Hierarchical Volumetric Multi-view Stereo Reconstruction of Manifold Surfaces based on Dual Graph Embedding

Alexander Hornung and Leif Kobbelt
Computer Graphics Group, RWTH Aachen University
http://www.rwth-graphics.de

Abstract

This paper presents a new volumetric stereo algorithm to reconstruct the 3D shape of an arbitrary object. Our method is based on finding the minimum cut in an octahedral graph structure embedded into the volumetric grid, which establishes a well defined relationship between the integrated photo-consistency function of a region in space and the corresponding edge weights of the embedded graph. This new graph structure allows for a highly efficient hierarchical implementation supporting high volumetric resolutions and large numbers of input images. Furthermore we will show how the resulting cut surface can be directly converted into a consistent, closed and manifold mesh. Hence this work provides a complete multi-view stereo reconstruction pipeline. We demonstrate the robustness and efficiency of our technique by a number of high quality reconstructions of real objects.

1. Introduction

The faithful and consistent reconstruction of three dimensional real world objects from images or video sequences remains a great challenge in computer vision. One particularly promising approach allowing for reconstructions of very high quality is the combination of volumetric scene representations with combinatorial optimization [16, 22].

In general such approaches are based on a discretized volume in which the object to be reconstructed is embedded. Then for each volumetric element (voxel) the likelihood of being intersected by the object surface is estimated, *e.g.*, by computing its photo-consistency based on color variances [7, 18], or normalized cross-correlation [22]. Earlier reconstruction methods [18] then extracted a voxel-based approximate surface by a thresholding approach, *i.e.*, the set of most photo-consistent voxels represents the object surface. While this is often sufficient for applications such as new view synthesis [14], the resulting scattered set of voxels does not support any global smoothness or topological constraints due to the purely local labeling of voxels. This makes it especially difficult to extract a consistent surface for further processing of the geometry.

Recently several papers, *e.g.* [2], have shown how the problem of computing proper segmentation surfaces in-between the voxels of a volumetric scene representation can be casted into a combinatorial optimization problem for graph cuts. By constructing a specifically designed graph structure, the globally optimal solution to such a problem can be efficiently computed as the minimum cut of the respective graph. In [22] it is shown how this approach can be applied to volumetric multi-view stereo reconstruction, yielding 3D reconstructions of very high quality.

However, one drawback of these combinatorial approaches is that the accuracy of the solution is bounded by the resolution of the domain discretization which cannot be too high due to restricted memory capacities and computation time. Moreover, when dealing with more complex objects and imperfect acquisition conditions, we would like to use a highly redundant set of images to obtain a robust reconstruction. This increasing amount of input data obviously increases the need for more efficient algorithms.

Furthermore there exists a subtle but important issue in using the above mentioned segmentation approaches for volumetric 3D object reconstruction. A photo-consistency measure generally integrates, for a specific region in space, the likelihood of being *intersected* by the surface [5, 7, 18]. Hence it is well defined only for a proper integration domain, *e.g.*, geometrically valid entities such as voxels. The above segmentation based approaches however generate a graph structure which associates graph nodes with voxel centers and graph edges with voxel faces. This poses the question of properly defining the edge weights of the embedded graph: If we compute photo-consistency values for the voxels, these values have to be re-mapped to the graph edges, *e.g.*, by taking the average consistency of two face connected voxels. This, however, is equivalent to applying a low-pass filter to the voxel consistency values and hence reduces the effective resolution of the reconstruction.

In [22] the photo-consistency is computed for the graph edges instead of voxels. But while the projection of a voxel and hence the respective integration domain for the photo-consistency estimation is well defined, there is no such definition for projecting graph edges into images and computing their photo-consistency. A simple integration of point samples over 1D edges or 2D faces would introduce a directional bias in the 6-connected grid, and the geometrical interpretation becomes even more unclear if one wants to extend this approach to larger voxel neighborhoods to allow for better surface integral approximations [2, 22]. So while these approaches are generally well suited for problems where voxels have to be segmented into different classes by a contour or surface at voxel boundaries, they do not support the extraction of a proper surface *intersecting* the interior of voxels, nor is it clear how the photo-consistency of the corresponding graph edges should be defined.

In this paper we present a new multi-view stereo approach for efficient, high quality volumetric 3D reconstruction of solid objects, which resolves the above mentioned problems. Our technique is based on a new octahedral graph structure which is embedded into the volumetric grid, such that voxel photo-consistency values are properly associated with graph edges. The minimum cut of this graph yields a manifold surface, which is globally optimal in terms of an energy functional of the photo-consistency and the area of the reconstructed surface, supporting smooth and consistent object reconstructions of high quality. Moreover, our specific graph design supports a hierarchical surface extraction, which allows us to efficiently process even high volumetric resolutions and a large number of input images. Finally, to provide a complete reconstruction pipeline, we show how to extract a manifold triangle mesh directly from the set of intersected surface voxels yielding high quality meshes which are directly available for further geometry processing.

2. Related Work

Our method is inspired by a number of important results from the fields of volumetric 3D reconstruction as well as efficient energy minimization techniques using graph cuts.

Based on the principles of voxel coloring presented by Seitz and Dyer [18], Kutulakos and Seitz [12] formulated the N-view reconstruction problem of an arbitrary Lambertian shape in their theory of space carving. This approach has shown to allow for highly efficient implementations using, *e.g.*, hardware accelerated solutions [14], which makes it especially suitable in the context of new-view synthesis and image-based rendering techniques. However, one problem in common with most techniques related to voxel coloring is the fact, that they cannot guarantee geometric constraints such as spatial coherence of the reconstructed object surface, since all classifications are made purely local at voxel level.

Different methods based on deformable polygonal surfaces have been proposed to directly reconstruct a proper manifold. [6] present a level-set method which enables automatic changes of the surface topology due to its representation as an implicit function. [8] use a stochastic refinement of the visual hull, such that it approximately satisfies the photo-consistency constraint. [13] improve the convergence by integrating 3D and 2D data into the optimization process. Reconstructions of very high quality from high resolution input images based on deformable models have been presented by [5]. Although these gradient based methods are able to yield high quality reconstructions of consistent object surfaces, they often have a large computational complexity on the one hand, and they are in general sensitive to local minima in the target functional . Hence, especially in the presence of noise, they cannot guarantee convergence to a globally optimal surface, which best satisfies all constraints such as photo-consistency, smoothness, or a minimal surface area.

Recently, efficient methods for energy minimization based on graph cuts were proposed for several vision related problems, some of which can guarantee finding the global optimum of the involved energy functional. A solution to the voxel occupancy problem from visual hulls based on a segmentation using binary labeling was proposed by [20]. [3] showed how to extend the optimization of binary labels to an arbitrary number of discrete labels by iterated graph cuts with applications in disparity based stereo. Generalized multi-camera stereo reconstruction using graph cuts was presented by [11]. The types of energy functionals which can be minimized using iterated graph cuts was analyzed in [10]. [2] show how to extend the graph construction to compute globally minimal geodesic contours or surfaces for arbitrary Riemannian metrics.

Applications of graph cut based energy minimization to volumetric stereo reconstruction were presented in [16, 19, 22]. [16] present a volumetric reconstruction technique for computing globally optimal and smooth surfaces in specific stereo setups. They show results for several stereo setups with remarkably detailed 3D reconstructions. However, the method is restricted to one rectangular surface patch and cannot be generalized to other surface topologies. For example, for the reconstruction of a closed surface, they have to join separately computed patches by which they lose the optimality near the patch boundaries.

[19] present a graph cut based reconstruction approach which combines strict enforcement of silhouette constraints with photo-consistency constraints. However this approach currently does not support objects of arbitrary genus, and furthermore does not respect concavities that are not present in the silhouette, but which coincide with rim curves.

The approach presented in [22] is closest related to our work: It extends the Riemannian minimal surface idea [2]

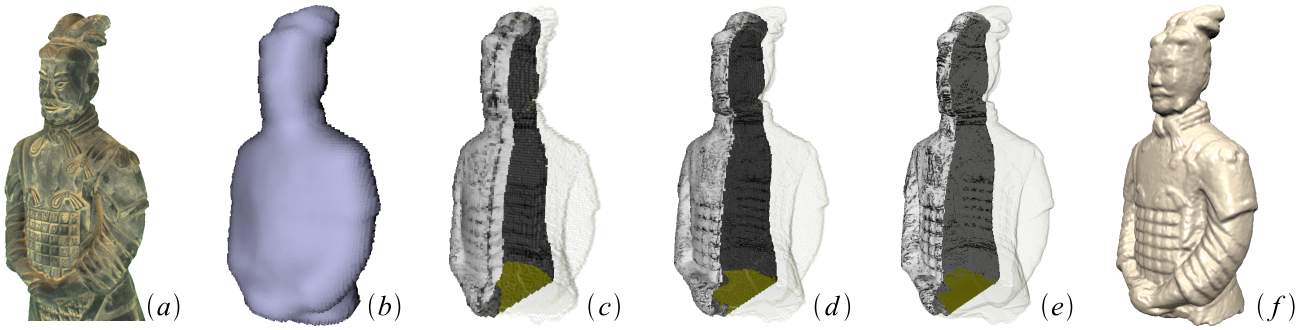

Figure 1. Given a set of input images (a) our hierarchical 3D reconstruction algorithm works on an adaptively refined grid. It first computes an approximation to the visual hull of the object (b) at a user specified resolution level (128^3 in this example). Then on successive refinement levels, photo-consistency values are computed for each voxel within a crust around the estimated surface position (c-e) (refinement levels 128^3, 256^3, and 512^3 respectively). A new graph based algorithm is used to extract the globally optimal surface on each level, which is then used as a surface proxy on the next refinement level. At the desired target resolution our method supports the direct extraction of a closed and manifold triangle mesh (f).

to general, multi-view volumetric stereo. Starting with an approximate surface, *e.g.*, the visual hull, they compute the visibility and photo-consistency of scene points in the proximity of this surface. Using graph-based surface extraction they are able to reconstruct a smooth, detailed surface of very high quality by segmenting the volume into inside and outside voxels. However, due to their specific graph structure, this method poses a number of open questions concerning the computation of edge weights within the embedded graph (*c.f.* Sect. 1).

Finally, most of the above mentioned techniques do not exploit hierarchical algorithms for increased efficiency, but perform all computations directly at the volumetric target resolution, such that it could be time consuming in practice to apply them at high volumetric resolutions with many input cameras. A first description and analysis of hierarchical graph cuts based on 6-connected grids for fast segmentation has recently been presented in [15].

3. Hierarchical Volumetric Reconstruction

Our work targets at reconstructing objects from arbitrary video sequences, for example created by using a turn-table setup or simply a hand-held camera. Similar to [22] the input to our algorithm is a number of calibrated and foreground segmented input images $I_0 \dots I_{n-1}$, *e.g.*, obtained by using structure from motion [1] and interactive segmentation techniques [17]. Instead of extracting the surface directly at the volumetric target resolution we present a hierarchical method using a sequence of adaptively refined voxel grids V^l (Fig. 1). The basic idea is to compute a voxelized, intermediate surface proxy $S^l_{opt} \subseteq V^l$ on each refinement level l, which is then refined to a new proxy S^{l+1} on the next level to constrain the volumetric region for the surface computation in V^{l+1}. This leads to a significantly reduced space and time complexity in comparison to previous work.

We start with an initial bounding cube V^0 containing the object to be reconstructed. As mentioned in [5, 7, 22] it is important to compute an initial surface proxy, *e.g.*, to estimate the visibility of each voxel in every image. Hence our algorithm starts by computing a voxelization of the object's visual hull: Beginning with $l = 0$ we iteratively refine the volume V^l and test for each voxel $v \in V^l$ if it projects to the background in one of the input images. This is done up to a level l_0, where all relevant object features such as thin structures or holes in the target model are reasonably approximated (Fig. 1, 5, 6). We typically set $l_0 = 7$ corresponding to a resolution of 128^3 voxels. However, if necessary as for example for the Dragon model (Fig. 7), the effect of choosing a different level l_0 can be easily tested by the user since the computation of the visual hull generally is a matter of a few seconds ([7], Table 1). We then define our initial voxelized surface proxy S^{l_0} as the outer boundary voxels of the voxelized visual hull, corresponding to the so called *base surface* described in [22].

Our hierarchical reconstruction algorithm then proceeds as follows: At a given level l we define a crust V^l_{crust} of voxels around the current surface proxy $S^l \subseteq V^l_{crust} \subseteq V^l$, which is supposed to contain the actual object surface, and which splits the volume V^l into three components: the crust itself, an exterior, and an interior component (Fig. 2 a). This crust V^l_{crust} can easily be computed by applying a number d of morphological dilation steps to S^l. In our experiments we generally set $d = 2$ for all levels l. This is justified by assuming an approximation error of the surface proxy S^{l-1} on the previous level up to the size of one voxel. The only exception is the initial level l_0, where $V^{l_0}_{crust}$ contains all non-background voxels of V^{l_0} since the visual hull of an object can be quite far away from the true object surface, *e.g.*, in concave regions [5, 22]. Based on the current sur-

face proxy S^l we then compute the visibility for each voxel $v \in V^l_{crust}$ in each image I_j using a hardware-accelerated technique described in [7]. Alternatively this can be done as described in [22], or by using simple ray-casting techniques. Although our hierarchical method allows for iterative improvements of the visibility information, our experiments showed, that an initial visibility estimation at level l_0 is sufficient as long as all relevant features such as holes in the object are present.

The next central step is to compute the photo-consistency $\phi(v)$ for all voxels $v \in V^l_{crust}$. For maximal flexibility and efficiency we apply the method described in [7] which estimates the photo-consistency per voxel based on an illumination invariant voxel-supersampling approach for Lambertian surfaces and geometrically unconstrained camera setups. Basically this approach creates a number m of object space samples p^0 to p^{m-1} and a camera-weight w_j for each voxel v and image I_j. For each of these samples p^i a (normalized) color value $\hat{\mathbf{c}}^i_j$ in each image I_j is computed. The final photo-consistency is then computed as the sum of weighted, normalized color variances per sample:

$$\phi(v) = \frac{1}{m}\sum_i \phi(p^i), \;\; \phi(p^i) = \sum_j w_j(\hat{\mathbf{c}}^i_j)^2 - \left(\sum_j w_j \hat{\mathbf{c}}^i_j\right)^2 . \tag{1}$$

Since $\phi(v)$ measures the photo-consistency based on color variances, it has values close to zero for photo-consistent voxels, while it increases for inconsistent voxels. Although the approach presented in [7] leads to the best results in our hierarchical reconstruction setting, techniques based on normalized cross-correlation of image patches as described in [5, 22] or even simple color variances as originally proposed in [18] also lead to acceptable results, depending on the properties of the input model and the images.

Having consistency values for all voxels within crust V^l_{crust} we apply our new graph-cut based surface extraction algorithm to refine the current surface proxy S^l to an optimized approximation S^l_{opt}, which better fulfills photo-consistency and smoothness constraints on the current level l. The details of this step are described in Sect. 4. If l is our desired target resolution, we are now able to extract a manifold triangle mesh from S^l_{opt} as described in Sect. 5. Otherwise we refine the adaptive grid for all voxels $v \in S^l_{opt}$ to a new surface proxy S^{l+1}, and start a new iteration by building a new crust V^{l+1}_{crust}.

This iterative, hierarchical reconstruction setting allows us to efficiently process large input data in terms of volumetric resolution and number of input images, since the photo-consistency as well as the extraction of S^l have to be done only within a thin crust of voxels instead of the whole volume, while the globally optimal properties of the final reconstruction are preserved.

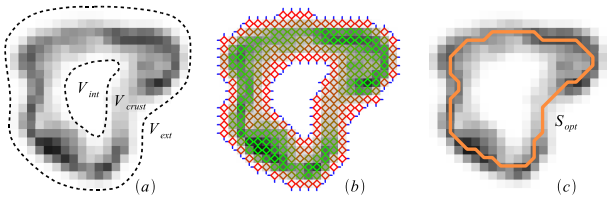

Figure 2. This image illustrates the graph based contour computation in 2D. (a) shows a voxel grid of consistency values where the darker colors indicate higher photo-consistency values. Typically these consistency values have many local minima and maxima due to color ambiguities in homogeneously textured object regions, image noise, erroneous visibility estimation, or imperfect camera calibration. In (b) the dual 2D graph is embedded into the voxel grid. Green colors correspond to low edge weights while red edges have higher weights. The source and sink node are connected to the inner and outer boundaries respectively. Despite the local inconsistencies the globally optimal graph cut correctly reconstructs surface S_{opt} as shown in (c).

4. Surface Reconstruction by Graph Cutting

In this section we will present a new graph structure for our hierarchical surface reconstruction approach, which establishes a well defined relationship between voxel photo-consistency values, the embedded graph, and the reconstructed surface. We show that this relation is achieved by representing each voxel by its dual octahedral subgraph. We will drop level indices l for a simplified notation, since all voxels $v \in V^l_{crust}$ are at the same refinement level.

Our goal is to find a surface S_{opt} within the current crust V_{crust} which approximates the true but unknown object surface. We computed the likelihood for each voxel to be intersected by this surface by assigning photo-consistency values $\phi(v)$ to each voxel v. Hence the reconstructed surface S_{opt} has to be defined by a set of surface-intersected voxels $S_{opt} \subseteq V_{crust}$. As already discussed there is an important difference of this surface definition to graph based *segmentation* approaches such as [2, 22], since these methods generate graph edges (and hence the segmentation boundary) in-between voxels (or pixels in the 2D case). This specific graph structure does not allow for a well defined assignment of voxel photo-consistency values to the graph edges.

We will introduce a new embedded graph structure with the properties that 1) the resulting surface S_{opt} is optimal with respect to an energy functional (Eq. 2) of the integrated photo-consistency and surface area and hence closely approximates the actual object surface, 2) edge weights are well defined by the photo-consistency of a corresponding voxel, and 3) a minimum cut through this graph directly yields a closed, 2-manifold surface represented by a set of surface-intersected voxels S_{opt}.

Similar to previous work the energy functional we want

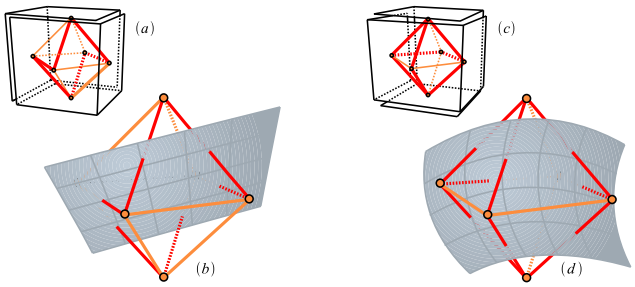

Figure 3. Splitting the faces of a cube into an exterior and an interior part is equivalent to a graph cut in the dual octahedron (a),(c). Simple configurations with 4 or 6 cut-edges correspond to locally planar cut surfaces in the geometric embedding (b) while complex configurations with 8 or more cuts correspond to locally curved cut surfaces (d).

to minimize for a surface S is a weighted sum of the integrated photo-consistency and the area of the surface:

$$E(S) = \int_S \phi(x)\, dx + \int_S a\ dS \tag{2}$$

The desired optimal surface S_{opt} should minimize this functional for a good approximation of the true object surface.

Within our voxel grid V consisting of cubical voxels with square faces, the crust V_{crust} properly separates the interior component V_{int} of V from the exterior component V_{ext} (Fig. 2), since V_{crust} is a face-connected (6-neighborhood) set of voxels. Suppose now we have an arbitrary closed surface S_{opt}. For each voxel in V_{crust} we can label its faces as interior or exterior depending on which side of the surface they lie. Faces that are intersected by the surface are labeled as interior by default. The important observation now is that if we want to separate the interior faces from the exterior faces for a single voxel, we have to cut along a sequence of edges of the voxel (Fig 3 a,c).

Based on this observation we build the following graph structure. For each voxel face in V_{crust} we define a node in the graph. Within each voxel v we connect the six nodes corresponding to the six faces in an octahedron-fashion and assign the consistency value of the voxel plus a surface area constant $\phi(v) + a$ to all twelve edges. Due to the duality of the cube (voxel) and the octahedron, we have a one-to-one correspondence between the edges of both. Hence the above voxel cut (along edges) which separates interior from exterior faces is equivalent to a graph cut in the octahedron which separates the corresponding nodes (Fig 3 b,d).

The global graph embedded in V_{crust} consists of the subgraphs of all voxels. The graph source is connected to all nodes, whose associated faces lie at the interface to the exterior component V_{ext}, while all nodes at the interior component V_{int} are connected to the sink (Fig 2 b).

Computing the minimum cut of this graph yields a set of cut edges C which minimizes the sum of edge weights, and which defines a closed surface that separates the graph into two components by splitting those voxels $S_{opt} \subseteq V_{crust}$ which are most likely intersected by the true object surface.

Besides pure consistency maximization the geometric smoothness of the resulting cut surface is enforced by two aspects of the graph embedding: First, within each voxel, a configuration which cuts as few octahedron edges as possible is preferred (since they all have the same weight). These preferred configurations cut either four or six edges and correspond to planar cuts while more complex cuts with eight or more edges correspond to curved configurations (Fig 3).

Second, since the cost of a separating cut is the sum of the weights of all edges that are split, the optimal solution might rather contain inconsistent voxels if this leads to a better global solution than summing over a larger number of voxels with better consistency.

Since the edges are embedded fairly uniform in the voxel space, the sum over all edge weights $\phi(v)+a$ can be considered as a decent, discretely sampled surface integral approximation of the consistency function and the surface area. Therefore a cut through this graph minimizes the discretization of the energy functional (2):

$$E(C) = \sum_{e \in C} \phi(e) + \sum_{e \in C} a \tag{3}$$

$E(C)$ discretizes $E(S)$ up to a constant scaling factor which depends on the voxel size. The trade-off between local consistency and global surface area minimization can be controlled to a certain extend by applying a suitable transfer function $\phi'(x) = \phi(x)^s$ with a smoothness factor $s > 0$, and by changing the surface area constant a. Since the graph complexity depends only quadratically on the current volumetric resolution due to our hierarchical approach, different choices of these parameters can be evaluated by the user with only short delays. However our experiments revealed that when using consistency measures such as [5, 7, 22], which generate a relatively distinct consistency maximum (Fig. 1 c-e), the energy functional is dominated by the photo-consistency values. This allows us to keep these two parameters constantly set to $s = 4$ and $a = 10^{-5}$ for all resolution levels and experiments. This graph-based surface reconstruction supports our initially formulated requirements concerning the minimized energy functional, the definition of edge weights, and the extraction of surface-intersected voxels S_{opt}.

5. Manifold Extraction

After the graph-cut algorithm has found the set of surface-intersected voxels S_{opt}, we have to extract this manifold surface and convert it into a triangle mesh to allow for further processing of the reconstructed 3D model.

One solution would be to apply existing iso-surface reconstruction techniques such as [9]. This however would

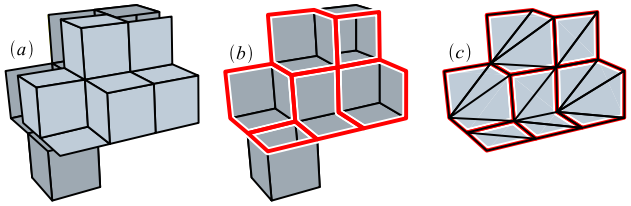

Figure 4. (a) shows a set of surface voxels from S_{opt}, which are split by the graph cut into inside (dark) and outside (bright) faces. The cut-edges within each voxel define a loop of split-edges (b). Each loop defines a polygonal face of the output surface. By triangulating these polygons we can extract the final output mesh (c).

imply a conversion of our voxel based surface into a 3D scalar field, and the resulting surface would not necessarily respect the surface topology defined by the cut edges C anymore. Instead we derive a simple and efficient algorithm for the manifold extraction directly from the geometric interpretation of our graph embedding and the computed cut.

Consider a voxel $v \in S_{opt}$ which contains some cut-edges after the cut surface has been computed for the octahedral photo-consistency graph (Fig. 3). Because of the one-to-one correspondence between the octahedron edges and the voxel (cube) edges, the set of *cut-edges* in the octahedron defines a loop of *split-edges* in the cube, splitting it into inside and outside faces (Fig. 4 a,b). Although one could create a proper mesh directly by triangulating either the set of inside or outside faces, this approach would not generate a minimal number of vertices and triangles.

Instead we can consider each of these loops of split-edges as one (non-planar) polygonal face of the output manifold. Collecting all these loops for each intersected voxel yields a polygonal mesh representation $\mathcal{M}_p$ of the final surface whose vertices lie at the voxel corners. By triangulating each of these polygons, *e.g.*, using simple triangle fans, we obtain the desired triangle mesh $\mathcal{M}_t$ (Fig. 4 c). Since the vertices $v \in \mathcal{M}_t$ are initially placed at the voxel corners the resulting reconstruction shows discretization artifacts (Fig. 5 b). We eliminate these artifacts in a post-processing step by applying a mesh smoothing filter [21] $v \leftarrow v + \triangle v$ to the mesh vertices, with the operator $\triangle$ being the discrete Laplacian defined on the mesh (Fig. 5 c). By restricting this smoothing to a maximum distance of one voxel the original reconstruction error is preserved. This method is guaranteed to efficiently extract a smooth, closed and manifold mesh.

As a side note we would like to mention that one can alternatively generate a polygonal mesh dual to $\mathcal{M}_p$, which places a vertex at each voxel center instead at voxel corners. In this case we have to generate a polygon for each 2^3 block of voxels, that contains at least 3 cut voxels. While this approach creates a reduced number of vertices, it is significantly more involved to implement for hierarchical grids.

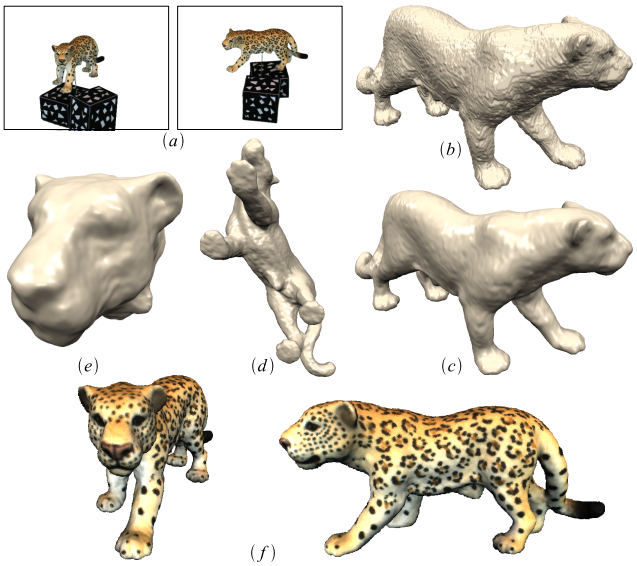

Figure 5. The Leo model was reconstructed using 86 images captured with an uncalibrated turn-table setup (a). (b) shows the mesh extracted from the graph cut surface. Discretization artifacts due to the underlying voxel grid can be effectively eliminated by applying a Laplacian smoothing filter (c). Invisible parts of a model like the belly or the feet of the Leo (d) are smoothly closed by the computed cut surface. A reconstruction (e) of the head from the same image set shows, that even very fine details such as the nostrils and the concavities of the ears are properly reconstructed. Textured and relighted versions of the model are shown in (f).

6. Results

In this section we will present the results of our method applied to several real-world data sets. Our reference system was a Linux-based Intel Pentium 4 CPU with 3.2 GHz and 2 GB of main memory. We captured video streams of the Warrior, the Leo, and the Dragon model with an uncalibrated turn-table setup and an image resolution of 1024×768. The Bahkauv statue was captured using a hand-held video camera with a resolution of 720×576. However the effective image size of the objects was often significantly smaller because of the necessity to include calibration objects within the images (*e.g.*, Fig. 5). We calibrated and segmented the video streams using standard structure from motion [1] and image segmentation techniques[17].

Table 1 presents the quantitative results for each of the models, including the complexity of the input data, timings, and the resulting mesh complexity. Our experiments show that the photo-consistency estimation is the dominating computational factor. This fact underlines the benefit of our hierarchical approach since it significantly reduces the number of processed voxels, keeping the computation times acceptable even for high volumetric resolutions.

The Bahkauv (Fig. 6) was reconstructed from only 27

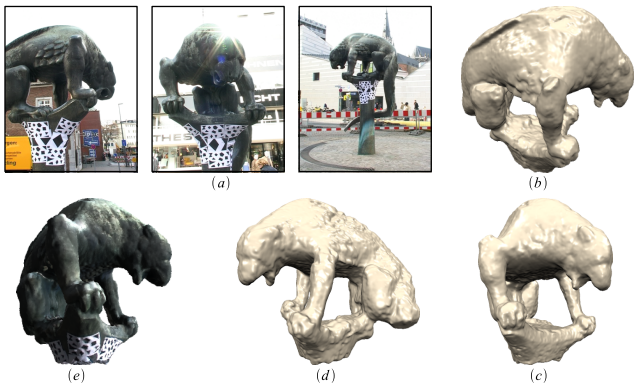

Figure 6. Reconstruction of the Bahkauv statue from only 27 images (a). The images were captured using a simple hand-held video camera on a roughly circular path. Despite the specular surface and other illumination artifacts we are able to reconstruct 3D meshes of fairly high quality (b), (c). However, some regions with a significant deviation from the Lambertian model in the input images lead to slightly noisy surface properties (d). (e) shows a textured and relighted version of the model.

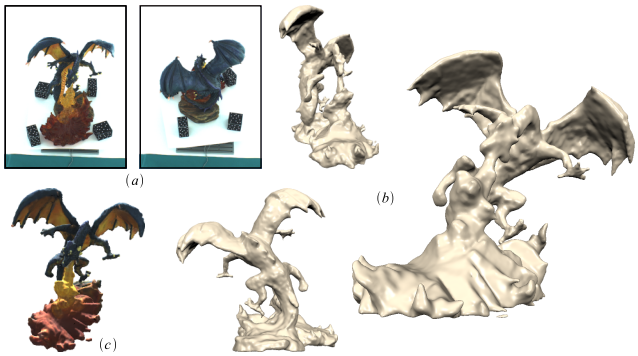

Figure 7. The dragon model is particularly difficult to reconstruct because of the fine details such as the claws as well as partially specular and transparent surface areas. We used 141 images (a) to increase the robustness of our technique and succeeded in reconstructing a fairly detailed model (b). However, our Lambertian photo-consistency measure leads to visible artifacts, *e.g.*, for the fire which is made from colored glass and hence causes complex light scattering effects. A relighted and textured version of the model is shown in (c).

images. Despite the difficult acquisition conditions with non-Lambertian, weakly textured surface properties and other illumination artifacts we achieve quite an acceptable reconstruction with less than 10 minutes overall computation time. Previous non-hierarchical approaches generally report reconstruction times in the range of one to several hours for comparable input complexity and reconstruction quality on similar reference systems. For the reconstruction of the Leo model (Fig. 5) we used 86 images. Although a quite acceptable reconstruction can already be obtained with about 30-40 images, we used a higher number of images to better reconstruct details such as the Leo's ears. With less cameras they became slightly truncated due to their relatively small image size and our imperfect camera calibration. The Dragon model represents the most complex reconstruction in terms of object structure and surface properties. We used 141 input images for increased robustness and a starting level $l_0 = 8$ corresponding to a volumetric resolution of 256^3 voxels to better carve out fine details such as the claws in the initial surface proxy. Fig. 8 shows a reconstruction of the Warrior at a high resolution of 1024^3. Despite the relatively low image resolution in comparison to the volumetric resolution our method is able to reconstruct fine surface details. For ground truth evaluation we computed the Hausdorff distance [4] of our reconstruction to a laser scan, resulting in a very low mean (max) error of 0.1% (1.9%) with respect to the bounding box diagonal. Since we explicitly generate the graph data structure for the surface computation, we used a machine with 4 GB of main memory and a reduced crust thickness for the final surface computation step. Despite this high resolution the computation times are still quite acceptable. However, one important goal of our future work will be the reduction of the memory requirements of our method. Our used input data and results are available at *http://www.rwth-graphics.de*.

Model	Bahkauv	Leo	Warrior	Warrior	Dragon
Images	27	86	71	71	141
Level l_0	$7(128^3)$	$7(128^3)$	$7(128^3)$	$7(128^3)$	$8(256^3)$
Target level	$9(512^3)$	$9(512^3)$	$9(512^3)$	$10(1024^3)$	$9(512^3)$
Visual hull	1.9 s	2.4 s	2.2 s	2.2 s	10.5 s
Visibility	2.9 s	6.5 s	5.9 s	5.9 s	35.9 s
Consistency	5.4 m	9.7 m	7.0 m	19.1 m	13.6 m
Graph cut	1.3 m	41.9 s	37.4 s	3.2 m	1.6 m
Meshing	45.0 s	17.5 s	25.5 s	2.6 m	1.2 m
Overall	9.5 m	12.8 m	10.5 m	27.8 m	20.4 m
Vertices	639 K	298 K	388 K	1570 K	575 K

Table 1. The time and space complexity of each reconstruction is mainly influenced by the number of input images, the initial level l_0, and the desired target resolution. We provide accumulated timings in seconds and minutes respectively for each of the algorithm's main phases. The overall reconstruction time additionally includes user interaction and other involved processing steps such as normal computation and the crust generation. The number of mesh vertices at the target resolution is given in the last row.

7. Conclusion

We presented a new algorithm for the reconstruction of a 3D shape from a set of video frames based on graph cut minimization. Our octahedral graph structure establishes a well defined relationship between the photo-consistency of a voxel and the edge weights of an embedded octahedral

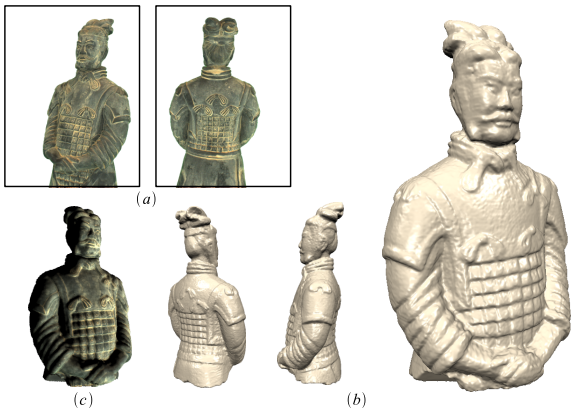

Figure 8. The warrior model was reconstructed from 71 images (a) at a high volumetric resolution of 1024^3. The resulting mesh has a complexity of 1570 K vertices (b). (c) shows the textured and relighted model.

subgraph. Our algorithm is efficient, especially if the number of input images and/or the volumetric resolution is large, because our hierarchical procedure effectively reduces the amount of voxels that have to be processed. This is achieved by using a coarse scale reconstruction to predict the region where the surface lies on the finer level. The final result of the graph cut computation can easily be converted into a closed and guaranteed manifold triangle mesh.

The current main limitation of our method to compute resolutions higher than 1024^3 is the fact that we explicitly generate the octahedral graph structure within the voxel grid, leading to a noticeable memory overhead. This overhead could be reduced by exploiting the special structure of our graph, *e.g.*, by using some indexing scheme in the voxel grid. The reconstruction quality is furthermore limited by our currently used Lambertian photo-consistency measure. Although our method allows us to effectively exploit a high number of input images to achieve robust reconstructions of difficult surfaces, we plan to investigate more sophisticated photo-consistency measures to reconstruct such surfaces more robustly. Finally we plan a more detailed evaluation of our graph construction in comparison to the current standard approaches (e.g, [2]) in terms of graph complexity, accuracy of the reconstruction, and metrication artifacts.

8. Acknowledgements

We would like to thank Martin Habbecke for his help during the generation of the presented data sets.

References

[1] P. A. Beardsley, A. Zisserman, and D. W. Murray. Sequential updating of projective and affine structure from motion. *IJCV*, 23(3):235–259, 1997.

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

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

[4] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: Measuring error on simplified surfaces. *Comput. Graph. Forum*, 17(2):167–174, 1998, http://vcg.sf.net.

[5] C. H. Esteban. *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, May 2004.

[6] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. In *Image Processing*, volume 7, pages 336–344, 1998.

[7] A. Hornung and L. Kobbelt. Robust and efficient photo-consistency estimation for volumetric 3d reconstruction. In *ECCV*, pages 179–190, 2006.

[8] J. Isidoro and S. Sclaroff. Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *ICCV*, pages 1335–1342, 2003.

[9] L. Kobbelt, M. Botsch, U. Schwanecke, and H.-P. Seidel. Feature sensitive surface extraction from volume data. In *SIGGRAPH*, pages 57–66, 2001.

[10] V. Kolmogorov and R. Zabih. What energy functions can be minimizedvia graph cuts?. *PAMI*, 26(2):147–159, 2004.

[11] V. Kolmogorov, R. Zabih, and S. J. Gortler. Generalized multi-camera scene reconstruction using graph cuts. In *EMMCVPR*, pages 501–516, 2003.

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

[13] M. Lhuillier and L. Quan. Surface reconstruction by integrating 3D and 2D data of multiple views. In *ICCV*, pages 1313–1320, 2003.

[14] M. Li, M. Magnor, and H.-P. Seidel. Hardware-accelerated rendering of photo hulls. *Computer Graphics Forum*, 23(3):635–642, September 2004.

[15] H. Lombaert, Y. Sun, L. Grady, and C. Xu. A multilevel banded graph cuts method for fast image segmentation. In *ICCV*, pages 259–265, 2005.

[16] S. Paris, F. Sillion, and L. Quan. A surface reconstruction method using global graph cut optimization. In *ACCV*, 2004.

[17] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, pages 309–314, 2004.

[18] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *CVPR*, pages 1067–1073, 1997.

[19] S. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *ICCV*, 2005.

[20] D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *CVPR*, June 2000.

[21] G. Taubin. A signal processing approach to fair surface design. In *SIGGRAPH*, pages 351–358, 1995.

[22] G. Vogiatzis, P. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *CVPR*, pages 391–398, 2005.