ROBUST AND UNBIASED FOREGROUND / BACKGROUND ENERGY FOR MULTI-VIEW STEREO

Zhihu Chen and Kwan-Yee K. Wong
Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong

Keywords: Multi-view stereo, Photo-consistency Energy, Foreground / Background energy, Graph-cuts.

Abstract: This paper revisits the graph-cuts based approach for solving the multi-view stereo problem, and proposes a novel foreground / background energy which is shown to be unbiased and robust against noisy depth maps. Unlike most existing works which focus on deriving a robust photo-consistency energy, this paper targets at deriving a robust and unbiased foreground / background energy. By introducing a novel data-dependent foreground / background energy, we show that it is possible to recover the object surface from noisy depth maps even in the absence of the photo-consistency energy. This demonstrates that the foreground / background energy is equally important as the photo-consistency energy in graph-cuts based methods. Experiments on real data sequences further show that high quality reconstructions can be achieved using our proposed foreground / background energy with a very simple photo-consistency energy.

1 INTRODUCTION

Multi-view stereo (MVS) is a key technique in computer vision for reconstructing a dense 3D geometry of an object from images taken around it. It has many applications such as preservation of arts, animation, and augmented reality. Many research works on MVS have therefore been carried out in the past few decades. This results in a huge pool of sophisticated algorithms (Seitz et al., 2006).

In this paper, we are going to revisit the graph-cuts based approach for solving the multi-view stereo problem. Being one of the most popular MVS algorithms, the graph-cuts based approach has been receiving a lot of attentions in recent years (Vogiatzis et al., 2005; Kolmogorov and Zabih, 2002; Lempitsky et al., 2006; Tran and Davis, 2006; Sinha and Pollefeys, 2005; Ladikos et al., 2008; Vogiatzis et al., 2007; Hernández et al., 2007). Graph-cuts based methods generally solve the problem by defining a foreground / background energy for each voxel in a discretized volume, and a photo-consistency energy between adjacent voxels. A specialized graph G is then constructed with each voxel defining a node in this graph. There are also two additional nodes, namely the source s and the sink t representing the foreground and background respectively, in G which are connected to all other nodes. The photo-consistency energy is used to define the weights for the links between adjacent voxel nodes, whereas the foreground / background energy is used to define the weights for the links between the voxel nodes and the sink t / source s. The object surface is estimated by minimizing the photo-consistency energy associated with the surface and the foreground / background energy associated with the (foreground / background) label of each voxel. This corresponds to finding the s-t min-cut of G which partitions G into two parts, namely S and T, with a minimal cost such that $s \in S$ and $t \in T$.

Unlike most existing works which focus on deriving a robust photo-consistency energy, this paper targets at deriving a robust and unbiased foreground / background energy. It has been noted in previous works that the foreground / background energy is important in preserving both protrusions and concavities in the reconstructed surface. By introducing a novel data-dependent foreground / background energy, we show that it is possible to recover the object surface from noisy depth maps *even in the absence of the photo-consistency energy*. To the best of our knowledge, this is the first time that a reconstruction is achieved without using a photo-consistency energy in graph-cuts. This demonstrates that the foreground / background energy is as important as the photo-consistency energy in graph-cuts based methods. Experiments on real data sequences further show that high quality reconstructions can be achieved using our proposed foreground / background energy with a

Chen Z. and K. Wong K..
ROBUST AND UNBIASED FOREGROUND / BACKGROUND ENERGY FOR MULTI-VIEW STEREO.
DOI: 10.5220/0003308604510458
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 451-458
ISBN: 978-989-8425-47-8

very simple photo-consistency energy.

The rest of the paper is organized as follows. Section 2 gives a brief literature review on graph-cuts based MVS methods. Section 3 describes our proposed algorithm in detail. In particular, a novel unbiased data-dependent foreground / background energy is introduced. In Section 4, experimental results on real data sequences as well as evaluation results are presented. Finally, Section 5 concludes our main contributions.

2 RELATED WORK

Kolmogorov and Zabih (Kolmogorov and Zabih, 2002) were amongst the first to formulate the multi-view stereo problem as an energy minimization problem, and reconstruct the 3D object by solving the minimization problem using graph-cuts. They proposed an energy formulation which could handle the visibility problem and impose spatial smoothness while preserving discontinuity. In (Vogiatzis et al., 2005), Vogiatzis et al. handled the visibility problem by exploiting the visual hull to approximate the visibility of voxels. They also introduced a *uniform ballooning term* to avoid the elimination of protrusions in the reconstruction. In (Sinha and Pollefeys, 2005), Sinha et al. enforced the silhouette constraints while minimizing the photo-consistency energy and the smoothness term. Lempitsky et al. (Lempitsky et al., 2006) estimated visibility based on the positions and orientations of local surface patches, and used graph-cuts to minimize the photo-consistency energy and the uniform ballooning term on a CW-complex. In (Tran and Davis, 2006), Tran and Davis added a set of predefined locations as constraints in graph-cuts to improve the performance. In (Vogiatzis et al., 2007), Vogiatzis et al. used Parzen window method to compute the depth maps robustly, and formulated the photo-consistency energy using a voting scheme (Esteban and Schmitt, 2004) based on these depth maps. In (Hernández et al., 2007), Carlos et al. proposed a data-dependent *intelligent ballooning term* based on the probability of invisibility of a voxel. The use of the intelligent ballooning can solve the over-inflated problem caused by the use of a data-independent *uniform ballooning term*.

Most of the aforementioned methods focus on tackling the visibility problem in the computation of the photo-consistency energy (Kolmogorov and Zabih, 2002; Lempitsky et al., 2006; Vogiatzis et al., 2007). Only two (Vogiatzis et al., 2005; Hernández et al., 2007) of them consider the foreground / background energy. In (Vogiatzis et al., 2005), Vogiatzis et al. pointed out that the energy-minimizing surface might suffer from a lack of protrusions present in the object if only the photo-consistency energy is considered. They therefore introduced the *uniform ballooning term* which favors a large volume inside the visual hull. Such a ballooning term is in fact a special form of the foreground / background energy. It only defines a background energy inside the visual hull, and the foreground energy is simply set to zero. Voxels inside the visual hull are therefore biased to be in foreground. By including this term in the energy function, protrusions in the object can then be reconstructed. However, depending on the weights assigned to this term, it may also result in an over-inflated reconstruction. Besides, the visual hull of the object may not be always available, especially in a complex background. In (Hernández et al., 2007), Carlos et al. formulated an intelligent ballooning term based on the overall probability of invisibility of a voxel, and their method can reconstruct both protrusions as well as concavities in the object. However, as the overall probability of invisibility of a voxel is computed as the product of its probabilities of invisibility in individual views, such a ballooning term is not robust. For instance, if the probability of invisibility of a voxel in one view is inaccurately calculated due to image noise, its overall probability of invisibility will be seriously affected. Besides, such a ballooning term is also biased. If the probability of invisibility of a voxel is small in one view, its overall probability of invisibility will become small, and therefore the voxel is biased to be in the background.

In this paper, instead of proposing yet another robust photo-consistency energy, we target at deriving a novel foreground / background energy that is both *unbiased* and *robust* against noisy depth maps. We believe that the foreground / background energy plays an equally important role as the photo-consistency energy in graph-cuts based methods. In fact, by using our proposed robust and unbiased foreground / background energy, we will demonstrate later in this paper that it is possible to reconstruct an object without even using the photo-consistency energy term. This further strengthens our belief in the importance of the foreground / background energy. This also means that a robust foreground / background energy can actually compensate the errors caused by the inaccuracy of the photo-consistency energy.

3 ALGORITHM DESCRIPTION

The input to our method is a sequence of images $I = \{I_1, I_2, \cdots, I_N\}$ taken around an object, together

with a set of the corresponding camera projection matrices $P = \{P_1, P_2, \cdots, P_N\}$, and a bounding box for the object. Note that we also refer $\{I_1, I_2, \cdots, I_N\}$ to as different views. Like other graph-cuts based methods, we formulate the 3D surface as an energy function, and solve the reconstruction problem by minimizing the energy function using graph-cuts. Our energy function E consists of three parts, namely the photo-consistency energy E_{surf}, the foreground energy E_{fore}, and the background energy E_{back}, and is given by

$$E(S) = E_{surf}(S) + E_{fore}(V(S)) + E_{back}(\overline{V}(S)), \quad (1)$$

where S denotes the object surface, $V(S)$ the object volume enclosed by S (also refers to as the foreground volume), and $\overline{V}(S)$ the background volume. In the following subsections, we will describe each of these energy terms in detail.

3.1 Photo-consistency Energy

The photo-consistency energy for a given surface S is defined as

$$E_{surf}(S) = \iint_S \rho(x)\, dA, \quad (2)$$

where $\rho(x)$ is a dissimilarity measure used to determine the degree of dissimilarity of a point x as observed in different views. The greatest challenge in the computation of $E_{surf}(S)$ is the problem of visibility. We need to determine the set of images I_{vis} ($I_{vis} \subseteq I$) in which a point x is visible. In (Vogiatzis et al., 2005), this problem is tackled by utilizing the visual hull of the object to approximate the visibility of nearby points. However, such an approximation will fail disgracefully in regions of concavities on the object surface. Besides, this approach cannot deal with self-occlusions and the visual hull might not be always available.

In this paper, we adopt a robust voting scheme as described in (Esteban and Schmitt, 2004) to compute $E_{surf}(S)$. Concretely, for each 3D point x inside the bounding box of the object, each view I_i will cast a vote $VOTE_i(x)$ for x. All the $VOTE_i(x)$ are then combined together using the formula

$$\rho(x) = e^{-\mu \sum_{i=1}^{N} VOTE_i(x)}. \quad (3)$$

To compute $VOTE_i(x)$ for a 3D point x inside the bounding box, we march along the corresponding optic ray

$$o_i(d) = c_i + \frac{x - c_i}{\|x - c_i\|} d \quad (4)$$

inside the bounding box, where c_i is the camera center for view I_i, and d is the depth value for a 3D point along the optical ray in view I_i (x corresponds to the depth value d_x). For each depth value d, we project the corresponding 3D point $o_i(d)$ onto a set $\mathcal{N}(i)$ of M closest views, and compute the NCC values using a window of size $m \times m$ centered at the projections of $o_i(d)$ on I_i and $I_{j \in \mathcal{N}(i)}$ with sub-pixel accuracy. We then combine these M NCC values into a single score $\mathcal{C}(d)$ and cast a vote to x using the following formula

$$VOTE_i(x) = \begin{cases} \mathcal{C}(d_x) & \text{if } \mathcal{C}(d_x) \geq \mathcal{C}(d)\ \forall d \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

There are different methods for calculating $\mathcal{C}(d)$. In (Esteban and Schmitt, 2004), $\mathcal{C}(d)$ is simply the average of the M NCC values. In (Vogiatzis et al., 2007), Vogiatzis et al. pointed out that the global maximum may not be necessarily corresponding to the correct depth, and they used a Parzen window technique to combine all the local maxima of the NCC values. As the focus of this paper is the foreground / background energy, we simply define $\mathcal{C}(d)$ as the average value of the largest $M/2$ NCC values. Although this strategy will result in noisy depth maps, experimental results show that high quality reconstructions can be achieved using the robust and unbiased foreground / background energy introduced in the next subsection.

3.2 Foreground/Background Energy

If the energy function only consists of the photo-consistency energy, protrusions and concavities in the object will be removed in the reconstructed surface. In order to prevent this situation, the energy function must include the foreground / background energy. The foreground energy of a 3D point x is the cost of assigning x to the foreground, and the background energy of x is the cost of assigning x to the background. In (Vogiatzis et al., 2005), the foreground / background energy is formulated as a uniform ballooning term in which $E_{fore}(V(S)) = 0$ and $E_{back}(\overline{V}(S)) = b \iiint_{\overline{V}(S)} dV$, where b is a weight parameter. This can be considered as the shape prior of the object which favors a large volume. However, voxels inside the visual hull are biased to be in the foreground with such a data-independent uniform ballooning term, and therefore it cannot deal with deep concavities in the object.

In (Hernández et al., 2007), Hernández et al. proposed a data-aware intelligent ballooning term based on the overall probability of invisibility of a 3D point. This formulation is theoretically correct, but will be very sensitive to noise in practice. Theoretically speaking, the overall probability of invisibility of a 3D point x inside the object should be close

to one. However, due to image noise, the "correct depth" for the ray passing through x may be incorrectly estimated in some views. The probability of invisibility of x in those views may become close to zero, making its overall probability of invisibility also close to zero. x will therefore be biased to be in the background. For this reason, the intelligent ballooning term in (Hernández et al., 2007) is not robust to noise and could only work in the case when the depth maps computed from different views are very accurate. This in turn implies a very robust photo-consistency energy is needed. This, however, is very difficult to achieve in practice.

As voting scheme has been proven to be robust to noise, we propose to use voting scheme to formulate the foreground / background energy. Intuitively, if a 3D point $V1$ is associated with a depth d_{V1} in view I_i which is larger than the correct depth d_{corr}, it is likely that $V1$ will be invisible in I_i. Meanwhile, if a 3D point $V2$ is associated with a depth d_{V2} in I_i which is smaller than d_{corr}, it is likely that it will be visible in I_i (see Fig. 1). In this paper, we approximate the correct depth by the estimated depth. In order to utilize the visibility information, each view will cast a vote for the 3D point x as follows:

$$B\_VOTE_i(x) = \begin{cases} 1 & \text{if } d_x < d_{corr} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The foreground / background energy of a 3D point x is then defined as

$$energyForeground(x) = 1 - e^{-\lambda \sum_{i=1}^{N} B\_VOTE_i(x)} \quad (7)$$

$$energyBackground(x) = e^{-\lambda \sum_{i=1}^{N} B\_VOTE_i(x)} \quad (8)$$

Note that $energyForeground(x) + energyBackground(x) = 1$. Finally, the foreground / background energy for the surface is formulated as

$$E_{fore}(V(S)) = b \iiint_{V(S)} energyForeground(x)dV \quad (9)$$

$$E_{back}(\overline{V}(S)) = b \iiint_{\overline{V}(S)} energyBackground(x)dV \quad (10)$$

where b is a weight parameter, and a value between 0.1 to 0.2 is suitable for all our experiments. λ is also a constant value, and it mainly depends on the number of views in the dataset. The proposed foreground / background energy is unbiased. Whether x is more probable to be in the foreground or background depends on the values of $energyForeground(x)$ and $energyBackground(x)$. If $energyForeground(x)$ is smaller, x is more probable to be in the foreground. Otherwise, it is more probable to be in the background. Moreover, if the depths for x in some views are incorrect, it will only influence the votings for x from those views. Therefore, the proposed foreground / background energy is also more robust to image noise.

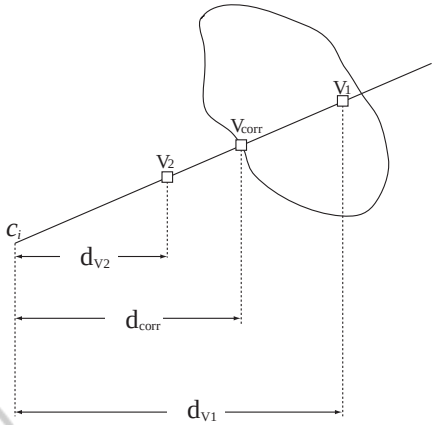


Figure 1: Consider a ray originated from the camera center c_i of view I_i that passes through the bounding box. Along this ray, the score $C(V_{corr})$ for voxel V_{corr} is a maximum, and hence the depth corresponding to V_{corr} is the estimated correct depth, denoted by d_{corr}. Voxel $V1$ has a depth d_{V1} which is larger than d_{corr}. $V1$ is therefore deemed to be invisible in view I_i as it is occluded by the voxel V_{corr}. On the contrary, voxel $V2$ has a depth d_{V2} which is smaller than d_{corr}. $V2$ is therefore deemed to be visible in view I_i. In this case, $B\_VOTE_i(V2) = 1$ and $B\_VOTE_i(V1) = 0$.

4 EXPERIMENTAL RESULTS

In the following subsections, implementation details of the proposed algorithm and experimental results will be described in detail. Note that, unlike the method presented in (Vogiatzis et al., 2005), the algorithm proposed in this paper does not require using the visual hull as an initialization.

4.1 Graph Structure

A 3D space slightly larger than the bounding box is quantized into voxels of size $h \times h \times h$. As illustrated in Fig. 2, a graph G is constructed with each voxel defining a node. Two nodes are connected if their corresponding voxels are 6-neighbor of each other. G also includes two additional nodes, namely the source s which is fixed in the foreground and the sink t which is fixed in the background. All voxel nodes are connected to both s and t. The weight of the link between a voxel node x_i and s is defined as $b \times energyBackground(x_i)$ and the weight of the link between x_i and t is defined as $b \times energyForeground(x_i)$. If voxel node x_i and voxel node x_j are connected, the weight of the link is defined as

$$w_{ij} = \rho\left(\frac{x_i + x_j}{2}\right). \quad (11)$$

Moreover, if the center of a voxel is outside the bounding box, the weight of the link between the correspond node and t is reset to infinity and the weight of the link between the correspond node and s is reset to zero.

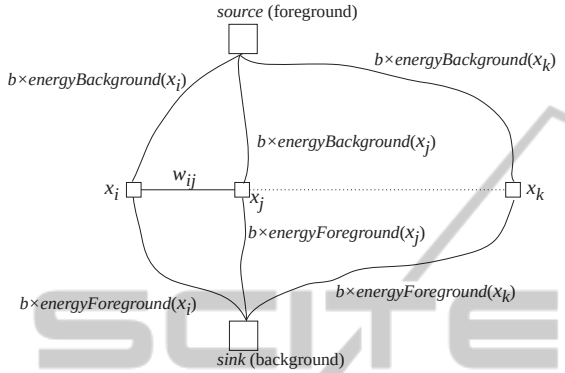


Figure 2: Graph structure for MVS problem. Each node in the graph represents a voxel. Two nodes are connected if the corresponding voxels are 6-neighbor of each other. The weight of the link between two connected nodes is defined by the dissimilarity measure $\rho(x)$ computed at the midpoint between the centers of the corresponding voxels. All nodes are also connected to two special nodes, namely the source and the sink, with weights of the links defined as $b \times energyBackground(x_i)$ and $b \times energyForekground(x_i)$ respectively.

4.2 Results and Evaluations

The graph-cuts algorithm proposed in (Boykov and Kolmogorov, 2004) is used to segment the graph into two parts: *source* part (foreground) and *sink* part (background). Marching cube algorithm (Lorensen and Cline, 1987) is then used to generate a triangulated mesh from the foreground voxels, and Taubin smooth (Taubin, 1995) is used to smooth the resulting mesh.

We have applied our algorithm to several datasets on a system with Intel(R) Core(TM) 2 Duo CPU E6750 @ 2.66GHZ 2.67GHZ and 8GM RAM. We have also submitted our reconstruction results of two standard datasets, namely the *temple sequence* and *dinosaur sequence*, to the evaluation website (Seitz et al.,) for comparison. The results are compared in terms of accuracy and completeness. The accuracy metric used is defined as the distance d such that 90% of the reconstruction is within d mm of the ground truth, while the completeness metric used is defined as the percentage X such that X% of the ground truth is within 1.25 mm of the reconstructed result. Table 1 shows the evaluation results. Details of the sequences used and the reconstruction results are discussed in the following subsections.

Temple. The *temple sequence* was captured in Standford Spherical Gantry and is available from the website (Seitz et al.,). It consists of 312 images and the image resolution is 640×480. The *templeRing sequence* is a sparse sequence sampled from the *temple sequence* and consists of only 47 images. Both sequences were used in our experiments.

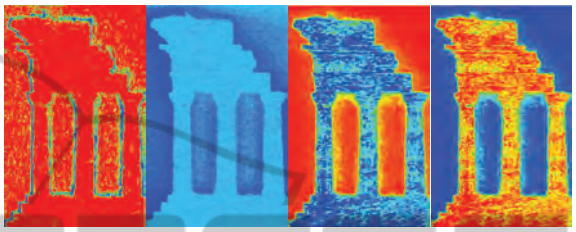


Figure 3: Comparison between the proposed foreground / background energy and the intelligent ballooning term proposed in (Hernández et al., 2007). From left to right: (a) photo-consistency energy, (b) intelligent ballooning term, (c) proposed foreground energy, and (d) proposed background energy. Blue color indicates a low energy value and red color indicates a high energy value. Note that the foreground energy values are lower inside the object and higher outside the object. The background energy values, on the other hand, are higher inside the object and lower outside the object.

Fig. 3 compares our foreground / background energy with the intelligent ballooning term[1] proposed in (Hernández et al., 2007). In this figure, blue color indicates a low energy value and red color indicates a high energy value. Fig. 3(a) shows a plot of the photo-consistency energy for a particular cross section, which reveals that the original depth maps are quite noisy. Fig. 3(b) shows a plot of the intelligent ballooning term for the same cross section. It can be observed that, although still separable, energy values inside the temple are very close to those in the background. Besides, the space in the upper right corner as well as that between the pillars are contaminated by noises. Fig. 3(c) and (d) show our proposed foreground and background energy, respectively, for the same cross section. It can be seen that energy values inside the temple are significantly different from

[1] In (Hernández et al., 2007), the probability density function of invisibility in each view is computed by $pdf(d) = \alpha * Uniform(d, dMin, dMax) + (1-\alpha) * Gaussian(d, mean_d, sigma_d)$, where $dMin, dMax$ are the depths of intersections of the optic ray with the bounding box, $mean_d$ is the measured depth, $sigma_d$ corresponds to the length of 1 pixel back-projected in 3D, and α is a predefined constant outlier ratio.

Table 1: Evaluation results for our reconstructions of the *templeRing sequence*, *temple sequence* , and *dinoRing sequence*. The accuracy value is the distance d such that 90% of the reconstruction is within d mm of the ground truth. The completeness value is the percentage X such that X% of the ground truth is within 1.25 mm of the reconstructed result.

Name	Images	Image Size	accuracy(mm)	completeness(%)
templeRing	47	640×480	0.54	99.6
temple	312	640×480	0.47	99.3
dinoRing	48	640×480	0.46	95.1

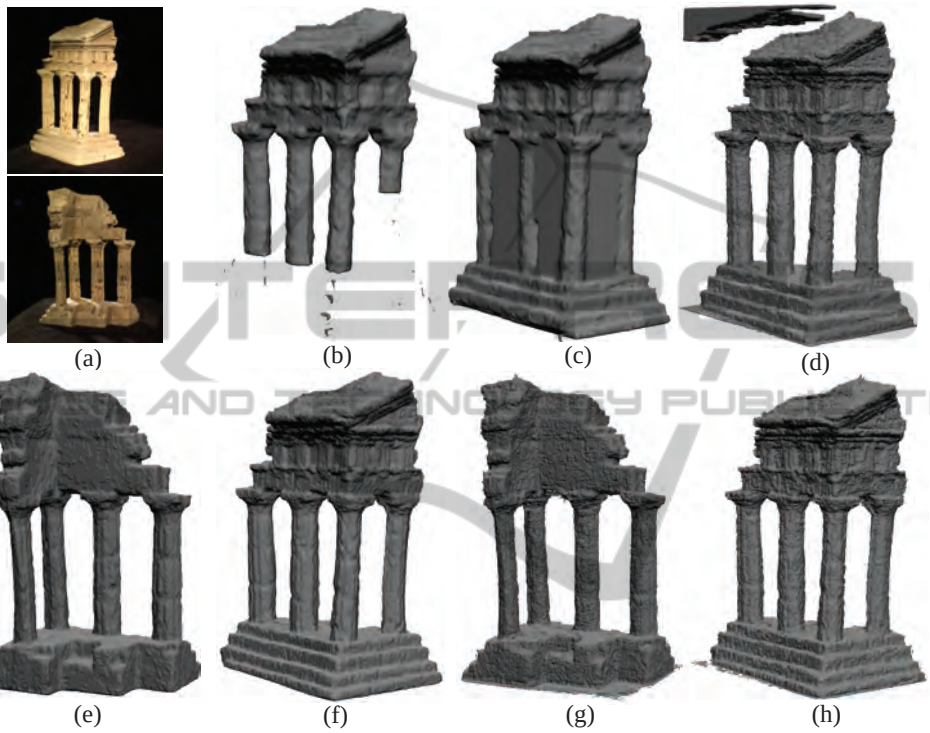


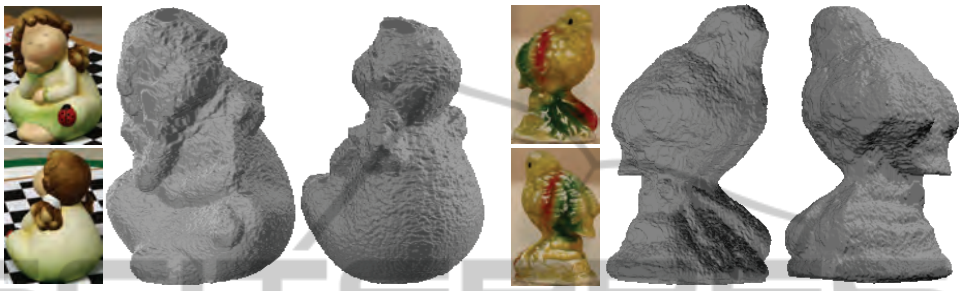
Figure 4: (a) Two images from the *templeRing sequence*. Reconstruction results using (b) photo-consistency energy with a small uniform ballooning term, (c) photo-consistency energy with a large uniform ballooning term, (d) photo-consistency energy with intelligent ballooning term, (e,f) photo-consistency energy with the proposed foreground / background energy, and (g, h) only the proposed foreground / background energy.

those in the background, and this allows an excellent segmentation of the foreground and background.

Fig. 4 shows the reconstruction results of the *templeRing sequence* using photo-consistency energy with a small uniform ballooning term (Fig. 4(b)), photo-consistency energy with a large uniform ballooning term (Fig. 4(c)), photo-consistency energy with intelligent ballooning term (Fig. 4(d)), photo-consistency energy with the proposed foreground / background energy (Fig. 4(e) and (f)), and only the proposed foreground / background energy (Fig. 4(g) and (h)), respectively. From Fig. 4(b) and (c), it can be seen that a small uniform ballooning term would result in an incomplete model, whereas a large uniform ballooning term would cause an over-inflated model. Using the intelligent ballooning term in place of the uniform ballooning term could alleviate this problem, and the temple was reconstructed with details. Nonetheless, it can be seen in Fig. 4(d) that there are some errors near the top and bottom of the reconstruction, which can be explained by the plot of the intelligent ballooning term shown in Fig. 3(b). The best reconstruction result was obtained using photo-consistency energy with the proposed foreground / background energy (see Fig. 4(e) and (f)). The roof, the stairs and the concavities at the bottom of the temple were all accurately reconstructed. The accuracy of the reconstruction is 0.54 mm and the completeness is 99.6%. To further test the usefulness of the proposed foreground / background energy, we have carried out a reconstruction using only the proposed foreground / background energy. Not surprisingly, good result was achieved since the foreground / background energy alone allows an excellent segmentation of the fore-

Table 2: Comparison between the reconstruction results of the standard *temple sequence* and *templeRing sequence* obtained by (Vogiatzis et al., 2005), (Vogiatzis et al., 2007) and the proposed method.

Method	*temple*(312)		*templeRing*(47)		Initialization
	acc.[*mm*]	comp.[%]	acc.[*mm*]	comp.[%]	
Proposed	**0.47**	**99.3**	**0.54**	**99.6**	Bounding Box
Vogiatzis (Vogiatzis et al., 2005)	1.07	90.7	0.76	96.2	Visual Hull
Vogiatzis (Vogiatzis et al., 2007)	0.5	98.4	0.64	99.2	Visual Hull



(a) (b)

Figure 6: Reconstruction results of (a) *girl sequence*, and (b) *bird sequence*.

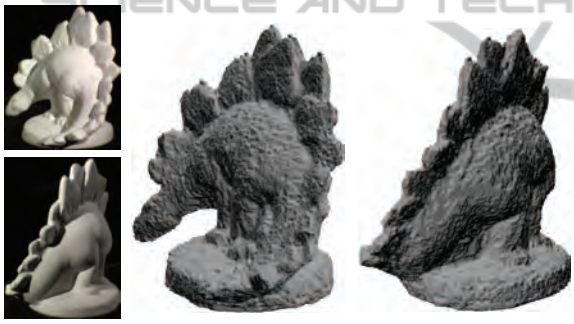


Figure 5: Two images from the *dinoRing sequence* and the reconstruction result rendered in two different viewpoints.

ground and background (see Fig. 3(c) and (d)).

We have also carried out a reconstruction of the *temple sequence* using photo-consistency energy with the proposed foreground / background energy. Table 2 compares our reconstruction results of the *temple sequence* and *templeRing sequence* with that obtained by (Vogiatzis et al., 2005) and (Vogiatzis et al., 2007), both of which use a uniform ballooning term. Although a bounding box instead of the visual hull is used in our algorithm, our results are still better than that obtained by (Vogiatzis et al., 2005) and (Vogiatzis et al., 2007) in terms of accuracy and completeness.

Dinosaur. The *dinoRing sequence* consists of 48 images and the image resolution is 640×480. This sequence is relatively more difficult than the *temple sequence* due to the lack of features on the body of the dinosaur. Nonetheless, the proposed algorithm still produced good result. Fig. 5 shows two images from the sequence and the reconstruction result rendered in two different viewpoints. The accuracy of the reconstruction result is 0.46 mm and the completeness is 95.1%.

Girl and Bird. Both the *girl sequence* and the *bird sequence* consist of 60 images taken around the object The *girl sequence* contains specular reflections and is lack of features, especially in the face, hands and feet. The *bird sequence*, on the other hand, has many protrusions and concavities at the bottom part of the bird model. Our algorithm could recover both objects with details (see Fig. 6).

5 CONCLUSIONS

This paper revisits the graph-cuts based approach for solving the multi-view stereo problem. Unlike most existing works which focus on deriving a photo-consistency energy that is robust against invisibility, this paper targets at deriving a robust and unbiased foreground / background energy. A voting scheme is adopted to derive a novel data-dependent foreground / background energy which is shown to be unbiased, robust against noisy depth maps, and allows an excellent segmentation of the foreground and background. In fact, we have demonstrated that it is possible to recover the object surface using only the proposed foreground / background energy (i.e., without using the photo-consistency energy term in graph-cuts). This demonstrates that the foreground / background energy

is equally important as the photo-consistency energy in graph-cuts based methods. Experiments on real data sequences further show that high quality reconstructions can be achieved using our proposed foreground / background energy with a very simple photo-consistency energy.

REFERENCES

Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 1124–1137.

Esteban, C. H. and Schmitt, F. (2004). Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, pages 367–392.

Hernández, C., Vogiatzis, G., and Cipolla, R. (2007). Probabilistic visibility for multi-view stereo. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 1–8.

Kolmogorov, V. and Zabih, R. (2002). Multi-camera scene reconstruction via graph cuts. In *Proc. European Conf. on Computer Vision*, pages 82–96.

Ladikos, A., Benhimane, S., and Navab, N. (2008). Multi-view reconstruction using narrow-band graph-cuts and surface normal optimization. In *Proc. British Machine Vision Conference*, pages 143–152.

Lempitsky, V. S., Boykov, Y., and Ivanov, D. V. (2006). Oriented visibility for multiview reconstruction. In *Proc. European Conf. on Computer Vision*, pages 226–238.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, pages 163–169.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. Multi-view stereo evaluation web page. http://vision.middlebury.edu/mview/.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 519–528.

Sinha, S. N. and Pollefeys, M. (2005). Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. *Proc. Int. Conf. on Computer Vision*, pages 349–356.

Taubin, G. (1995). A signal processing approach to fair surface design. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 351–358.

Tran, S. and Davis, L. (2006). 3d surface reconstruction using graph cuts with surface constraints. In *Proc. European Conf. on Computer Vision*, pages 219–231.

Vogiatzis, G., Hernández Esteban, C., Torr, P. H. S., and Cipolla, R. (2007). Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 2241–2246.

Vogiatzis, G., Torr, P. H. S., and Cipolla, R. (2005). Multi-view stereo via volumetric graph-cuts. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 391–398.