http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (1): 34-40, 2010
ISSN 1812-5638

Robustly Shape from Calibrated Images under Complex Illumination

Shuangtao Ma, Jiuqiang Han and Xinman Zhang
School of Electronics and Information Engineering, Xian Jiaotong University,
Xian City, Shaanxi Province, Zip Code 710049, China

Abstract: Focusing on shape from calibrated images under complex illumination, a two-step reconstruction policy is presented, which combines bounding edge model and level set method. Bounding edge model is used to sample visual hull, while controlling noise by projection color variance threshold. Based on level set method, evolution process of partial differential equation can shrink to target surface from the bounding edge samples, while controlling filter effect by a revisory coefficient in energy model. A technique called varying time steps TVD Runge-Kutta is proposed to reinitialize level set function, which guarantees upwind design of each Euler step. The experiment results show that the deformable initial surface can automatically shrink to model of object from the samples with accuracy and robustness under complex illumination that high light and shadow exist synchronously.

Key words: Machine vision, three-dimensional reconstruction, shape from silhouettes, level set method

INTRODUCTION

Shape recovery from images has been one of the most studied topics in computer vision and under a wild application background, e.g., inverse engineering, image-based modeling and three dimensional measurements. This great number of applications has motivated a great deal of research and led to many significant achievements.

Many algorithms have been developed for constructing volumetric models from a set of silhouette images, that is, shape from silhouettes method and they can be categorized into two groups: the discrete approach (also called visual hull approach) and the differential approach. The discrete approach takes the 3D reconstruction from silhouettes as a problem computing the intersection of the volumetric cones back projected from silhouettes in different views. The main drawback of this approach is that it cannot deal with the self-occluding edges and the reconstructed surface is not smooth. The differential approach is based on the assumption that the object surface is smooth enough to allow a differential analysis of the properties of the 2D silhouettes and of the 3D surface to be constructed. In this method, strong assumptions are usually made about the motion of camera and the local shape of the object to get a closed-form solution and it cannot deal with sparse views. See the work of Dyer (2001) and Slabaugh *et al*. (2001) for two surveys on volumetric-based methods.

In the past two decades, stereo vision algorithms have been investigated in computer vision in order to infer 3D structure from images captured with different viewpoints. The most challenging problem in stereo reconstruction is the establishment of visual correspondence among images. This is a fundamental operation that is the starting point of most geometric algorithms for 3D shape reconstruction and motion estimation. Although, significant progress has been made in dense two-frame stereo matching (Scharstein *et al*., 2001) for a comprehensive survey, producing accurate results near depth discontinuities remains a challenge. In general, dense stereo techniques can be classified as local or global, depending on whether they rely on local window-based computations or the minimization of a global energy function. In local-based methods, the disparity computation at a given point depends only on intensity values within a finite window. Clearly, these techniques assume that all pixels within the window have the same disparity and thus are sensitive near object boundaries. Attempts to alleviate this problem include the use of adaptive windows (Kanade and Okutomi, 1994) and shift able windows (Kang *et al*., 2001).

Thus, shape from silhouettes and shape from stereo are the two classical methods. Shape from silhouettes (Ehsan *et al*., 2007) based on volume carving can estimate robustly object surface, but it fails on the non-convex part. Shape from stereo (Hirschmuller and Scharstein, 2007) concludes depth information using correspondence

Corresponding Author: Shuangtao Ma, No. Mail Box 1854, Xian Jiaotong University, Xian City, Shaanxi Province, Zip Code 710049, China

of features between two/multi images, which introduces the ambiguity issue during stereo match. Esteban and Schmitt (2003) combined shape form silhouettes and shape from stereo, coarse surface can be estimated by shape from silhouettes and then a local shape from stereo is used to conclude the concave parts of surface, but, ambiguity issue during stereo match is not resolved.

We fuse shape from silhouettes and shape from stereo essentially using bounding edge model; propose a two-step reconstruction policy and avoid the stereo match process. First, samples of object surface can be obtained through resolving a 1D line-based variance minimum problem, then an energy model is proposed to reconstruct a smooth surface from the samples based on a variation level set method and a Partial Differential Equation (PDE) is thus deduced. We propose a fast method to construct the initial surface, which saves the temporal cost of reconstruction. Level set method is used to resolve the initial value problem of the PDE, where we discretize space by Weighted Essentially Non-Oscillatory (WENO) technique and discretize time by Total Variation Diminishing (TVD) Runge-Kutta method. A TVD Runge-Kutta method with varying time steps is proposed to reinitialize the signed distance function, which guarantees the upwind design of each Euler step. Evolution process of the PDE can shrink to target surface by controlling a revisory coefficient in the energy model.

SAMPLING OBJECT SURFACE

Bounding edge model proposed by Cheung *et al.* (2003, 2005) can provide robust samples of object surface. And it has the following Fundamental Property of Visual Hull (FPVH):

FPVH: Each bounding edge of the visual hull touches the objects at least one point.

The FPVH allows us to use bounding edges to store and represent the key shape information of the object. Thus, we can combine the FPVH and color stereo on the bounding edges to extract 3D points on the surface of objects, called Colored Surface Points (CSPs) of the object:

- **Step 1:** Capture some images (internal and extern parameters of the camera are calibrated) at different view-points
- **Step 2:** Extract silhouettes of the object from the images obtained from pre-step
- **Step 3:** Search a point on the boundary of a silhouette image, project it into 3D space through the corresponding camera center and we can get a ray; select the part on the ray whose projection lies completely inside the silhouettes for any other image plane as a bounding edge
- **Step 4:** Search the parameterized bounding edge and select the point which has the minimum variance of projected color in all the color images as a CSP and the variance must be less than a appointed threshold α
- **Step 5:** If the samples on the object surface called CSPs are enough, then go to end; else go the step 3

Although, increasing the number of samples on the boundary of silhouette images, the surface can be represented more completely. In fact, due to noises, inaccuracies in color balancing we felicitously select a threshold α to define CSPs based on the method of Harten *et al.* (1987); if we select a higher threshold α, the more CSPs will be obtained with a higher noises level, conversely, if we select a lower threshold α, noises level can be satisfied, but we can only obtain the less CSPs.

ENERGY MODEL OF OBJECT SURFACE

Supposing points set $P = \{p_i; i = 1,..., N\}$ comprised of all the CSPs obtained from the pre-section, for any point in 3D space $x \in R^3$, we define:

$$d(x) = dist(x,P) = \min_{1\le i\le N} dist(x,p_i) \quad (1)$$

where, dist (x, p_i) is the Euclidean distance between x and p_i. So energy model can be defined as:

$$E(\Gamma) = (1-\alpha)\int_\Gamma d^2(x)dA + \alpha\int_\Gamma dA \quad (2)$$

where, α, we call it revisory coefficient here, can control the model filter ability and smoothing reconstructed surface. Thus target integration surface is depicted as:

$$\Gamma(t) = \{x(t) \in R^3 : \bar{d}(x,t) = 0\} \quad (3)$$

where, the level set function $\bar{d}(x,t)$ is the sign distance function:

$$\bar{d}(x,t) = \begin{cases} -d(x,t) = -dist(x,\Gamma(t)), & x \in \Omega(t) \\ 0, & x \in \Gamma(t) \\ d(x,t) = dist(x,\Gamma(t)), & x \in R^3 \setminus \Omega(t) \end{cases} \quad (4)$$

where, $\Omega(t)$ represents the domain surrounded by $\Gamma(t)$.

Using the Zhao (1996) result, geometry flow can be constructed as:

$$\frac{\partial \bar{d}}{\partial t} = (d^2 + \alpha)\mathrm{div}(\frac{\nabla \bar{d}}{\| \nabla \bar{d} \|}) \cdot \| \nabla \bar{d} \| + 2(1-\alpha)\nabla(d^2) \cdot \nabla \bar{d} \quad (5)$$

The first item of energy Eq. 2 represents position potential energy, through minimizing it, the target integration surface Γ(t) can shrink to CSPs step by step; the second item represents elasticity potential energy, through minimizing it, the CSPs noises can be filtered and the target integration surface Γ(t) can be smoothed. Thus, the threshold α is the weight of elasticity potential energy in energy model; and it control the filter ability of level reconstruction algorithm, it's range is (0, 1), in experience, we select it in $0.0 \le \alpha \le 0.3$.

CONSTRUCTING INITIAL SURFACE

To construct the initial surface, distance function in Eq. 5 must be resolved firstly. We calculate distance between any point in 3D space and the points set P by solving the Eikonal function, that is the boundary problem $\|\nabla d(x)\| = 1$, $d(x \epsilon P) = 0$. Using the equal distance mesh, that is, $\Delta x = \Delta y = \Delta z = h$, we select $d(x) = \varepsilon$ as the initial surface, where, $\varepsilon = 1.2\ h$ generally. We construct the initial sign distance function with following algorithm, which saves the temporal cost of reconstruction:

- **Step 1:** Any point in 3D mesh space is tagged as tag = 1
- **Step 2:** The point which absolutely lies out of target surface is tagged as tag = 1 and its neighbors which are tagged as tag = -1 are putted into the heap
- **Step 3:** The point in the heap having the maximal distance is found out, if it has a neighbor whose distance is less than ε, then it and its all neighbors in the heap are tagged as tag = 0 and go to step 4; else it is tagged as tag = -1 and is deleted from the heap, the neighbor which is tagged as tag = -1 and is not in the heap is putted into the heap. Go to step 3
- **Step 4:** If tag (x), then $\bar{d}(x) = 0$; else if tag (x) = 1, then $\bar{d}(x) = d(x) - \varepsilon$; else $\bar{d}(x) = -(d(x) - \varepsilon)$

SOLVING DIFFERENTIAL EQUATION

Equation 5 is consisted of two items which depend on the different numeric calculating domain. For decreasing numeric vibration and the better resolution, different discrete schemes are used for them for:

$$L(\bar{d}) = (d^2 + \alpha)\mathrm{div}(\frac{\nabla \bar{d}}{\| \nabla \bar{d} \|}) \| \nabla \bar{d} \| \quad (6)$$

Center-difference method with 2-order accuracy is used to calculate the derivative. For the Hamilton-Jacobi part $H(\nabla \bar{d}) = 2(1-\alpha)\nabla(d^2)\nabla \bar{d}$, upwind design policy is used, that $\bar{d}_x$ uses its left derivative $\bar{d}_x^-$ or right derivative $\bar{d}_x^+$ depends on the sign of $(d^2)_x$. If $(d^2)_x > 0$, $\bar{d}_x$, uses the left derivative $\bar{d}_x^-$; if $(d^2)_x < 0$, $\bar{d}_x$ uses the right derivative $\bar{d}_x^+$. For $\bar{d}_y, \bar{d}_z$, the similar processes are used. Unilateral derivative is calculated through WENO technique proposed by Shu (1997).

During the iterations, level set maybe deposit to a small domain, indeed shrink to nothing. Therefore, it is need to reinitialize the sign distance function repeatedly, that is, keep the zero-value surface and replace the current sign distance function by another level set function with the better property.

Since, target surface description do not depend on the special level set function, we solve the Eq. 7 with a varying time steps TVD Runge-Kutta technique we called, which guarantees upwind design of each Euler step:

$$\begin{cases} \dfrac{\partial \bar{d}}{\partial t} = \mathrm{sign}(\bar{d}_0)(1 - \| \nabla \bar{d} \|) \\ \bar{d}(x,0) = \bar{d}_0(x) \end{cases} \quad (7)$$

where, sign is the sign function, Following Harten *et al.* (1987), $\mathrm{sign}(\bar{d}_0)$ is approximated dynamically by:

$$s(\bar{d}) = \frac{\bar{d}}{(\bar{d}^2 + \| \nabla \bar{d} \|^2 h^2)^{1/2}}$$

At the mesh point (ih, jh, kh), the discrete iteration scheme of the function (7) is:

$$\hat{\bar{d}}_{ijk}^{n+1} = \bar{d}_{ijk}^n + (\Delta t)_1^n L(\bar{d}_{ijk}^n) \quad (8)$$

$$\hat{\hat{\bar{d}}}_{ijk}^{n+2} = \hat{\bar{d}}_{ijk}^{n+1} + (\Delta t)_2^n L(\hat{\bar{d}}_{ijk}^{n+1}) \quad (9)$$

$$\bar{d}_{ijk}^{n+1} = \gamma^n \bar{d}_{ijk}^n + (1-\gamma^n) L(\hat{\hat{\bar{d}}}_{ijk}^{n+2}) \quad (10)$$

where, $\bar{d}_{ijk}^n$ and $\bar{d}_{ijk}^{n+1}$ represent the value of $\bar{d}_{ijk}$ at n, n+1 time step, respectively. The first and second Euler time step $(\Delta t)_1^n$ and $(\Delta t)_2^n$ are selected satisfying the CFL condition like as:

$$\begin{cases} (\Delta t)_1^n \le \dfrac{1}{3} \dfrac{h}{\max_{i,j,k} | p_{ijk}^n |} \\ (\Delta t)_2^n \le \dfrac{1}{3} \dfrac{h}{\max_{i,j,k} | (p_{ijk}^n)_1 |} \end{cases} \quad (11)$$

Where,

$$p_{ijk}^{n} = \frac{\bar{d}_{ijk}^{n}}{((\bar{d}_{ijk}^{n})^2 + \| \nabla(\bar{d}_{ijk}^{n}) \|^2 \, h^2)^{1/2}}$$

$$(p_{ijk}^{n})_1 = \frac{\hat{\bar{d}}_{ijk}^{n+1}}{((\hat{\bar{d}}_{ijk}^{n+1})^2 + \| \nabla(\hat{\bar{d}}_{ijk}^{n+1}) \|^2 \, h^2)^{1/2}}$$

The combination coefficient in Eq. 10 is:

$$\gamma^{n} = \frac{((\Delta t)_1^{n})^2 + ((\Delta t)_2^{n})^2}{((\Delta t)_1^{n} + (\Delta t)_2^{n})^2} \qquad (12)$$

RESULT AND DISCUSSION

The experiments are performed in PC (CPU: Intel Pentium Double core 2.0 GHz; RAM: 4096 MB; OS: Ubuntu Linux 8.10). C++ Language is used to descript the algorithm proposed by this study.

Controlling the position of a powerful light source and the rotation angle ϖ of the non-convex object, 72 images were captured under various and complex illumination, sixteen of them show as Fig. 1. The size of image is 640×480 or 480×640 pixels. Where, in Fig. 1a, the powerful light source is in the left of object, thus the shadow lies in the right part; in Fig. 1b, the powerful light

Fig. 1: Input images. (a) The powerful light source is in the left, ϖ = 0°, 90°, 180°, 270°, (b) the powerful light source is in the right, ϖ = 30°, 120°, 210°, 300°, (c) the powerful light source is in the top,ϖ = 0°, 90°, 180°, 270° and (d) the powerful light source is in the top, ϖ = 30°, 120°, 210°, 300°. The images data come from our capturing process under specified illumination

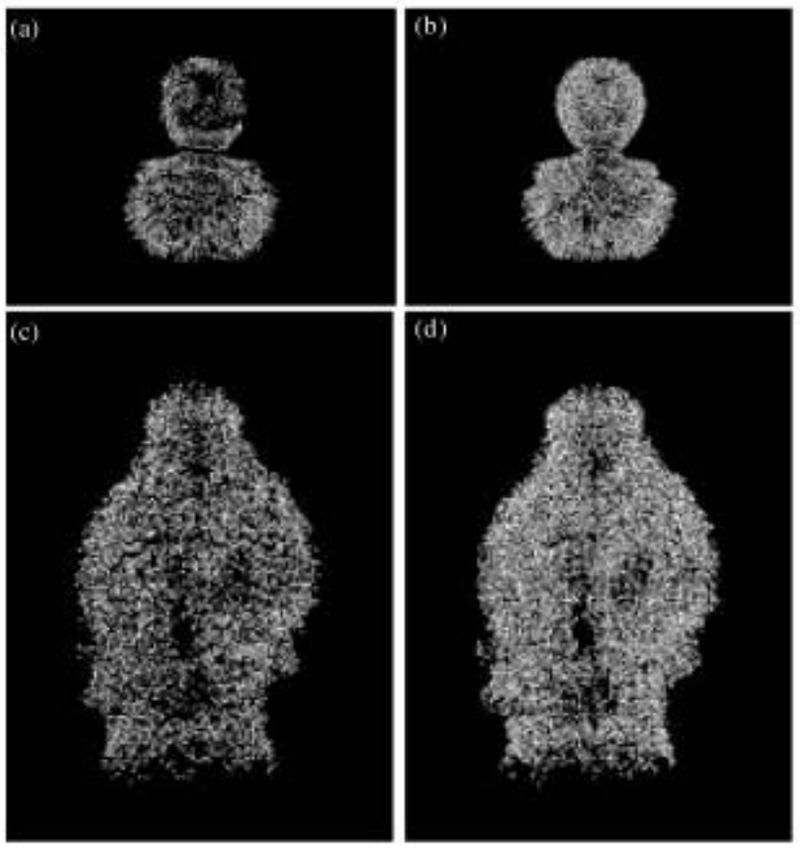

Fig. 2: CSPs distribution. (a) 4089 CSPs, (b) 10654 CSPs, (c) 5431 CSPs and (d) 10859 CSPs. The image data come from our object surface sampling process under two noises level

source is in the right of object, thus the shadow lies in the left part and in Fig. 1c and d, the powerful light source is in the top of object, thus no shadow lies there.

We divide input images into two groups according to different object in image, Duck or Buddha. For the first group, 4089 CSPs were obtained when color variance threshold $\alpha = 0.1$ and 18 images were used; 10654 CSPs were obtained when color variance threshold $\alpha = 0.15$ and 36 images were used. For the second group, 5431 CSPs were obtained when color variance threshold $\alpha = 0.125$ and 18 images were used; 10859 CSPs were obtained when color variance threshold $\alpha = 0.25$ and 36 images were used. Here α represents the different noise level. The CSPs show locally the profile of the object, distribution of CSPs is shown as Fig. 2a-d.

At the four different noise level, Marching Cube algorithm is used to visualize the sign distance field and OpenGL is used to implement the 3D representation. The two initial surfaces reconstructed from the two CSPs set are shown as Fig. 3a-d.

As you see, the initial surfaces are un-accuracy and coarse, then begin the level set iteration process, while searching narrow-band is set to 12 h and the sample noise is filtered, as result, the sign distance function shrink to CSPs set step by step. The reconstruction results are shown as Fig. 4a-d.

Time-consuming of our algorithm is show as Table 1, where, the time of sampling object surface, time of initialize the surface and time of surface moving are illustrated.

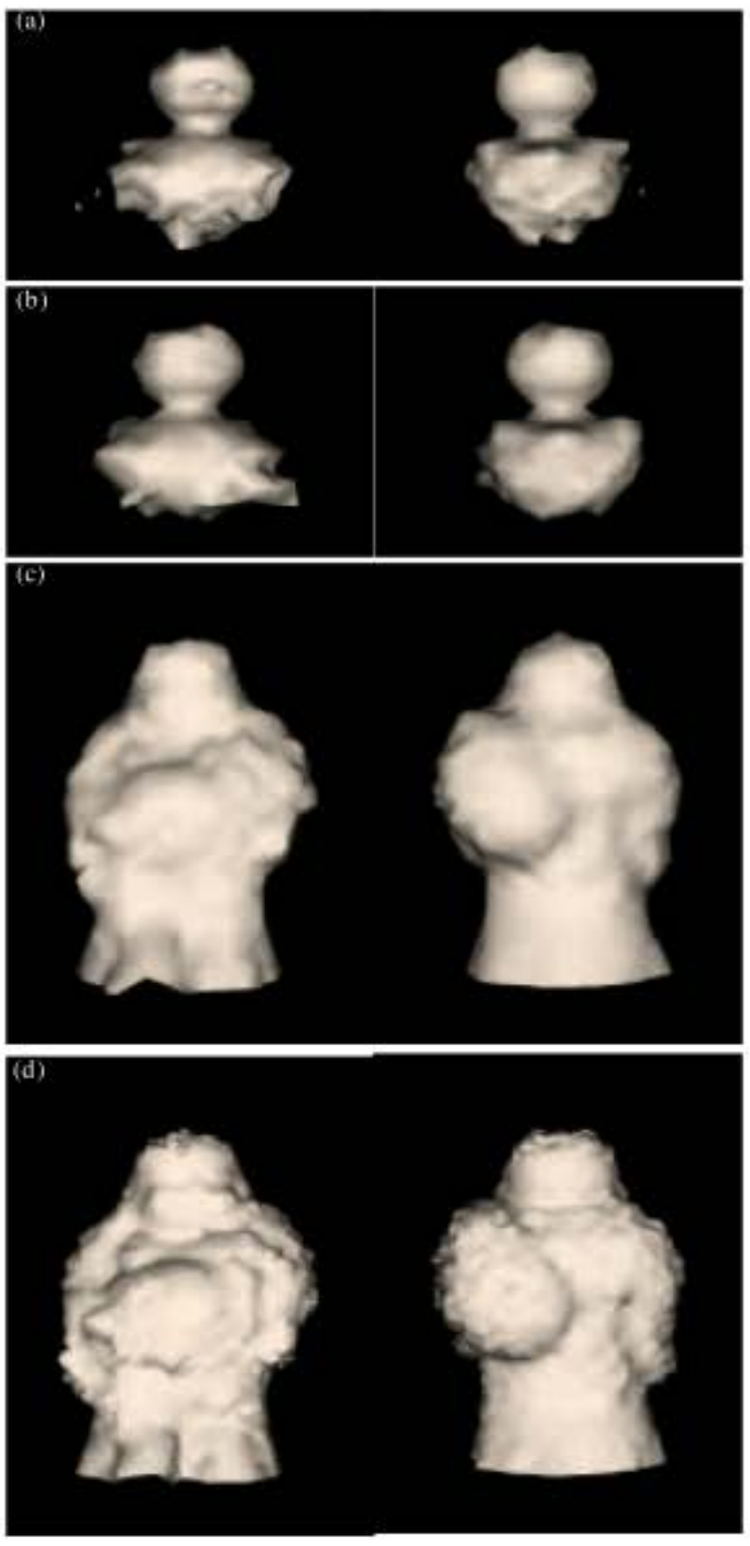

Fig. 3: Initial surface models. (a) 4089 CSPs, (b) 10654 CSPs, (c) 5431 CSPs and (d) 10859 CSPs. The images data come from our initial surface reconstruction using the CSPs set

Table 1: Time-consuming of our algorithm

No. of Images	α	No. of samples	Time of sampling (sec)	Time of reconstruction (sec)		Total time (sec)
				Initialization	Shrink	
18	0.100	4089	6.92	0.82	1.04	8.78
36	0.150	10654	16.54	1.27	2.56	20.37
18	0.125	5431	9.19	1.09	1.36	11.64
36	0.250	10859	16.83	1.29	2.60	20.72

The data in table come from our sampling and reconstruction experiments

To illustrate our algorithm performance, we had implemented Jin *et al*. (2008) algorithm at the same time. Under the same input conditions, the reconstruction results are shown as Fig. 5a-d. As with Jin's algorithm, the object surface can be reconstructed with smaller number of input image and lower noise level; a bigger error occur with bigger number of input image and higher noise level.

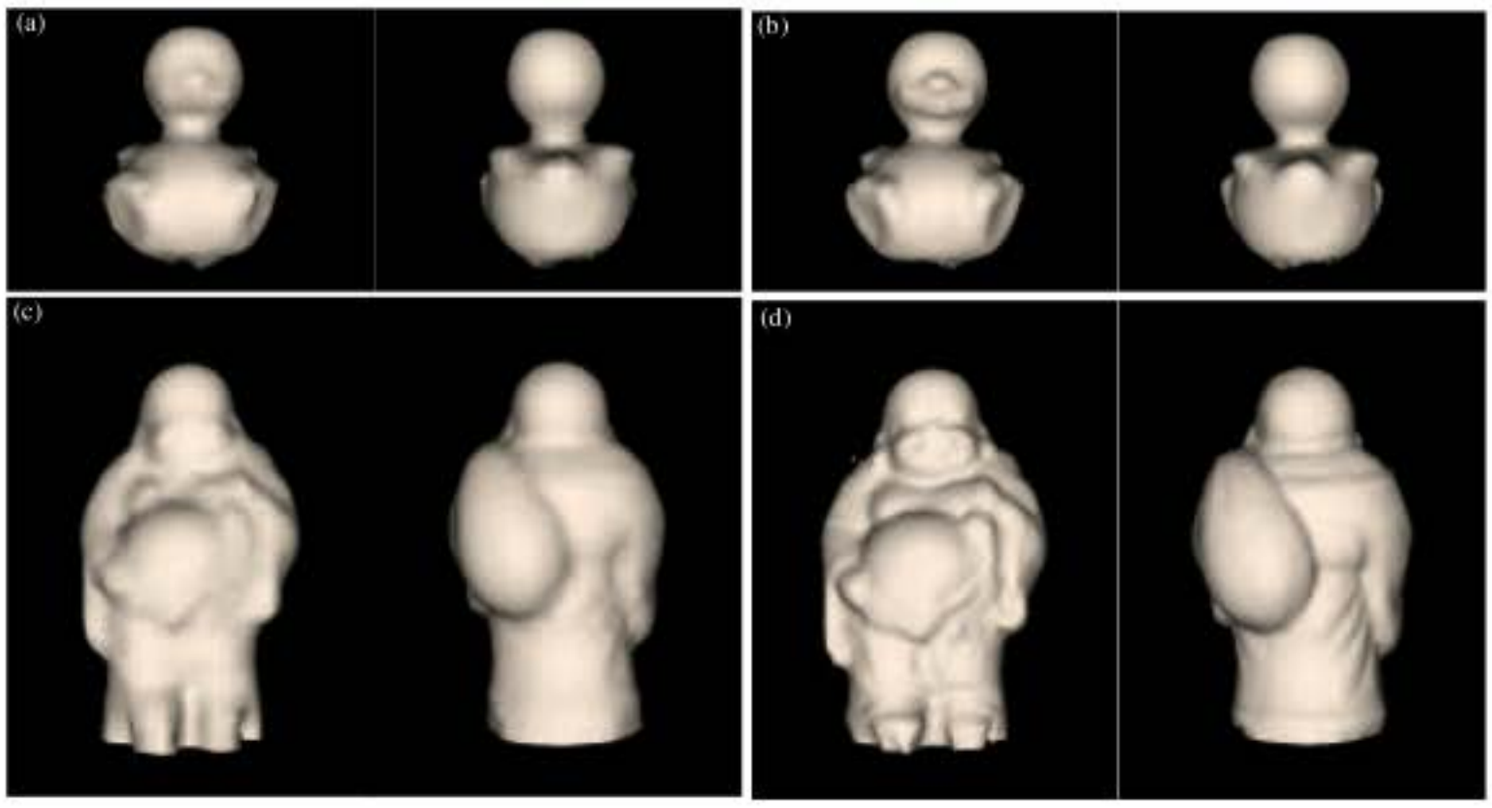

Fig. 4: Reconstruction results of our algorithm. (a) 18 images, (b) 36 images, (c) 18 images and (d) 36 images. The images come our reconstruction algorithm implementation

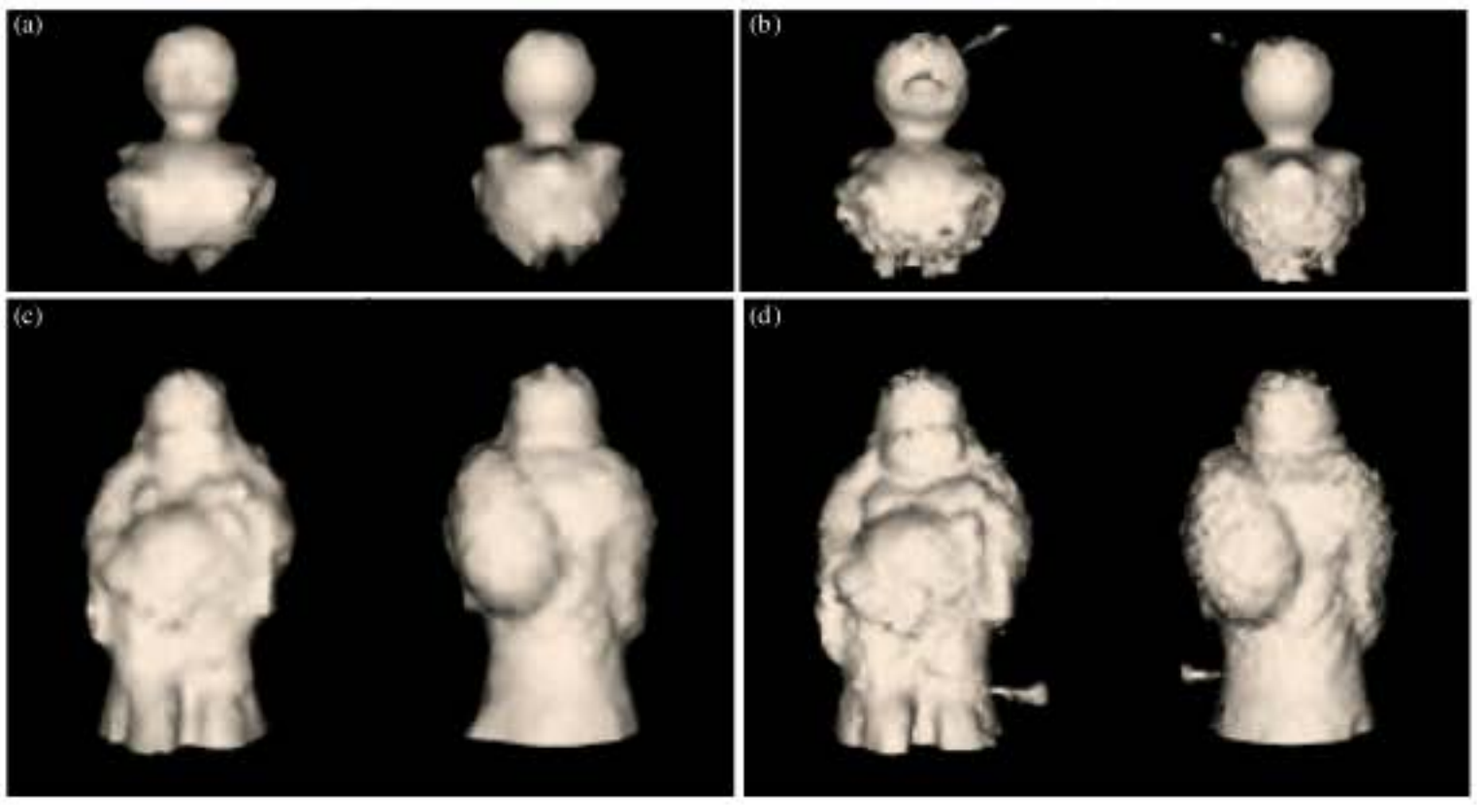

Fig. 5: Reconstruction results of Jin algorithm. (a) 18 images, (b) 36 images, (c) 18 images and (d) 36 images. The images data come from Jin algorithm implementation

Table 2: Time-consuming comparison between our algorithm and Jin's algorithm

No. of images	α	Jin's algorithm (sec)	Our algorithm (sec)	Saved time (%)
18	0.100	14.32	8.78	38
36	0.150	35.34	20.37	42
18	0.125	18.19	11.64	36
36	0.250	35.12	20.72	41

The data in the table come from our algorithm and Jin algorithm implementation

Comparison table of the both algorithm time-consuming is shown in Table 2. Our algorithm saved 36~42% reconstruction time than Jin's algorithm.

CONCLUSION

A two-step reconstruction method was proposed, firstly bounding edge model was used to obtain CSPs and then level set method was used to reconstruct surface of object. A technique called varying time steps TVD Runge-Kutta was used to guarantee upwind design at each Euler step. Result of experiment show that the algorithm we proposed could turn from the calibrated images to the model of object surface with accuracy and robustness under complex illumination. Future work in this field may

include more sophisticated handing of non-Lambertian scenes, new methods for reconstruction from un-calibrated images and more computationally efficient methods for real-time reconstruction.

ACKNOWLEDGMENT

This study is supported by Chinese National Natural Science Foundation (No. 60602050).

REFERENCES

Cheung, G.K.M., S. Baker and T. Kanade, 2003. Visual hull alignment and renfinement across time: A 3D reconstruction algorithm combining shape-from-silhouette with stereo. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Monona Terrace Convention Center, Madison, Wisconsin, Jun. 16-22, IEEE Computer Society, pp: 375-382.

Cheung, G.K.M., S. Baker and T. Kanade, 2005. Shape from silhouette across time part I: Theory and algorithms. Int. J. Comput. Vision, 62: 221-247.

Dyer, C.R., 2001. Volumetric Scene Reconstruction from Multiple Views, in Foundations of Image Understanding. Norwell, Kluwer, ISBN: 0-7923-7457-6, pp: 469–489.

Ehsan, A., P. Jean-Philippe and P. Florent, 2007. Spatio-temporal shape from silhouette using four-dimensional delaunay meshing. Proceedings of the 7th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, Oct. 14-21, IEEE Computer Society, pp: 1-8.

Esteban, C.H. and F. Schmitt, 2003. Silhouette and stereo fusion for 3D object modeling. Proceedings of the 4th International Conference on 3D Digital Imaging and Modeling, Banff, Alberta, Cananda, Oct. 6-10, IEEE Computer Society, pp: 46-53.

Harten, A., B. Engquist and S. Osher, 1987. Uniformly high order accurate essentially non-oscillatory schemes, III. J. Comput. Phys., 71: 231-303.

Hirschmuller, H. and D. Scharstein, 2007. Evaluation of cost functions for stereo matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis MN, Jun. 17-22, IEEE Computer Society, pp: 1-8.

Jin, H., D. Cremers and D. Wang, 2008. 3-D reconstruction of shaded objects from multiple images under unknown illumination. Int. J. Comput. Vision, 76: 245-256.

Kanade, T. and M. Okutomi, 1994. A stereo matching algorithm with an adaptive window: Theory and experiment. IEEE Trans. Pattern Anal. Mach. Intell., 16: 920-932.

Kang, S., R. Szeliski and J. Chai, 2001. Handling occlusions in dense multi-view stereo, computer vision and pattern recognition. CVPR2001, Kauai Marriott Hawaii, 1: 102-110.

Scharstein, D., R. Szeliski and R. Zabih, 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision, Dec. 9-10, Kkauai, Hawaii, USA., pp: 131-140.

Shu, C.W., 1997. Technical report on essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolie conservation laws. Hamptom: NACA Langlcy Research Center. Page: 206-235.

Slabaugh, G., B. Culbertson, T. Malzbender and R. Schafer, 2001. A survey of methods for volumetric scene reconstruction from photographs. Proceedings of the International Workshop on Volume Graphics, June 2001, Stony Brook, New York, USA., pp: 81-100.

Zhao, H., 1996. A variation level set approach to multiphase motion. J. Comput. Phys., 127: 179-195.