Multi-view Stereo via Volumetric Graph-cuts

G. Vogiatzis[1]   P. H. S. Torr[2]   R. Cipolla[3]

Dept. of Engineering, University of Cambridge, Cambridge,CB2 1PZ, UK[13]

Dept. of Computing, Oxford Brookes University, Wheatley, Oxford OX33 1HX, UK[2]

`{gv215`[1]`, cipolla`[3]`}@eng.cam.ac.uk   philiptorr@brookes.ac.uk`[2]

Abstract

This paper presents a novel formulation for the multi-view scene reconstruction problem. While this formulation benefits from a volumetric scene representation, it is amenable to a computationally tractable global optimisation using Graph-cuts. The algorithm proposed uses the visual hull of the scene to infer occlusions and as a constraint on the topology of the scene. A photo consistency-based surface cost functional is defined and discretised with a weighted graph. The optimal surface under this discretised functional is obtained as the minimum cut solution of the weighted graph. Our method provides a viewpoint independent surface regularisation, approximate handling of occlusions and a tractable optimisation scheme. Promising experimental results on real scenes as well as a quantitative evaluation on a synthetic scene are presented.

1   Introduction

This paper considers the problem of reconstructing the dense geometry of a 3D scene from a number of images, calibrated for pose and intrinsic parameters. This is a computer vision problem that has been extensively studied. Work on the field can be categorised according to the mathematical representation of the scene geometry. There are two main classes of techniques according to this categorisation: (1) techniques that recover depth-maps with respect to an image plane and (2) volumetric methods that represent the volume directly, without any reference to an image plane.

In the first class of methods, a reference image is selected and a disparity or depth value is assigned to each of its pixels using a combination of image correlation and regularisation. An excellent review for image based methods can be found in Scharstein and Szeliski [13]. These problems are often formulated as minimisations of Markov Random Field (MRF) energy functions providing a clean and computationally-tractable formulation, for which good ap-proximate solutions exist using Graph cuts [2, 7, 12, 6] or Loopy Belief Propagation [16]. They can also be formulated as continuous PDE evolutions on the depth maps [15]. However, a key limitation of these solutions is that they can only represent depth maps with a unique disparity per pixel, i.e. depth is a function of image point. Capturing complete objects in this manner requires further processing to merge multiple depth maps [10] which is a complicated and error-prone procedure. A second limitation is that the smoothness term imposed by the MRF is defined on image disparities or depths and hence is viewpoint dependent– if a different view is chosen as the reference image the results may be different.

The second class comprises of methods that use a *volumetric representation* of shape. This is a more natural framework in which multiple viewpoints can be easily integrated and surface smoothness can be enforced independent of viewpoint. In this class are well-known techniques like Space Carving [9] and level set stereo [5]. While these methods are known to produce high quality reconstructions their convergence properties in the presence of noise are not well understood and are susceptible to instabilities or local minima. For Space Carving in particular, there is also no simple way to impose surface smoothness.

The approach described in this paper combines the advantages of both classes described above. We adopt a volumetric scene representation but pose the reconstruction problem as finding the minimum cut of a weighted graph. This computation is exact and can be performed in polynomial time. Our method requires an approximate base surface, obtained from the visual hull. This is used as the source of occlusion information and as a hard constraint on the topology of the scene surface. The benefits of our approach are the following:

1. General surfaces and objects can be fully represented and computed as a single surface.

2. The representation and smoothness constraint is image and viewpoint independent.

3. Multiple views of the scene can be used with occlusions approximately modelled.

4. Optimisation is computationally tractable, using existing max-flow algorithms.

1.1 Related work

The inspiration for the approach presented in this paper is the recent work of Boykov and Kolmogorov [1] which establishes a theoretical link between maximum flow problems in discrete graphs and minimal surfaces in an arbitrary Riemannian metric. In particular the authors show how a continuous Riemannian metric can be approximated by a discrete weighted graph so that the max-flow/min-cut solution for the graph corresponds to a local geodesic or minimal surface in the continuous case. The application presented in that paper is interactive 2D or 3D image segmentation where the user is asked to approximately specify the object and background regions in an image which then become the *source* and *sink* sets of the graph. The segmentation is then obtained by computing the minimum cut between the two sets. The work presented here extends these ideas to multi-view stereo reconstruction, offering solutions to some of the difficulties that occur within this domain, namely the determination of the source and sink sets as well as occlusion reasoning.

Another related approach is the use of level sets [5] for stereo reconstruction. In that work a 3D surface, represented as the zero level set of a 3D scalar field, is evolved using continuous PDE techniques, until it is photo-consistent with a number of images. While level sets can represent arbitrary surface topologies, the resulting optimisation methods give local minima of the energy function, which can be sensitive to initialisation. This work poses the reconstruction problem as a computation of maximum flow in a discrete graph, for which global optimisation methods exist.

Space carving [9] is a technique that starts from a volume containing the scene and greedily carves out non photo-consistent voxels from that volume until all remaining visible voxels are consistent. It uses a discrete representation of the surface but does not enforce any smoothness constraint on the surface which often results in quite noisy reconstructions. Furthermore, as all voxel removal decisions affect subsequent ones, it is very conservative in carving out voxels which implies that reconstructed surfaces tend to be *fatter* than the true scene. Also using a discrete quantisation of space, Snow *et al* [14] showed how visual hull extraction from silhouettes can be cast as a binary MRF problem which can solved exactly with a minimum cut computation. Unfortunately, even though it seems similar, the type of graph used in that work cannot be applied to the problem of shape from stereo because of the problem of occlusion which is not present when dealing with silhouettes.

Finally, recent work by Paris *et al* [11], proposes the use of a global Graph cut optimisation to minimise a discretised version of a continuous functional for surface reconstruction. While we also use Graph cuts to discretise a similar continuous problem, their technique differs in two aspects: (i) it assumes that a plane separates the cameras from the scene and (ii) represents the scene as multiple depth-maps. This means that a full circumnavigation of an object, such as the sequence of the third experiment shown in section 4, which does not satisfy either of these requirements, cannot be reconstructed.

In previous work [17] we argued for the use of the base surface for occlusion reasoning and placing topological constraints on the true surface. In that work, a discrete height map above the base surface is optimised with a Belief Propagation algorithm, and while promising results are presented, the height map representation is weak in regions of high curvature or corners in the base surface. In this work we provide a new method that alleviates this problem.

The rest of the paper is laid out as follows: Section 2 describes in detail how the scene surface is represented as an interface between two boundary surfaces. In section 3 we describe the cost functional associated with any candidate surface, as well as how this functional is approximated with a discrete flow graph. Section 4 presents experimental results on synthetic and real scenes, section 5 compares of our method with level-sets and section 6 concludes with discussion of the paper's main contributions.

2 Graph-cuts for volumetric stereo

This paper extends the Riemannian minimal surface idea of [1] for multi-view volumetric stereo. In that work points in space are successfully labelled as foreground and background while also regularising the interface between the two volumes. In the multi-view volumetric stereo domain the corresponding labelling problem is to decide if points in space are inside or outside the scene that is being reconstructed. Two main difficulties immediately arise. Firstly, the problem of occlusion, i.e. the fact that distant points in space might occlude each other, implies that the state of each point in space (inside/outside the scene) is affected by potentially distant points. This does not naturally fit into the Graph-cut minimisation framework. Secondly, it is not obvious how to define connections to the source and sink, which corresponds to defining a likelihood for each point being inside or outside the scene.

To overcome both these difficulties, a *base surface* [17] is used. This is a surface that captures the approximate geometry of the scene and *contains* the scene. The visual hull, which is the intersection of the cones generated by the silhouettes of the scene is an ideal example of such an approximate surface. In situations where this cannot be obtained
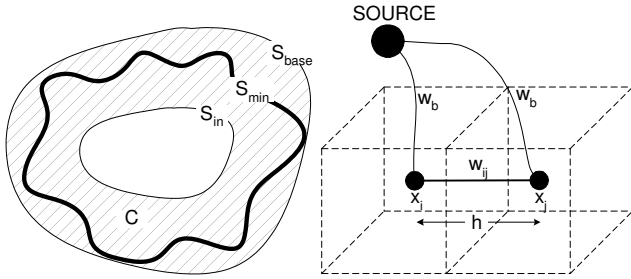
Figure 1. *Surface geometry and flow graph construction. On the left: a 2D slice of space showing the base surface S_{base} and the inside boundary surface S_{in}. The shaded region is the volume C of voxels that become nodes in the flow graph. The thick line is the true surface of the scene S_{min} that is represented as the minimum cut in the flow graph. On the right: the correspondence of voxels in C with nodes in the graph. Each voxel is connected to its neighbours as well as to the source.*

(e.g. the scene cannot be circumnavigated) then it may be sufficient to estimate a small number of sparse image correspondences and triangulate them to obtain an approximate mesh (we include such a scene in the results section).

Let the base surface be denoted by S_{base} and define for all $\mathbf{x} \in \mathbb{R}^3$ the signed distance function $d(\mathbf{x})$ as the distance from $\mathbf{x}$ to the closest point on S_{base}, positive if $\mathbf{x}$ lies in the direction of the surface normal at the closest point and negative otherwise. The inner boundary surface for some positive constant D_{in} can be defined as:

$$S_{in} = \left\{ \mathbf{x} \in \mathbb{R}^3 : d(\mathbf{x}) = -D_{in} \right\} \tag{1}$$

and the volume between the two is just

$$C = \left\{ \mathbf{x} \in \mathbb{R}^3 : -D_{in} \leq d(\mathbf{x}) < 0 \right\}. \tag{2}$$

The surface model used in this work will require the minimal surface S_{min} to lie between S_{in} and S_{base} so that $S_{min} \subseteq C$. The geometric configuration of the base, inner and minimal surfaces is shown in figure 1 (left). The next subsection introduces the cost functional associated with a candidate surface, which will subsequently be minimised.

3 Surface cost functional

The input to our method is a sequence of images $I_1, ..., I_N$ calibrated for camera pose and intrinsic parameters. The photo-consistency of a potential scene point $\mathbf{x}$ can be evaluated by comparing its projections in the images where it is visible. If the point is on the true surface, these projections should be conforming with some surface model assumptions. Determining the visibility of a point is therefore a central part of this calculation, since, if a point is occluded in an image, its projection would violate any model

assumptions. Using the base surface for obtaining visibility information is one of the contributions of this work.

The visibility of a point $\mathbf{x}$ on surface S is represented by $\mathcal{V}(\mathbf{x}, S)$, the set of images from which $\mathbf{x}$ would be visible if the scene consisted of a surface S.

In [9] it was shown that if S_{base} contains a surface S and $\mathbf{x}$ is on S_{base} then $\mathcal{V}(\mathbf{x}, S)$ is a superset of $\mathcal{V}(\mathbf{x}, S_{base})$. It was also shown that if $\mathbf{x}$ is found inconsistent with $\mathcal{V}(\mathbf{x}, S_{base})$, it is also inconsistent with any superset of $\mathcal{V}(\mathbf{x}, S_{base})$. This implies the visibility induced by S_{base} can be used to detect inconsistent 3D points that lie on S_{base}. For a point that lies in the volume C, i.e. not on S_{base}, this corollary is no longer valid. Let $\mathbf{s}(\mathbf{x})$ be the point on S_{base} closest to $\mathbf{x}$. For a small distance D_{in}, $\mathbf{s}(\mathbf{x})$ and $\mathbf{x}$ will in general be quite close to each other, so one can hope that $\mathcal{V}(\mathbf{x}, S)$ will be the same as $\mathcal{V}(\mathbf{s}(\mathbf{x}), S_{base})$. This visibility reasoning even though approximate, has been justified by experiments.

The photo-consistency measure $\rho(\mathbf{x})$ used to determine the degree of consistency of a point $\mathbf{x}$ with the images is computed as follows: For all pairs of images in $\mathcal{V}(\mathbf{s}(\mathbf{x}), S_{base})$, we compute the normalised cross-correlation score between patches centred on the projections of x. For each NCC score, let $\mathbf{V_1}$ and $\mathbf{V_2}$ be the viewing directions from $\mathbf{s}(\mathbf{x})$ to the two cameras and $\mathbf{N}$ the surface normal of S_{base} at $\mathbf{s}(\mathbf{x})$. The NCC scores are more likely to be unreliable (a) due to projective warping for very big baselines and (b) due to violations of the occlusion approximation for viewing angles close to 90 degrees. To avoid these situations we simply exclude an NCC score if $\arccos(\mathbf{V_1} \cdot \mathbf{V_2}) > 45°$ or $\arccos(\mathbf{N} \cdot \mathbf{V_1})) > 60°$ or $\arccos(\mathbf{N} \cdot \mathbf{V_2})) > 60°$. The rest are averaged to produce an average score $c(\mathbf{x})$, lying between -1 and 1. This average is then mapped to the nonnegative interval $[0, 1]$ using

$$\rho(\mathbf{x}) = 1 - \exp\left(-\tan\left(\frac{\pi}{4}(c(\mathbf{x}) - 1)\right)^2 / \sigma^2\right). \tag{3}$$

This is only one of the possible ways to smoothly map $c(\mathbf{x})$ to nonnegative values, parameterising by σ the fidelity of the surface to the data. Determining the optimal such mapping however remains unclear and should be further investigated. Nevertheless the definition of (3) has proved to work well in practice. The cost functional associated with the photo-consistency of a candidate surface S is the integral of $\rho(\mathbf{x})$ on the surface

$$E_{surf}[S] = \iint_S \rho(\mathbf{x}) dA. \tag{4}$$

Surface smoothness is implicitly enforced in (4) since minimising $E_{surf}[S]$ corresponds to finding the minimal surface with respect to a Riemannian metric. Larger values of the parameter σ in (3) lead to a surface that is less smooth but which passes through photo-consistent points
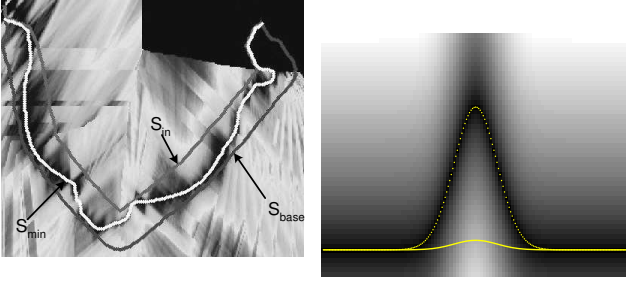
Figure 2. *Slice of Riemannian metric.* *Left: 2D slice of the Riemannian metric corresponding to the photo-consistency cost $\rho(.)$ (with low values being darker) showing intersections of S_{in}, S_{base} and S_{min} for the synthetic face experiment. Right: A slice of a synthetic Riemannian metric. The two lines represent two possible surfaces, the actual surface of the scene (dotted line) and the minimal surface (continuous line) returned as S_{min} by our algorithm. In the real surface the cost per unit area is smaller but the total surface integral of the cost is higher and hence the erroneous but actually lower cost surface is selected.*

while smaller values of σ lead to a smoother surface with smaller Euclidean surface area. The performance of the algorithm is quite stable with respect to the value of σ which was kept constant at 0.05 for all experiments presented in this paper.

A side-effect of minimising Riemannian surface area is that it may result in removing protrusions present in the scene. This is because, while cost per unit area is lower along the protrusion, the total surface integral of the cost may be quite large. At the same time a surface with the protrusion flattened out, may have lower total cost, despite the fact that it passes through high cost regions. Figure 2 (right) shows an example of how this occurs. The phenomenon is also illustrated in one of the experiments (fig. 7). It is worth noting that methods based on geodesic active contours or level sets would face the same difficulty if used to retrieve the global minimum instead of just stopping at the first strong local minimum which is how they are typically used. To counterbalance the protrusion flattening problem, an inflationary (*ballooning*) term is added. The motivation for this type of term in the active contour domain is given in [3], but intuitively, it can be thought of as a shape prior that favours objects that fill the space of the visual hull more, everything else being equal. Let $V(S) \subset C$ be the volume between S_{base} and a candidate surface S. The ballooning term is proportional to the magnitude of the volume $V(S)$:

$$E_{vol}[S] = \lambda \iiint_{V(S)} dV \qquad (5)$$

where λ is a weight parameter. Minimising E_{vol} maximises the magnitude of the volume enclosed by S. The effect of

the E_{vol} term is to inflate the surface, competing with the effect of E_{surf} which is the minimisation of Riemannian surface area. The weight of the ballooning term at the moment has to be selected by hand which implies that care has to be taken to avoid over-inflation. In practice the correct parameter is obtained with just a few trial runs but an automatic mechanism for determining λ is a future research goal. Figure 2 (left) shows a 2D slice of a scalar cost field $\rho(.)$ with dark intensities corresponding to lower costs (i.e. photo-consistent points). The two boundary surfaces S_{in} and S_{base} are shown, as well as S_{min}, the minimal surface separating them.

The reconstructed surface is obtained by solving the optimisation

$$S_{min} = \arg \min_{S \subseteq C} E_{surf}[S] + E_{vol}[S] \qquad (6)$$

which is achieved by embedding the functional in a flow graph that will be described in the next sub-section.

3.1 Graph structure

To obtain a discrete solution to (6) 3D space is quantised into voxels of size $h \times h \times h$. The graph nodes consist of all voxels whose centres are in C, i.e. between the inner boundary and base surfaces. For the results presented in this paper these nodes were connected with a regular 6-neighbourhood grid, but at the expense of using more memory to store the graph, bigger neighbourhood systems can be used which provide a better approximation to the continuous functional (6). Now assume two voxels centred at $\mathbf{x_i}$ and $\mathbf{x_j}$ are neighbours. Then the weight of the edge joining the two corresponding nodes on the graph will be

$$w_{ij} = \frac{4\pi h^2}{3} \rho \left(\frac{x_i + x_j}{2} \right) \qquad (7)$$

where ρ is the matching cost function defined in equation (3). See [1] for the derivation of the weight w_{ij} in the case of a 6-neighbour regular grid. In addition to these weights between neighbouring voxels there is also the ballooning force edge connecting every voxel to the source node with a constant weight of $w_b = \lambda h^3$. Finally, the voxels that are part of S_{in} and S_{base} are connected with the source and sink respectively with edges of infinite weight. The configuration of the graph is shown in figure 1 (right).

It is worth pointing out that the graph structure described above is a simple binary MRF. Variables correspond to voxels and can be labelled as being *inside* or *outside* the scene. The singleton clique potential is just 0 if the voxel is outside and w_b if it is inside the scene while the pairwise potential between two neighbour voxels i and j is a Potts energy equal to w_{ij} if the voxels have opposite labels and 0 otherwise. As a binary MRF with a *regular* energy function [8] it can be solved exactly in polynomial time using Graph-cuts.

Method	MSE (pixels)	% of correct disparities
Space Carving [9]	1.913	69.7%
2-view BP [16]	1.626	74.3%
Relief-surf [17]	0.829	78.6%
Volumetric GC	0.780	79.1%

Table 1. *Quantitative comparison for synthetic scene. Comparison of our method, 2-view belief propagation and [17] against ground truth data. Both methods that use the base surface prior on shape significantly outperform belief propagation and our method marginally outperforms both.*

4 Experiments

To quantitatively analyse the performance of the volumetric Graph-cuts algorithm presented here with ground truth, an experiment on a synthetic scene was performed (fig. 3). A textured model of a human face was rendered from 8 view points. For these 8 images the silhouettes of the face were obtained and the visual hull reconstructed. The performance of volumetric Graph-cuts was tested against an implementation of a 2-view stereo algorithm using Loopy Belief Propagation (LBP) similar to [16] and Space Carving [9]. To compare against a method that also assumes a base surface, the Relief-surface method of [17] was run on the same sequence. The quality of the results of all four methods was estimated in terms of the 7 disparity maps, corresponding to the 7 pairs of consecutive views in the 8 images. Specifically, five sets of the 7 disparity maps were obtained from (i) Space Carving (ii) the 2-view LBP algorithm (iii) Relief-surfaces (iv) Volumetric Graph-cuts and (v) the ground truth. The results shown on Table 1 show the measured mean square error of the four stereo methods against the ground truth, expressed in pixels, as well as the percentage of correct disparities (rounded to one pixel). They indicate that the existence of a base surface as well as the combination of multiple images in a single volumetric framework greatly improves the accuracy of the reconstruction. They also show a slight improvement of the accuracy in the case of the present method against Relief surfaces. However by visually inspecting the results of both these methods (fig. 3) it becomes apparent that the Volumetric Graph-cuts technique is far less dependent on the base surface. A slight misalignment of the real nose of the face with respect to the 'nose' of the visual hull (shown in the cost volume slice of figure 1) results in a 'seam' artifact in the Relief-surface reconstruction, which is not present in the Volumetric Graph-cut surface.

The second experiment (figure 4) involves three images of a real scene of a stone carving and illustrates one possible source of a base surface in cases where the visual hull cannot be obtained. This usually occurs when, as in this ex-
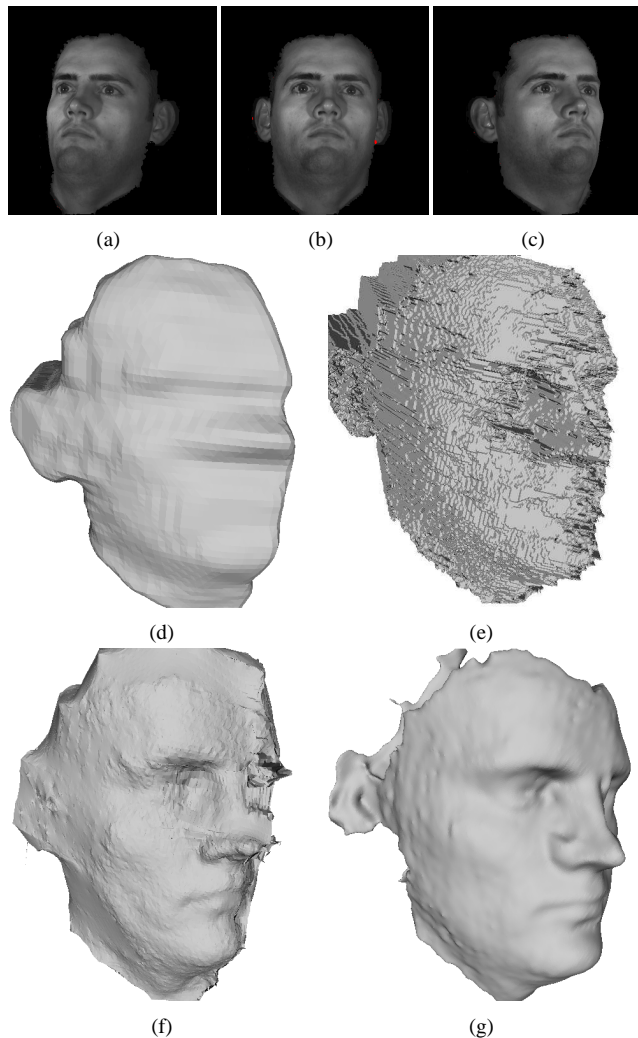

(a) (b) (c)

(d) (e)

(f) (g)

Figure 3. *Face (synthetic scene). (a)-(c) Three images of the synthetic face sequence where a 3D face model has been rendered from 8 viewpoints. (d) The visual hull generated from silhouettes of the face. (e) The result of space carving. (f) The result of the Relief Surface reconstruction from [17] in which the 'seam' artifacts across the face are present. (g) The reconstructed face model using our method, with the concavities of eyes and mouth correctly recovered.*
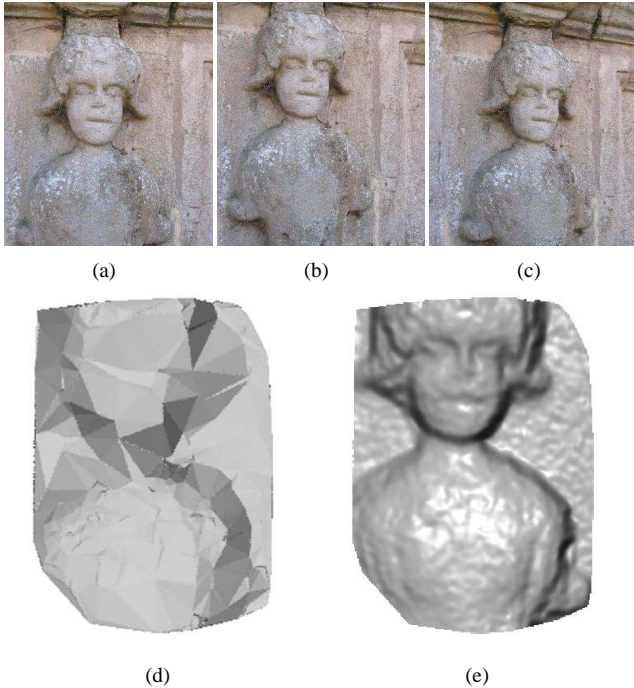
(a)   (b)   (c)

(d)   (e)

Figure 4. *Stone carving (real scene).* *(a)-(c) The three images used for the reconstruction. (d) The base surface obtained by triangulating a sparse set of correspondences. (e) The reconstructed model using our method with details of the face recovered.*

Figure 5. *House (real scene).* *Left: the visual hull (S_{base}) generated from silhouettes extracted from the images. Right: the very noisy result of Space Carving [9].*

ample, the scene cannot be circumnavigated. In this scene a number of feature correspondences are obtained across the three images and Delaunay triangulated in the middle image to obtain a 3D triangular mesh. S_{base} and S_{in} are defined at constant offsets at either side of this mesh. Figure 4 shows the triangular mesh as well as the Volumetric Graph-cuts reconstruction which retrieves the surface details of the scene.

For the third experiment a larger sequence was used, of 36 images circumnavigating a toy house (figures 5 and 6). The silhouettes of the house in all images were obtained using a graph-cut based, interactive segmentation technique similar to the one described in [1]. Figure 5 shows the visual hull, S_{base}, obtained by intersecting the cones generated by the silhouettes. Space is quantised to a $560 \times 460 \times 360$ grid of cubical voxels, 0.4mm in length. Parameter D_{in} is set to 15mm. The reconstructed scene is computed using the method described here in approximately 40 minutes, on a Pentium IV 2.8GHz with 2Gb of RAM. The results are compared against Space Carving [9] operating on a voxel grid of the same quantisation. The result of Space Carving is very noisy because of the lack of regularisation by a surface model. There are also holes which are caused by the accidental carving of a number of voxels which leads
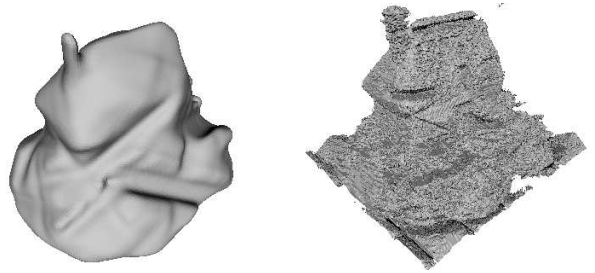
to a cascade-effect of carving out large parts of the volume. In contrast, the surface obtained by Volumetric Graph-cuts captures the correct scene geometry.

Figure 7 gives highlights the significance of the ballooning term. It shows the surface returned by the method when λ, the weight of the term is 0. The result is the flattening of the protrusion corresponding to the house's roof. Even though the cost per unit surface area is smaller along the roof than along the flat surface, the cost saved in total surface area when taking the 'shortcut' and flattening the roof is bigger. If λ is set to 0.8 this tips the balance in favour of the correct surface.

For the fourth and final experiment a sequence of was used, of 41 images circumnavigating a small clay horse (figure 8). In spite of the homogeneous texture of this object our algorithm is still able to recover most of the surface detail.

5   Relation to level set stereo

The cost functional optimised in this paper bears strong resemblance to the functional optimised in level set stereo [4]:

$$E_{level}[S] = \iint_S \rho_{level}(\mathbf{x}, \mathbf{N}(\mathbf{x}), S)dA. \qquad (8)$$

where S is the surface, $\mathbf{N}(\mathbf{x})$ denotes the surface normal at a point $\mathbf{x}$ on the surface and ρ_{level} is the matching cost. This cost uses normalised cross-correlation where the image patches are appropriately warped between viewpoints using the surface normal $\mathbf{N}(\mathbf{x})$. Similarly to our approach, level sets stereo relies on the idea of Riemannian minimal surfaces for regularisation. The differences from our approach are (a) that ρ_{level} depends on the normal to the surface and (b) that the matching cost depends upon the entire surface S. This is necessary to deal with occlusions in the level set framework, because the matching cost depends only on images whose cameras are unoccluded from $\mathbf{x}$ by the current surface S.

Figure 7. *The effect of the 'ballooning' term.* *Left: view of the reconstructed house model without the 'ballooning' term ($\lambda = 0$). Even though the photo-consistency cost $\rho(.)$ is higher per unit surface area along the collapsed roof, the fact that euclidean surface area is considerably smaller means that it is the optimal solution. Right: introducing the ballooning term ($\lambda = 0.8$) counterbalances this effect, forcing the optimal surface to be the real roof.*

Figure 6. *House (real scene).* *Left column: Images of the toy house sequence. Right column: Similar viewpoints of the reconstructed model using our method.*
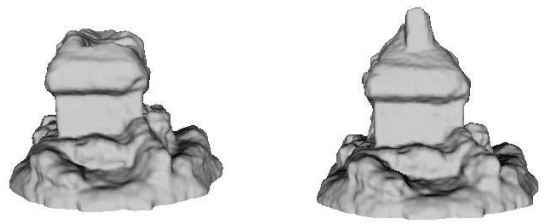

Figure 8. *Clay horse (real scene).* *Left column: Images of the clay horse sequence. Right column: Similar viewpoints of the reconstructed model using our method.*

Point (a) does not seem to be crucial for the success of level sets. In fact, in [4] the authors examine the case of viewpoints with baseline small enough that patch warping can be ignored as in our method. Point (b) however is a significant point of departure between the two models. Our method has no concept of a current surface during the optimisation phase, and therefore has to make the visibility assumption of section 3.

6 Conclusion

This paper has presented a new volumetric formulation of multi-view stereo. A continuous photo-consistency functional is defined on surfaces, and a discrete approximation is formulated as a flow graph. The minimal surface under this functional is obtained by computing the minimum cut solution of the graph. The method uses an approximate base surface obtained from the visual hull of the scene which can be thought of as a coarse prior on shape. This prior is used in two different ways: (i) as a hard constraint, by assuming that the true surface will be between the base surface and a parallel inner boundary surface and (ii) as the source of occlusion information, by assuming that each voxel has the same visibility as the nearest point on the base surface, if that surface was the volume causing occlusions. Furthermore, a prior on the total volume of the reconstructed object is applied which, all else being equal, will put preference on objects that fill the space of the visual hull. This is necessary to counterbalance the effect of the minimisation of Euclidean surface area, which is implicit in the Riemannian minimal surface framework.

The experimental results presented, demonstrate the benefits of combining a volumetric surface representation with a powerful discrete optimisation algorithm such as Graph-cuts, so far only used in depth-map stereo. The resulting method can represent general scene surfaces and provides regularised and globally optimal solutions.

References

[1] Y. Boykov and Kolmogorov V. Computing geodesics and minimal surfaces via graph cuts. In *ICCV 2003*, pages 26–33, 2003.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.

[3] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *PAMI*, 15(11):1131–1147, November 1993.

[4] O. Faugeras, J. Gomes, and R. Keriven. Variational principles in computational stereo. In S. Osher and N. Paragios, editors, *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer-Verlag, 2003.

[5] O. Faugeras and R. Keriven. Variational principles, surface evolution, pdes, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):335–344, 1998.

[6] D. Geiger and H. Ishikawa. Segmentation by grouping junctions. In *Proceedings of CVPR,1998*, pages 125–131, 1998.

[7] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph-cuts. In *ECCV 2002*, volume 3, pages 82–96, 2002.

[8] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *PAMI*, 26(2):147–159, November 2004.

[9] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[10] P.J. Narayanan, P.W. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. In *ICCV98*, pages 3–10, 1998.

[11] S. Paris, F. Sillion, and L. Quan. A surface reconstruction method using global graph cut optimization. In *Proceedings of Asian Conference on Computer Vision*, January 2004.

[12] S. Roy and I. J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *Proceedings of ICCV 1998*, pages 735–743, 1998.

[13] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3):7–42, 2002.

[14] D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *CVPR 2000*, pages 345–353, 2000.

[15] C. Strecha, R. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *ICCV 2003*, pages 1194–1201, 2003.

[16] J. Sun, H,-Y Shum, and N. -N. Zheng. Stereo matching using belief propagation. In *Proceedings of ECCV,2002*, pages 510–524, 2002.

[17] G. Vogiatzis, P. H. S. Torr, S.M. Seitz, and R. Cipolla. Reconstructing relief surfaces. In *Proceedings British Machine Vision Conference*, pages 117–126, 2004.