Artif. Intell. J.* **78**(1–2), 147–177 (1995).

30. S. M. Seitx and C. R. Dyer, "Complete scene structure from four point correspondences," in *5th Intern. Conf. Computer Vision*, DC, USA, pp. 330–337 (1995).
31. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in $N - D$ images," in *Intern. Conf. Computer Vision*, Vancouver, British Columbia, Canada, Vol. 1, pp. 105–112 (2001).
32. Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," in *IEEE Int. Conf. Comput. Vis.*, Nice, France, Vol. 1, pp. 26–33 (2003).
33. C. Li et al., "Level set evolution without re-initialization: a new variational formulation," in *IEEE Conf. Comp. Vis. Pattern Recogn.*, San Diego, Vol. 1, pp. 430–436 (2005).
34. R. Cipolla and P. J. Giblin, *Visual Motion of Curves, and Surfaces*, Cambridge University Press, Cambridge, United Kingdom (1999).
35. http://vision.middlebury.edu/mview/data/.
36. S. M. Seitz et al., "A comparison and evaluation of multiview stereo reconstruction algorithms," in *Proc. Comp. Vision and Patt. Recog.*, New York, Vol. 1, pp. 519–528 (2006).
37. G. Chesi and Y.S. Hung, "Fast multiple-view L2 triangulation with occlusion handling," *Comput. Vis. Image Understand.* **115**(2), 211–223 (2011).
38. G. Csurka, D. Demirdjian, and R. Horaud, "Finding the collineation between two projective reconstructions," *Comput. Vis. Image Understand.* **75**(3), 260–268 (1999).
39. H. Radu et al., "Rigid and Articulated Point Registration with Expectation Conditional Maximization," *IEEE Trans. Pattern. Anal. Mach. Intell.* **33**(3), 587–602 (2011).

Graziano Chesi received the Laurea in information engineering from the University of Florence in 1997 and the PhD in systems engineering from the University of Bologna in 2001. He has served as associate editor for *Automatica*, *BMC Research Notes*, the *European Journal of Control*, the *IEEE Transactions on Automatic Control*, *the IEEE Transactions on Computational Biology*, and *Bioinformatics, and Systems and Control Letters*. Also, he has served as guest editor of the Special Issues on Positive Polynomials in Control (*IEEE Transactions on Automatic Control*), Systems Biology (*International Journal of Robust and Nonlinear Control*), and *Visual Servoing* (*Mechatronics*). He is author of the book *Homogeneous Polynomial Forms for Robustness Analysis of Uncertain Systems* (Springer, 2009), editor of the book *Visual Servoing via Advanced Numerical Methods* (Springer, 2010), and author of the book *Domain of Attraction: Analysis and Control via SOS Programming* (Springer, 2011).

Jian Zhang currently works with the Department of Diagnostic Radiology of Wake Forest University Medical Center as a research fellow. She received her PhD in electrical and electronic engineering from the University of Hong Kong in 2012. She was a visiting member at INRIA Grenoble Rhone-Alpes, Montbonnot, France, in 2009. Her research interests include 3-D reconstruction, feature matching, image analysis, and medical image processing.