Continuous Global Optimization in Multiview 3D Reconstruction

Kalin Kolev, Maria Klodt, Thomas Brox and Daniel Cremers

Computer Vision Group, University of Bonn,
Römerstr. 164, 53117 Bonn, Germany
{kolev,klodt,brox,dcremers}@cs.uni-bonn.de

February 4, 2009

Abstract

In this article, we introduce a new global optimization method to the field of multiview 3D reconstruction. While global minimization has been proposed in a discrete formulation in form of the maxflow-mincut framework, we suggest the use of a continuous convex relaxation scheme. Specifically, we propose to cast the problem of 3D shape reconstruction as one of minimizing a *spatially continuous* convex functional. In qualitative and quantitative evaluation we demonstrate several advantages of the proposed continuous formulation over the discrete graph cut solution. Firstly, geometric properties such as weighted boundary length and surface area are represented in a numerically consistent manner: The continuous convex relaxation assures that the algorithm does not suffer from metrication errors in the sense that the reconstruction converges to the continuous solution as the spatial resolution is increased. Moreover, memory requirements are reduced, allowing for globally optimal reconstructions at higher resolutions.
We study three different energy models for multiview reconstruction, which are based on a common variational template unifying regional volumetric terms and on-surface photoconsistency. The three models use data measurements at increasing levels of sophistication. While the first two approaches are based on a classical silhouette-based volume subdivision, the third one relies on stereo information to define regional costs. Furthermore, this scheme is exploited to compute a precise photoconsistency measure as opposed to the classical estimation. All three models are compared on standard data sets demonstrating their advantages and shortcomings. For the third one, which gives the most accurate results, a more exhaustive qualitative and quantitative evaluation is presented.

1. Introduction

1.1. Problem Statement

We consider the classical problem of inferring a dense 3D structure reconstruction of an object from a collection of calibrated 2D views. Being one of the fundamental problems in computer vision with many applications outside its field, it has gained considerable attention and remains an active research area. There are different types of approaches according to the exploited image information. All these methods aim at

modeling the inverse process of image formation. However, in a mathematical sense, the inverse projection mapping does not exist, since all 3D points along a visual ray are projected onto the same image point. This makes the problem of multiview 3D reconstruction ill-posed. Thus, additional constraints are needed in order to allow for reasonable shape retrieval. A straightforward condition covering a large range of real scenes is the Lambertian assumption. It states that the appearance of a scene does not depend on the viewing direction, i.e., intensity is reflected uniformly in all directions. This property is the basic justification for stereo-based approaches, which search for matching image regions and infer corresponding surface locations. Multiview stereo is known to produce highly detailed reconstructions with quality close to that of laser-scanned models (Seitz et al., 2006) and exhibits one of the most commonly used image cues.

The earliest dense multiview stereo algorithms use carving techniques to obtain a volumetric representation of the scene by repeatedly eroding inconsistent voxels (Seitz and Dyer, 1997; Kutulakos and Seitz, 2000). These methods introduce a bias towards maximal photoconsistent shapes and do not enforce smoothness, which often results in rather noisy reconstructions. Later, mathematically more elegant energy minimization techniques have become more popular.

1.2. Previous Work on Energy Models for Multiview Stereo

Variational methods for multiview 3D reconstruction inherit the active contour framework proposed originally for image segmentation (Kass et al., 1988). They pose the problem as one of modeling a continuous two-dimensional surface in space by minimizing an appropriate energy functional. This methodology allows to combine a data fidelity criterion on the unknown surface with desired properties like regularity, thus achieving a considerable increase in robustness to image noise.

The first approaches are based on the geodesic active contour model (Caselles et al., 1995; Kichenassamy et al., 1995) by measuring weighted surface area, where weights reflect local photoconsistency. The corresponding flow acts as a smoothness term, while at the same time attracting the evolving shape towards photoconsistent locations. Different techniques have been applied to model the surface: level sets (Faugeras and Keriven, 1998), triangle meshes (Hernandez and Schmitt, 2004; Duan et al., 2004) and graph cuts (Vogiatzis et al., 2005). A generalization of this approach has been developed in (Pons et al., 2007), which allows to replace the classical pointwise photoconsistency estimation with a global matching score on the entire image domain. A grave drawback of the minimal surface model is that it couples data fidelity and regularization. As a result, it is difficult to adjust the regularizing behavior (Soatto et al., 2003). In particular, the global minimum of the underlying functional is always the empty set, which makes the initialization crucial.

In order to react on the shrinking behavior of the minimal surface model an additional weighted balloon term preferring shapes of larger volumes has been introduced (Vogiatzis et al., 2005; Lempitsky et al., 2006). However, although the empty set can be excluded as a solution, oversmoothing effects still persist making it difficult to reconstruct simultaneously thin protrusions and deep concavities (Hernández et al., 2007; Kolev et al., 2007a).

A straightforward alternative to the ballooning model is the incorporation of data-aware volumetric terms instead of the constant one. Different techniques to achieve this have been proposed in the literature. In (Boykov and Lempitsky, 2006) the possibility to use surface orientation to define foreground/background subdivision of the volume based on the divergence of the corresponding vector field is demonstrated. The authors propose to approximate surface orientation based on the gradient of the photoconsistency map. While this method improves on the constant ballooning model, it still does not provide state-of-the-art reconstructions. Another approach consists in merging precomputed depth maps to label voxels as interior or exterior with respect to the estimated surface (Hernández et al., 2007; Zach et al., 2007). Although this technique could produce very high-quality reconstructions, it is suboptimal in the sense that the process of 3D modeling is split into two stages. Erroneous decisions in the first stage could propagate to the final estimate, especially at locations of specular reflections or low texture. Limitations of depth-map merging approaches are discussed in section 1.4 in more detail.

1.3. Previous Work on Optimization for 3D Reconstruction

Apart from the energy model at hand, another crucial issue concerning the quality of the reconstructions is the optimization.
The first approaches rely on local minimization to optimize the underlying functionals (Faugeras and Keriven, 1998; Duan et al., 2004; Hernandez and Schmitt, 2004), which makes them sensitive to initialization and local minima. The introduction of *discrete* global minimization in the context of the maxflow-mincut framework (Greig et al., 1989; Kolmogorov and Boykov, 2004) has brought a considerable increase in robustness and removed the need for an initial guess. The potential of graph cut optimization has rapidly attracted the attention of researchers (Vogiatzis et al., 2005; Lempitsky et al., 2006; Hornung and Kobbelt, 2006) and replaced local schemes like gradient descent. However, graph cuts come along with some important shortcomings. Firstly, globality is guaranteed only in a discrete sense. In particular, the representation of geometric quantities such as boundary length or surface area is based on an L_1-metric, which is dependent on the choice of the underlying grid. As a consequence, respective reconstruction algorithms are not rotationally invariant and produce metrication errors. Although such inaccuracies can be alleviated by increasing the neighborhood structure, they cannot be removed in a discrete framework. For any choice of connectivity there exists a metrication error, which persists despite resolution refinement. (Kirsanov and Gortler, 2004) proposes a strategy to couple spatial resolution and graph connectivity and shows that this scheme converges to the continuous solution when the discretization goes to infinity. However, the practical applicability of this approach in the field of multiview reconstruction is limited, since it involves building huge graph structures and thus entails considerable memory requirements. The relatively large memory consumption of graph cuts can be decisive when computing reconstructions at a high resolution. A possible remedy is to use an adaptive multi-resolution scheme (Sinha et al., 2007) or a sparse grid (Labatut et al., 2007), but these techniques can not give any guarantee about globality of the computed solution.

1.4. Motivation and Contributions

The main contribution of the present work is the development of a novel *continuous* global optimization technique for multiview reconstruction, which allows to avoid previously mentioned limitations of shape recovery via graph cuts. A preliminary version has been presented in a conference article (Kolev et al., 2007b) and is, to our knowledge, the first method for continuous global optimization in the context of multiview 3D reconstruction. It shares similarities to continuous maxflow formulations presented for the task of 2D image segmentation (Chan et al., 2006; Appleton and Talbot, 2005) and volumetric 3D segmentation (Appleton and Talbot, 2006). However, a direct comparison to the approach of (Appleton and Talbot, 2006) is not possible, since it does not allow to incorporate regional information, which makes it inappropriate to our energy models. In the context of multiview reconstruction the use of convex minimization schemes has been independently developed in (Zach et al., 2007). While (Zach et al., 2007) use it merely to merge sets of precomputed range images, in this work we show how classical non-convex formulations for multiview 3D reconstruction can be globally optimized in a continuous manner. All these techniques have been inspired by the pioneering works of (Hu, 1969), (Strang, 1983).
Additionally, we propose and compare different energy models amenable to the discussed optimization method. They all consist of a combination of a photoconsistency-based discontinuity term and regional labeling terms. The difference is in the way the photoconsistency function is computed as well as in the definition of regional costs. The first model has been introduced in (Kolev et al., 2007b). It is based on color information from the input images to construct foreground/background subdivision of the 3D volume. More precisely, it is built upon the probabilistic formulation of (Kolev et al., 2006). Since for computing photoconsistency one needs the visibility of surface points, the photoconsistency term is collapsed at the beginning. With the resulting approximate visibility information determined by the gradient of the corresponding signed distance function, we can globally optimize the energy that includes both constraints. The approach is related to the one introduced in (Vogiatzis et al., 2005). Beyond the fact that the optimization is performed in a continuous setting, the main difference is that the sought surface is not restricted to lie within some predefined band, which imposes different weighting of silhouette and stereo costs. In the second energy model the classical photoconsistency estimation is replaced by the voting scheme of (Hernandez and Schmitt, 2004), which results in more precise photoconsistency maps. The regional terms are constructed in the same way as in the first model. In contrast, the third model replaces the silhouette-based foreground/background subdivision of the volume by a more sophisticated one using stereo information, which allows to capture also surface indentations not "visible" by the image silhouettes. This approach has been originally introduced in (Kolev et al., 2007a). It is related to depth map fusion methods (Curless and Levoy, 1996; Hernández et al., 2007; Zach et al., 2007). However, our formulation is entirely volumetric and does not involve any preprocessing on the image domain. This entails a series of advantages:

– It avoids discretization problems that could arise in a per-pixel visual ray determination, since a ray through a pixel will generally not capture the volume subdivision.

– A crucial issue when measuring photoconsistency along viewing rays is the sampling rate of the discretization. A too dense sampling leads to high computational costs, whereas a too sparse sampling could result in a miss of the maximizing location. In the volumetric framework, the sampling is naturally given by the volume resolution.

– The computational time of the proposed mechanism does not depend on the resolution of the input images but only on the volume resolution.

The paper is laid out as follows. The next section contains a brief review of related continuous global optimization techniques in the context of image segmentation. In Section 3 we present and discuss the underlying energy models. Section 4 is devoted to the optimization technique including implementation details. We show experimental results and quantitative evaluation in Section 5 and conclude the paper with a brief summary in Section 6.

2. Convex Formulations of Image Segmentation

In a series of works (Chan et al., 2006; Chambolle, 2005; Bresson et al., 2005) image segmentation functionals, namely the two-phase piecewise constant Mumford-Shah model (Mumford and Shah, 1989) and the snakes (Kass et al., 1988) were addressed by means of convex formulations. The key idea is to represent region-integrals by means of a binary variable $u : \Omega \subset \mathbb{R}^2 \rightarrow \{0, 1\}$ indicating foreground/ background and subsequently to relax this constraint to a convex one. The weighted length term proposed in the snakes and the geodesic active contours (Caselles et al., 1995; Kichenassamy et al., 1995) can then be expressed by means of a weighted total variation (TV) norm (Rudin et al., 1992):

$$TV_g(u) = \int_\Omega g(|\nabla I|)\, |\nabla u|\, dx, \tag{1}$$

with an edge indicator function $g(|\nabla I|)$ that provides the local metric.
Since the space of binary functions is a non-convex space, also the respective optimization problems are non-convex. However, in (Chan et al., 2006) it was found that when minimizing the total variation norm over all real-valued functions $u : \Omega \rightarrow \mathbb{R}$, the values of $u(x)$ converge to $\pm\infty$ almost everywhere. Therefore the segmentation can be cast as a convex problem on the convex space of functions $u : \Omega \rightarrow [0, 1]$ by enforcing $0 \leq u(x) \leq 1$ via a convex penalizer (Chan et al., 2006)

$$\theta(u) := \max\left\{0, 2\left|u - \frac{1}{2}\right| - 1\right\}. \tag{2}$$

Minimization over the space of real-valued functions and subsequent thresholding will then lead to a global minimizer of the respective segmentation problem.
In this work, particularly in Section 4, we will revisit these ideas and show that under appropriate assumptions the multiview reconstruction problem can be cast as a spatially continuous convex optimization problem. Moreover we will introduce an efficient numerical solution by means of Successive Overrelaxation (SOR).

3. Continuous Energy Models for Multiview Reconstruction

In this section we present and discuss three different energy models for multiview reconstruction. All three functionals have the same structure combining on-surface photoconsistency and regional costs.
Let $V \subset \mathbb{R}^3$ be a volume, that contains the scene of interest, and $I_1, \ldots, I_n : \Omega \to \mathbb{R}^3$ a collection of calibrated color images with perspective projections $\pi_1, \ldots, \pi_n$. We are looking for some surface $\hat{S} \subset V$ that gives rise to these images. According to a certain surface estimate S, all points in V can be divided into two classes: lying inside S or belonging to the background, i.e. $V = R^S_{obj} \cup R^S_{bck}$, where R^S_{obj} denotes the interior and R^S_{bck} the exterior. Considering the given image content we can assign each point $x \in V$ photoconsistency costs $\rho(x) \in [0, 1]$ describing the probability of a voxel for lying on the surface, based on its projections onto the images, where it is visible. The basic idea is that under the Lambertian assumption points on the surface are expected to have consistent appearance on the images, whereas distant points will generally give inconsistent projections. In a similar manner, we can compute costs $\rho_{obj}(x), \rho_{bck}(x) \in [0, 1]$ describing probabilities of a point x to belong to R^S_{obj} and R^S_{bck}, respectively. Hence, we can formulate the following energy minimization problem:

$$\begin{aligned} E(S) &= \int_{R^S_{obj}} \rho_{obj}(x)\, dx + \int_{R^S_{bck}} \rho_{bck}(x)\, dx + \nu \int_S \rho(x)\, dS \\ \hat{S} &= \arg\min_{S \subset V} E(S). \end{aligned} \tag{3}$$

The first two terms of the functional impose correct subdivision of the volume into interior/exterior according to the respective regional costs. The last term acts as a constraint both for smoothness and photoconsistency by seeking the minimal surface with respect to a Riemannian metric. Hence, it can be considered as a weighted smoothness term. Note that the cost functions may also depend on the orientation of the surface estimate S in order to take distortion of the compared image patches into account. This dependency is suppressed here for simplicity.
In the following, the three energy models and the differences between them are discussed in more detail.

3.1. Energy Model I: Silhouette-based Regional Constraints & Classical Photoconsistency

This model relies on a classical foreground/background subdivision of the 3D space based on silhouette cues (Vogiatzis et al., 2005; Hornung and Kobbelt, 2006). However, in many cases, silhouettes are not easy to extract automatically. Thus, it is beneficial to consider a probabilistic model, which deals with uncertainty by taking all views into account. To this end, we introduced in (Kolev et al., 2006) the conditional probabilities for observing intensities $I_l(\pi_l(x))$ in images $1, ..., n$ as

$$\begin{aligned} P_{obj}(x) &:= P(\, \{I_l(\pi_l(x))\}_{l=1,\ldots,n} \mid x \in R^S_{obj}) \\ P_{bck}(x) &:= P(\, \{I_l(\pi_l(x))\}_{l=1,\ldots,n} \mid x \in R^S_{bck}). \end{aligned} \tag{4}$$

Note that in this formulation $P_{obj}(x)$ and $P_{bck}(x)$ will generally not sum to 1. Considering dependence of the image observations we can write

$$\begin{aligned} P_{obj}(x) &= \sqrt[n]{\prod_{i=1}^{n} P(I_i(\pi_i(x)) \mid x \in R^S_{obj})} \\ P_{bck}(x) &= 1 - \sqrt[n]{\prod_{i=1}^{n} [1 - P(I_i(\pi_i(x)) \mid x \in R^S_{bck})]}. \end{aligned} \tag{5}$$

The probability of a voxel being part of the foreground is equal to the probability that *all* cameras observe this voxel as foreground, whereas the probability of background membership describes the probability of *at least one* camera seeing background. The root is for normalization with respect to the number of camera views, since both products will converge to 0 for $n \to \infty$. Thus, dependency between single image observations is expressed in terms of their geometric mean. The simultaneous use of all available image information in a probabilistic manner leads to a considerable increase of robustness compared to the classical carving technique (see (Kolev et al., 2006)).
The foreground/background probabilities for the single image observations

$$\begin{aligned} P(I_i(\pi_i(x)) \mid x \in R^S_{obj}) &\sim \mathcal{N}(\mu_{obj}, \Sigma_{obj}) \\ P(I_i(\pi_i(x)) \mid x \in R^S_{bck}) &\sim \mathcal{N}(\mu_{bck}, \Sigma_{bck}). \end{aligned} \tag{6}$$

are modeled to be Gaussian distributed. The parameters of both models, i.e. mean vectors and covariance matrices, are estimated interactively by marking a small object and background region in one of the images. This is a requirement for the energy to be globally minimizable. Minimization of the first two terms in (3) results in the most probable surface with respect to the probability distributions P_{obj} and P_{bck}.
A classical methodology to derive cost terms based on probability values is to apply the negative logarithm. In our formalism this reads

$$\begin{aligned} \rho_{obj}(x) &= -\log P_{obj}(x) \\ \rho_{bck}(x) &= -\log P_{bck}(x). \end{aligned} \tag{7}$$

Both values could be additionally normalized to lie within $[0, 1]$.
Now, we will concentrate on the definition of the photoconsistency function ρ in the last term of (3). A basic requirement to compute this photoconsistency 3D map is that camera visibility is available. To this end, we minimize the energy with Euclidean regularizer $\rho(x) = 1$. From the resulting surface, one can compute a signed distance function $\phi : V \to \mathbb{R}$, which in turn allows for normal estimation $N_x = \frac{\nabla\phi}{|\nabla\phi|}$ to each voxel $x \in V$. Hence, visibility is determined by front-facing cameras according to the estimated normal direction. In particular, photoconsistency is computed in terms of the normalized cross-correlations by averaging over front-facing cameras

$$c(x) = \frac{1}{|\mathcal{N}(x)|} \sum_{(i,j) \in \mathcal{N}(x)} NCC(\pi_i(x), \pi_j(x)), \tag{8}$$

where $\mathcal{N}(x)$ denotes the set of all front-facing camera pairs according to the normal direction N_x. In particular

$$\mathcal{N}(x) = \{(i,j) \in \{1,...,n\}^2 \mid \angle(V_i, N_x) \leqq \gamma_{max}, \angle(V_j, N_x) \leqq \gamma_{max}, i \neq j\},$$

where V_k is the viewing direction of camera k. In our experiments we used $\gamma_{max} = 60°$. Another important issue associated with the computation of the matching score is estimation of patch distortion. A square patch in one of the images does not in general correspond to a square patch in the other images due to the nonlinear nature of the projection mapping. In order to take distortion into account, we locally approximate the surface by its tangent plane according to (Faugeras and Keriven, 1998). As a result, distortion can be estimated via a homography mapping. For a camera pair (i,j) it is given by

$$H_{ij} = R_{ij}^T - \frac{R_{ij}^T T_{ij} N_x^T}{N_x^T x}, \tag{9}$$

where $R_{ij} \in \mathbb{R}^{3\times3}$ and $T_{ij} \in \mathbb{R}^3$ denote the relative rotation and translation between the local coordinate systems of both cameras. All involved entities are defined in the coordinate system of the reference camera. Note that H_{ij} is only determined up to a scale factor. Now, we can compute a NCC score based on the proposed local distortion model

$$NCC(\pi_i(x), \pi_j(x)) = \frac{1}{c_1 c_2} \sum_{p\in\mathcal{P}} \langle I_i(p) - \bar{I}_i(\pi_i(x)), I_j(H_{ij}(p)) - \bar{I}_j(\pi_j(x))\rangle, \tag{10}$$

where $\mathcal{P}$ stands for a square patch around $\pi_i(x)$ in the reference image i, $\bar{I}_i$ and $\bar{I}_j$ are the corresponding mean values, and c_1, c_2 are normalization constants. Note that the size of each patch is determined according to its projection on the reference image plane rather than being set to a fixed size on the tangent plane. This avoids sampling problems on the image domain. For the experiments presented here we used 7×7 pixel windows. For each point $x \in V$ we get some measure $c(x)$ between -1 and 1, where 1 stands for perfect correlation. This value is then mapped to the unit interval $[0,1]$ using the following function proposed in (Vogiatzis et al., 2005):

$$f(s) = 1 - \exp\left(-\tan\left(\frac{\pi}{4}(s-1)\right)^2 / \sigma^2\right). \tag{11}$$

The parameter σ controls the fidelity of the surface to the data and exhibits a trade-off between smoothness and fitness to the observed measurements. We used $\sigma = 0.5$ in our experiments. Finally, we obtain $\rho(x) = f(c(x))$.

3.2. Energy Model II: Silhouette-based Regional Constraints & Denoised Photoconsistency

The classical photoconsistency estimation used by the previous model generally yields noisy measures due to homogeneity or repeatability of the texture pattern, which could result in noisy reconstructions. To this end, (Hernandez and Schmitt, 2004) suggests the use a more elaborate approach to increase the accuracy of the photoconsistency computation. The basic idea is to ask each camera to give a vote to a point in space. The

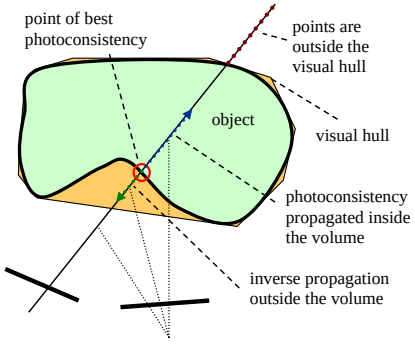

Figure 1. Propagation of photoconsistency. Illustration of the proposed approach to spread stereo information inside a volume. Spatial labeling of the volume as interior (blue arrow) or exterior (green arrow) is derived based on the location of maximal photoconsistency (red circle) along the depicted viewing ray.

vote is accepted only if the optimum is reached at the current point. This methodology leads to a considerable increase in the precision of the corresponding photoconsistency maps (see Fig. 2). This scheme is further developed and described in more detail in Section 3.3.
The current model combines the silhouette-based volume subdivision used by energy model I and the proposed denoised photoconsisteny estimation. It is similar to the approach of (Vogiatzis et al., 2007).

3.3. Energy Model III: Stereo-based Regional Constraints & Denoised Photoconsistency

A great limitation of both previously presented energy models is that they use silhouette-based regional terms, which do not capture surface indentations, since concavities do not affect the observed silhouettes. As a result, these functionals introduce a bias towards the maximal silhouette-consistent shape, i.e. the visual hull. In order to address this shortcoming, the current model replaces previous regional terms by more accurate ones (see Fig. 2). The basic idea is to propagate classical *on-surface* photoconsistency within the volume and thereby define confidence values for lying *inside* or *outside* the observed object (see Fig. 1). In the following, this approach is explained in more detail. The main difficulty in defining volume subdivision likelihoods is the fact that the state of each voxel in space (inside/outside the object) is affected by potentially distant points. We solve this problem by measuring photoconsistency along visual rays exploiting the following property of silhouette-consistent shapes:

Property: *Let S be an arbitrary surface, which is consistent with the silhouettes of a set of input images $I_1, \ldots, I_n$. Then, each visual ray passing through a point x in the interior of S intersects the real observed surface $\hat{S}$ at least once.*

If there exists a visual ray through a point x, which does not intersect the real surface $\hat{S}$, x does not project within the silhouette of the respective image. Hence, this point cannot lie in the interior of a silhouette-consistent shape. Note that the above property is fulfilled for the maximal consistent shape as well as for any subset of it. This leads to

the following idea. We can compute photoconsistency along each visual ray and take the position, where its maximum is reached, as a potential intersection with the real surface $\hat{S}$; see Figure 1. Of course, a viewing ray could intersect $\hat{S}$ more than once, but only the first intersection will be photoconsistent according to a certain set of neighboring cameras. Based on this observation, we can convert classical photoconsistency describing the likelihood for lying *on* the surface into regional terms representing an *interior/exterior* assignment.

We start with an initial silhouette-based surface approximation S_I computed as described in 3.1. Due to the above property we consider all voxels x lying in the interior $R_{obj}^{S_I}$ of S_I and corresponding visual rays passing through x. Let $r_j(x,t)$ be the visual ray of camera j, parametrized by t starting at the camera position. Let t_{cur} be the position of x along the ray. We measure photoconsistency along the ray according to another camera i:

$$C_i^j(x,t) = NCC(\pi_i(r_j(x,t)), \pi_j(r_j(x,t))). \tag{12}$$

The computation of the NCC score is given in (10). Once again, patch distortion is approximated by a local homography mapping defined by the normal direction N_x. Since it is expected that the orientation at the surface intersection point of a viewing ray corresponding to a front-facing camera is similar to that measured at x, the same normal direction N_x can be used to estimate distortion along the entire ray r_j. Note that the second term in (12) stays constant for varying t, since points on the ray r_j always project onto the same location in image I_j. This formulation can be extended to multiple cameras:

$$C^j(x,t) = \sum_{i=1}^{m} w_i^j(x) C_i^j(x,t). \tag{13}$$

We sum only over neighboring cameras according to the normalized viewing direction $V_j(x)$ of camera j. That is, camera i is excluded if $\alpha_i^j(x) := \angle(V_i(x), V_j(x)) > \alpha_{max}$ for some bounding angle α_{max}. The weights w_i^j are computed as

$$w_i^j(x) = \frac{\alpha_{max} - \alpha_i^j(x)}{\sum_{k=1}^{m} \alpha_{max} - \alpha_k^j(x)} \tag{14}$$

in order to compensate for non-linear projective warping and violations of the occlusion approximation. We set $\alpha_{max} = 45^\circ$ in all our experiments, but a more conservative smaller value could lead to a considerable reduction of computational time. As mentioned above, we determine the maximal photoconsistency along r_j together with the location, where it is reached:

$$\begin{aligned} C_{max}^j(x) &= \max_t C^j(x,t) \\ t_{max} &= \arg\max_t C^j(x,t). \end{aligned} \tag{15}$$

A natural choice for the sampling rate along the ray is the volume resolution, since it poses a constraint on the reconstructable surface details. Finally, we can define costs for interior/exterior assignment according to ray r_j as

$$\begin{aligned} \rho_{obj}^j(x) &= H(t_{max} - t_{cur}) \cdot (1 - f(C_{max}^j)) \\ &\quad + (1 - H(t_{max} - t_{cur})) \cdot f(C_{max}^j) \\ \rho_{bck}^j(x) &= H(t_{max} - t_{cur}) \cdot f(C_{max}^j) \\ &\quad + (1 - H(t_{max} - t_{cur})) \cdot (1 - f(C_{max}^j)), \end{aligned} \tag{16}$$

where H is the Heaviside function

$$H(z) = \begin{cases} 1, & if\ z \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

and f is defined in (11). The computed values depend on whether the maximal photoconsistency location t_{max} lies before or behind the current voxel t_{cur}. If for example $t_{max} < t_{cur}$, ρ^j_{obj} decreases and ρ^j_{bck} increases with the maximal measured photoconsistency C^j_{max} accounting for uncertainties because of mismatches. In effect, the Heaviside function H realizes this case differentiation. The final regional costs can be computed by simple averaging over single rays r_j, which yields

$$\begin{aligned} \rho_{obj}(x) &= \frac{1}{l} \sum_{j=1}^{l} \rho^j_{obj}(x) \\ \rho_{bck}(x) &= \frac{1}{l} \sum_{j=1}^{l} \rho^j_{bck}(x). \end{aligned} \tag{18}$$

In practice, only visual rays of front-facing cameras due to the normal N_x are considered, as described in 3.1. Note that $\rho_{obj}(x) + \rho_{bck}(x) = 1$ for all $x \in V$. In case of photometrically difficult scenes containing noise and shading effects, more sophisticated fusion strategies could be used. We experimented with a weighting procedure based on the variance of the measured photoconsistency values along viewing rays, but we could not observe any visible improvements in the reconstructions.

The process of maximization of photoconsistency along visual rays can be exploited in the computation of the on-surface costs $\rho(x)$. In this respect, the voting scheme proposed in (Hernandez and Schmitt, 2004) naturally fits in our framework. It brings about significant improvements in the localization of the on-surface cost values $\rho(x)$ compared to the classical method used in the preliminary conference version (Kolev et al., 2007a). The basic idea is that all potential causes of mismatches like occlusion, image noise, lack of texture etc. are uniformly treated as outliers in the matching process. Specifically, the photoconsistency value $\rho(x)$ for a given 3D point x is computed by asking every image j to give a vote for that location and subsequently fusing the votes to a final score

$$\rho(x) = \exp\{-\mu \sum_{j=1}^{l} \mathrm{VOTE}_j(x)\}, \tag{19}$$

where VOTE_j denotes the vote of camera j, and μ is a rate-of-decay parameter, which in our experiments was set to 0.15. The central idea is to allow a camera j to give a vote to the 3D location x only if the correlation measure along the corresponding viewing ray takes its maximum at x, i.e.

$$\mathrm{VOTE}_j(x) = \begin{cases} C^j_{max}(x) & if\ t_{max} = t_{cur} \\ 0 & otherwise \end{cases} \tag{20}$$

The presented voting scheme accounts for outliers due to occlusions, noise, or shading effects as well as matching ambiguities.

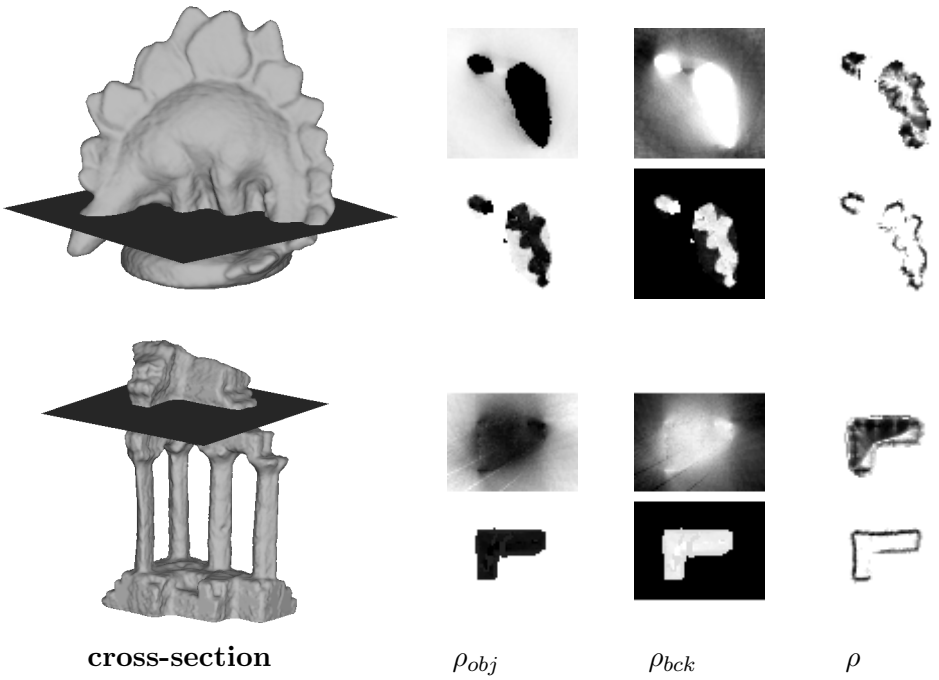

Figure 2. Comparison between the volumetric data terms used by energy models I, II and III on the "dinoRing" and "templeRing" data sets (see Fig. 3 and 4). Visualized are cross-sections through all data volumes (at resolution 64^3 for "dinoRing" and $64 \times 96 \times 48$ for "templeRing"): ρ_{obj}, ρ_{bck} and ρ, respectively. The traditional silhouette-based subdivision technique and photoconsistency estimation used by energy models I and II (upper row) are opposed to the more elaborate stereo-based approach and voting scheme used by models II and III (lower row). Intensity values correspond to estimated costs. Note that stereo information allows to capture surface indentations in contrast to silhouettes and hence produces more accurate regional terms. Note also that the voting scheme generally produces more precise photoconsistency maps but could fail in case of occlusions or ambiguous texture.

In order to accelerate the computation of the data terms of the three energy models, we used a banded multi-resolution scheme by starting with a coarse volume resolution and subsequently restricting the computations at finer levels. We carried out 1-4 iterations at each level and updated the data terms iteratively based on the orientation of the current surface estimate.

Fig. 2 shows a comparison between the data terms used by the models on the "dinoRing" and "templeRing" data sets (see Fig. 3 and 4). Visualized are cross-sections through all data volumes at the lowest resolution (64^3 for "dinoRing" and $64 \times 96 \times 48$ for "templeRing"): ρ_{obj}, ρ_{bck} and ρ, respectively. The traditional silhouette-based subdivision technique and photoconsistency estimation (upper row) are opposed to the more elaborate stereo-based approach and voting scheme (lower row). As expected, the naive silhouette-based method fails to produce accurate regional terms at concavities like the legs of the dino figurine or the back of the temple model in contrast to the stereo-based one. As a result, the corresponding subdivision and discontinuity costs compete each other in such areas, which makes their weighting a very challenging task (see Fig. 3 and 4). Moreover, the voting scheme used by the second and third energy models yields notably more precise photoconsistency maps by removing the

influence of repeated texture patterns or accidental matching. However, as a side effect, this approach could erroneously suppress photoconsistency in case of occlusions or ambiguous texture (see, for example, the vertical inside wall of the temple model), which in turn lets the regional terms play the decisive role (see Fig. 4).

4. Continuous Global Optimization

This section deals with the optimization of the energy functional proposed in (3), which exhibits the main contribution of the current article.

4.1. An Equivalent Convex Formulation

Energy (3) can be globally optimized for given data terms $\rho, \rho_{obj}, \rho_{bck}$. We build upon the optimization technique described in Section 2 by formulating (3) as a continuous convex optimization problem.
To this end, the surface S is represented implicitly by the characteristic function $u : V \to \{0, 1\}$ of R^S_{bck}, i. e. $u = \mathbf{1}_{R^S_{bck}}$ and $1-u = \mathbf{1}_{R^S_{obj}}$. Hence, changes in the topology of S are handled automatically without reparametrization. With the implicit surface representation we have the following constrained, non-convex energy minimization problem corresponding to (3):

$$\begin{gathered} E(u) = \int_V (\rho_{bck}(x) - \rho_{obj}(x))u(x)\, dx + \nu \int_V \rho(x)|\nabla u|\, dx \to \min, \\ \text{s. t. } u \in \{0, 1\}\,. \end{gathered} \tag{21}$$

The minimization problem stated in (21) is non-convex, since the optimization is carried out over a non-convex set of binary functions. However, relaxing the binary condition and extending the optimization to all functions $u : V \to \mathbb{R}$, where also intermediate values can be taken, will cause the values of $u(x)$ to converge to $\pm\infty$ almost everywhere. In order to circumvent this difficulty, one can restrict the domain by enforcing $0 \le u(x) \le 1$ via a convex penalizer $\theta(u)$:

$$E(u) = \int_V \Big(\rho_{bck}(x) - \rho_{obj}(x)\Big)u(x) + \nu\rho(x)|\nabla u| + \alpha\theta(u(x))\, dx, \tag{22}$$

where α has to be chosen sufficiently large in order to ensure that u does not leave the interval $[0, 1]$. A possible choice for θ is given in (2). This leads to a convex formulation, which allows for global optimization by using standard techniques like gradient descent. Hence, the above functional does not possess minima that are not global. However, it is not *strictly convex*, i.e., its global minimum is not unique. It turns out that the above energy functional has a very nice property that allows for global minimization of the original *non-convex* functional (21). It is stated by the following theorem, based on the work of (Chan et al., 2006):

Theorem: *If $u^* : V \to \mathbb{R}$ is any minimizer of the functional (22), then for any threshold $\mu \in (0,1)$ the binary function $\mathbf{1}_{\Sigma_\mu(u^*)}(x) : V \to \{0,1\}$ with $\Sigma_\mu(u) := \{x : u(x) > \mu\}$ is also a minimizer of (22).*

Proof: Let u^* be a global minimum of (22). The convex penalizer $\theta(u)$ (2) effects $u \in [0,1]$. We express the energy (22) in terms of the level sets of u and then minimize pointwise in μ. For these purposes we use the following layer cake representation of $u \in [0,1]$:

$$u(x) = \int_0^1 \mathbf{1}_{\{x:u(x)>\mu\}}\, d\mu$$

For the data fidelity term we obtain

$$\begin{aligned}
&\int_V \Big(\rho_{bck}(x) - \rho_{obj}(x)\Big)u(x)\,dx \\
=\ &\int_V \rho_{bck}(x)u(x)\,dx - \int_V \rho_{obj}(x)u(x)\,dx \\
=\ &\int_V \rho_{bck}(x)\int_0^1 \mathbf{1}_{[0,u(x)]}(\mu)\,d\mu\,dx - \int_V \rho_{obj}(x)\int_0^1 \mathbf{1}_{[0,u(x)]}(\mu)\,d\mu\,dx \\
=\ &\int_0^1\int_V \rho_{bck}(x)\mathbf{1}_{[0,u(x)]}(\mu)\,dx\,d\mu - \int_0^1\int_V \rho_{obj}(x)\mathbf{1}_{[0,u(x)]}(\mu)\,dx\,d\mu \\
=\ &\int_0^1\int_{V\cap\{x:u>\mu\}} \rho_{bck}(x)\,dx\,d\mu - \int_0^1\int_{V\cap\{x:u>\mu\}} \rho_{obj}(x)\,dx\,d\mu \\
=\ &\int_0^1\int_{V\cap\{x:u>\mu\}} \rho_{bck}(x)\,dx\,d\mu - \int_V \rho_{obj}(x)\,dx \\
&+\int_0^1\int_{V\cap\{x:u>\mu\}^c} \rho_{obj}(x)\,dx\,d\mu \\
=\ &\int_0^1\int_{\Sigma_\mu} \rho_{bck}(x)\,dx + \int_{V\setminus\Sigma_\mu} \rho_{obj}(x)\,dx\,d\mu - C
\end{aligned}$$

where $C := \int_V \rho_{obj}(x)\,dx$ is independent of u.
The coarea formula (Strang, 1983) for the TV_ρ-Norm (1) and the fact that $u \in [0,1]$ yield

$$\int_V \rho(x)\,|\nabla u|\ dx = \int_{-\infty}^{\infty} \mathrm{Per}_\rho(\{x : u(x) > \mu\})\,d\mu = \int_0^1 \mathrm{Per}_\rho(\Sigma_\mu)\,d\mu$$

where $\mathrm{Per}_\rho(\Sigma)$ is the perimeter of the set Σ weighted by ρ.
The layer cake formula for the penalizing function θ (2) yields for $u \in [0,1]$:

$$\int_V \theta(u(x))\,dx = \int_0^1 \mathbf{1}_{\{x:\theta(u)>\mu\}}\,d\mu = \int_0^1 \mathbf{1}_\emptyset\,d\mu = 0$$

Putting all together, we obtain the following level set representation of (22):

$$E(u) = \int_0^1 \left(\int_{\Sigma_\mu} \rho_{bck}(x)\,dx + \int_{V\setminus\Sigma_\mu} \rho_{obj}(x)\,dx + \nu\,\mathrm{Per}_\rho(\Sigma_\mu)\right) d\mu - C$$

Note that the sets Σ_μ depend on u but this dependence is suppressed here for simplicity of notation. Now, we can define

$$E_u(\Sigma_\mu) = \int_{\Sigma_\mu} \rho_{bck}(x)\, dx + \int_{V\setminus\Sigma_\mu} \rho_{obj}(x)\, dx + \nu \operatorname{Per}_\rho(\Sigma_\mu)$$

and minimize this functional pointwise in μ.
Let Σ^* be a global minimizer of E_{u^*}. This implies that

$$E_{u^*}(\Sigma_\mu) \geq E_{u^*}(\Sigma^*) \text{ for } \mu \in (0,1)\,.$$

And therefore

$$E(u^*) = \int_0^1 E_{u^*}(\Sigma_\mu) d\mu - C \geq E_{u^*}(\Sigma^*) - C = E(\mathbf{1}_{\Sigma^*}).$$

Hence, $\mathbf{1}_{\Sigma^*}$ is also a minimizer of (22).
□

Any "thresholded" (global) minimizer of (22) is binary and fulfills the constraints in (21). Trivially, it is also a global minimizer of the *non-convex* functional given in (21), since the only effective difference between both functionals is the domain of admissible functions.
Finally, we obtain the following approach for globally optimizing (21):

1. Find a minimizer u of (22).
2. Threshold the result: $R^S_{obj} = \{x \in V \mid u(x) < \mu \text{ for some } \mu \in (0,1)\}$.

In our experiments, we chose $\mu = 0.5$, but we obtained virtually the same results with $\mu \in [0.1, 0.9]$.
A necessary condition for a minimum of (22) is stated by the associated Euler-Lagrange equation

$$\begin{aligned} 0 &= (\rho_{bck} - \rho_{obj}) - \nu\rho \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) - \langle \nabla\rho, \frac{\nabla u}{|\nabla u|}\rangle + \alpha\theta'_\epsilon(u) \\ &= (\rho_{bck} - \rho_{obj}) - \nu \operatorname{div}\left(\rho\frac{\nabla u}{|\nabla u|}\right) + \alpha\theta'_\epsilon(u), \end{aligned} \tag{23}$$

where θ_ϵ is a regularized version of the derivative of θ with respect to its argument.

4.2. Fast Minimization by Successive Overrelaxation

One way to solve the nonlinear system in (23) is via gradient descent. However, gradient descent converges very slowly. Thus, we suggest to use a fixed point iteration scheme that transforms the nonlinear system into a sequence of linear systems. These can be efficiently solved with iterative solvers, such as Gauss-Seidel, successive over-relaxation (SOR), or even multi-grid methods.
First, we neglect the term $\alpha\theta'_\epsilon(u)$ and replace it by simply clipping values of u that fall out of the interval $[0,1]$. The only remaining source of nonlinearity in (23) is the diffusivity $g := \frac{\rho}{|\nabla u|}$. Starting with an initialization $u^0 = 0.5$, we can compute g

and keep it constant. For constant g, (23) is linear and discretization yields a linear system of equations, which we solve with the SOR method. This means, we iteratively compute an update of u at voxel i by

$$u_i^{l,k+1} = (1-\omega)u_i^{l,k} + \omega \frac{\nu \sum\limits_{j\in\mathcal{N}(i),j<i} g_{i\sim j}^l u_j^{l,k+1} + \nu \sum\limits_{j\in\mathcal{N}(i),j>i} g_{i\sim j}^l u_j^{l,k} - b_i}{\nu \sum\limits_{j\in\mathcal{N}(i)} g_{i\sim j}^l}. \tag{24}$$

$\mathcal{N}(i)$ denotes the 6-neighborhood of i and $b_i := \rho_{bck,i} - \rho_{obj,i}$ contains the constant part of (23) that does not depend on u, i.e., the righthand side of the linear system. Finally, $g_{i\sim j}$ denotes the diffusivity between voxel i and its neighbor j. It is defined as

$$g_{i\sim j}^l := \frac{g_i^l + g_j^l}{2}, \qquad g_i^l := \frac{\rho_i}{\sqrt{|\nabla u_i^l|^2 + \epsilon^2}}, \tag{25}$$

where $\epsilon := 0.001$ is a small constant that prevents the diffusivity to become infinite when $|\nabla u_i^l|^2 = 0$ and $|\nabla u_i^l|^2$ is approximated by standard central differences. The over-relaxation parameter ω has to be chosen in the interval $(0, 2)$ for the method to converge. The optimal value depends on the linear system to be solved. Empirically, for the system at hand, we obtained the fastest convergence rate for $\omega = 1.85$. After the linear solver yields a sufficiently good approximation (we iterated for $k = 1, ..., 10$), one can update the diffusivities and solve the next linear system. Iterations are stopped as soon as the energy decay in one iteration is in the area of number precision.

5. Experiments

5.1. Experimental Comparison of the Three Cost Functionals

First, we provide a comparison between the presented energy models I, II and III (see Section 3). They were tested on the well-known "dinoRing" and "templeRing" data sets, which are part of the Middlebury multiview stereo evaluation project (Seitz et al., 2006). The data sets contain 48/47 calibrated images of resolution 640×480 of a plaster dinosaur and a reproduction of a temple in Sicily. Both objects exhibit very different properties. While the dinosaur figurine is relatively smooth and weakly textured, the temple duplicate is well-textured but of complex geometry in terms of small-scale details and sharp corners.

Fig. 3 shows a comparison on the first data set including some of the input images and multiple views of the reconstructed surfaces at volume resolution 256^3. The first two energy models clearly fail to recover the concavities (e.g. at the legs) due to the use of silhouette-based regional terms, that act in contradiction to the stereo-based on-surface term. An increase of the weighting parameter ν will not lead to the desired effect, since this will also cut protruding parts (e.g. the spikes). ν was chosen as the largest value that retains all relevant surface details, but it is still insufficient to capture the concavities, even with the improved photoconsistency estimation. This limitation of such models has been observed by other researchers (Tran and Davis, 2006; Hernández et al., 2007) and addressed via different heuristics like search space

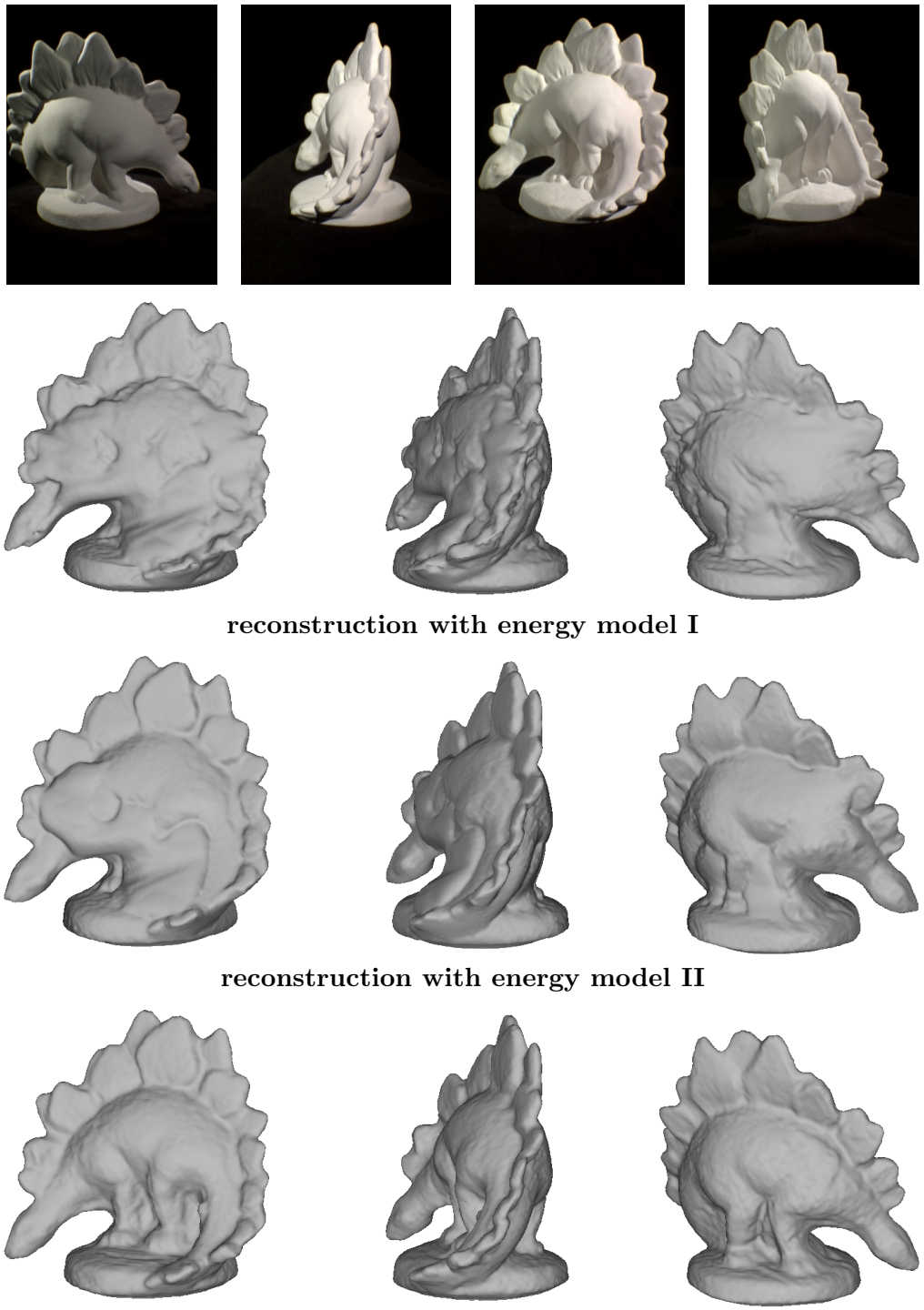

Figure 3. Comparison of energy models I, II and III on the "dinoRing" data set. 4 of 48 input images of resolution 640×480 and multiple views of the reconstructions obtained with the three energy models. Note that the first two models completely fail to recover deep concavities due to the limitations discussed previously. In contrast, energy model III is able to retrieve accurately deep indentations as well as thin protrusions.

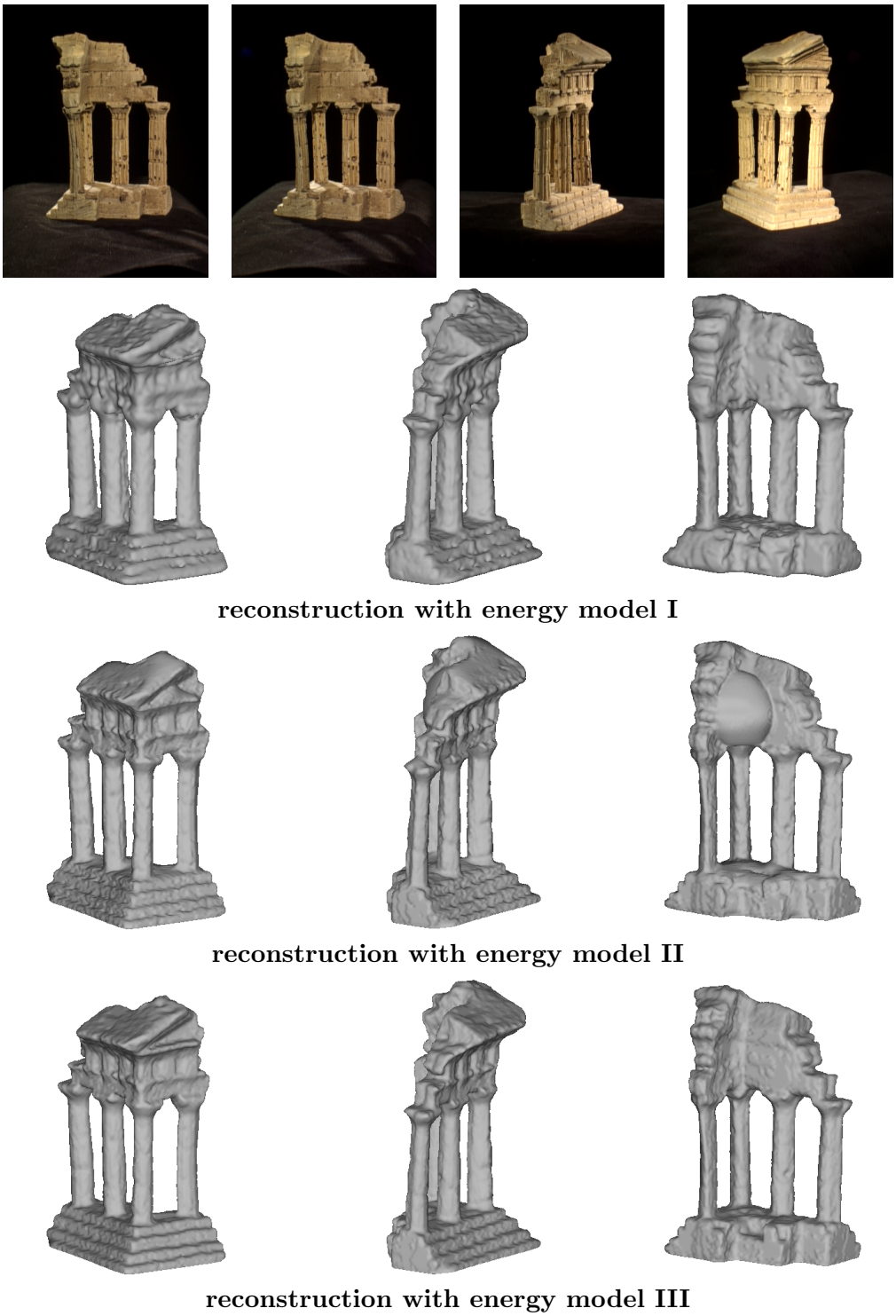

Figure 4. Comparison of energy models I, II and III on the "templeRing" data set. 4 of 47 input images of resolution 640×480 and multiple views of the reconstructions obtained with the three energy models. Although the first model captures the deep concavity at the back, it produces a very noisy reconstruction. The second model successfully suppresses noise but fails at locations of ambiguous texture. In contrast, the third model achieves the highest accuracy by retrieving all large-scale details.

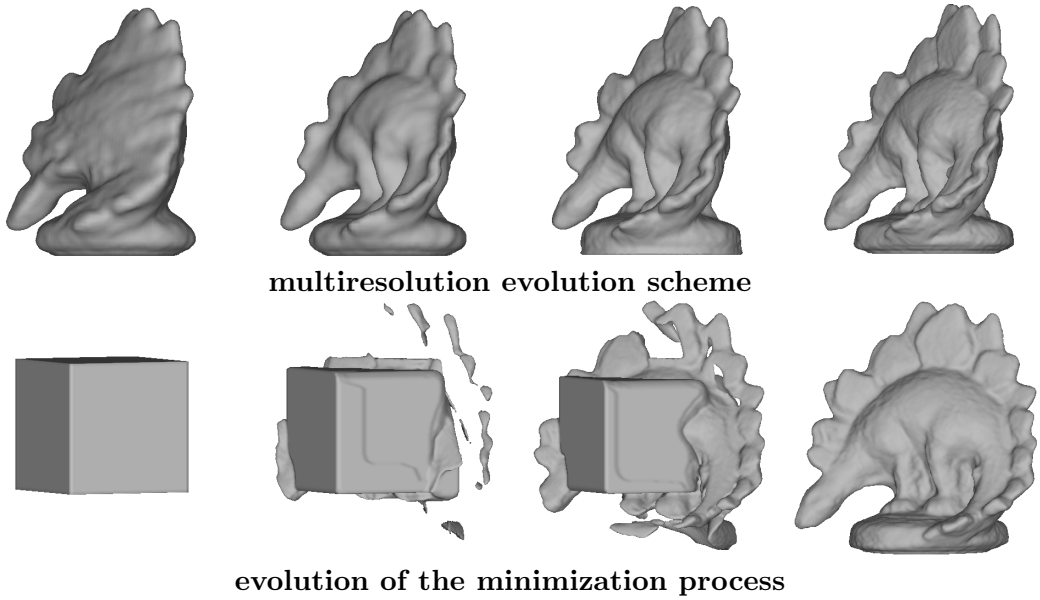

Figure 5. Surface evolution towards the final result. *First row:* Successively refined reconstructions with increasing resolutions of the volume: 64^3 (initialization and final result), 128^3 and 256^3. *Second row:* Surface evolution at the finest resolution, obtained by thresholding the evolving function u at 0.5 (see Section 4). In contrast to level set schemes the evolution process is not coherent.

restriction (Vogiatzis et al., 2005) or additional post-processing (Tran and Davis, 2006). Energy model III presents a data-driven formulation to circumvent the mentioned shortcomings and produces a visibly more accurate reconstruction.
Analogously, Fig. 4 depicts a comparison on the second data set. Here, the first energy model captures the deep indentation at the back, but the reconstruction is pretty noisy and imprecise due to the noisy photoconsistency map (see Fig. 2). Although the second model produces a generally more accurate reconstruction, it completely fails in areas of weak or ambiguous texture (for example the wall at the back; see Fig. 2). In contrast, the third model achieves the highest accuracy by generating a smooth shape preserving all large-scale details.
The computational times of the three methods, which were measured on a 2.66 GHz Intel Core2 architecture, range from 40-50 minutes for the first classical approach to more than 10 hours for the third one. Note that these runtimes can be reduced by a more conservative choice for the parameters α_{max} and γ_{max} and/or a GPU implementation, but such an analysis is beyond the scope of this article. Not surprisingly, the increased accuracy of the third model comes at the expense of increased computational efforts.

5.2. Analysis of Energy Model III

In the sequel, we give a more detailed evaluation of energy model III. As mentioned in Section 3.3 a banded multi-resolution scheme was applied in order to accelerate the computation of the data terms. Reconstructions at intermediate levels for the "dinoRing" data set are shown in Fig. 5. Moreover, the evolution of an initial surface

Table I. Quantitative evaluation on the Middlebury data sets (see Fig. 3 and 4).

data set	# images	completeness	accuracy	runtime
dinoSparseRing	16	98.3 %	0.53 mm	55min
dinoRing	48	99.4 %	0.43 mm	9h 48min
templeSparseRing	16	91.8 %	1.04 mm	1h 08min
templeRing	47	97.8 %	0.72 mm	10h 18min

towards the final result is depicted for the finest volume resolution of 256^3. Note that the final reconstruction does not depend on the initialization, since global minimization is performed. A closer look at the evolution process reveals the difference to local optimization techniques like level sets (Sethian, 1996). While the surface always evolves coherently for level set methods, there are no such constraints for the method proposed here as structures can appear and fade freely.

In Table I we give a quantitative evaluation of the proposed approach on four of the Middlebury datasets. Laser-scanned models of both objects are used as ground-truth in order to evaluate the quality of the reconstructions. The accuracy metric shown is the distance d (in millimeters) that brings 90% of the reconstructed surface within d from some point on the ground truth surface. The completeness score measures the percentage of points in the ground truth model that are within $1.25mm$ of the reconstructed model. The used volume resolution was 256^3 for "dino(Sparse)Ring" and $256 \times 384 \times 192$ for "temple(Sparse)Ring", respectively. See (Seitz et al., 2006) and the accompanying website for comparison to other methods. Note that the reconstruction of the dinosaur figurine, that exhibits a challenge to many previous approaches due to the lack of texture, demonstrate the potential of the proposed approach and ranks currently among the top-performers in both metrics. The reconstruction of the sufficiently textured temple replication, where most of the previous methods perform well, is less impressive but still satisfactory.

Finally, Fig. 6 and 7 illustrate two high-quality reconstructions on sequences with 33 images of resolution 1024×768. The volume resolution was set to $216 \times 288 \times 324$ ("bunny" sequence) and $240 \times 288 \times 360$ ("Beethoven" sequence) respectively, and the measured computational time was in the range of $2 - 4$ hours. The input images are challenging due to the presence of homogeneous texture ("bunny" sequence) or the absence of texture ("Beethoven" sequence). Despite these difficulties, which can introduce ambiguities in the matching process, the proposed approach produces accurate, highly detailed reconstructions.

5.3. Continuous vs. Discrete Global Shape Reconstruction

In this article, we introduced a spatially continuous global optimization technique for shape recovery from multiple views. Corresponding energy functionals (see (3)) describe a typical space partitioning problem and can be also optimized globally in a spatially discrete setting via graph cuts (Kolmogorov and Zabih, 2002). However, globality is guaranteed in a discrete manner, which does not preclude the presence of metrication errors. Compared to graph cuts, the proposed technique for continuous

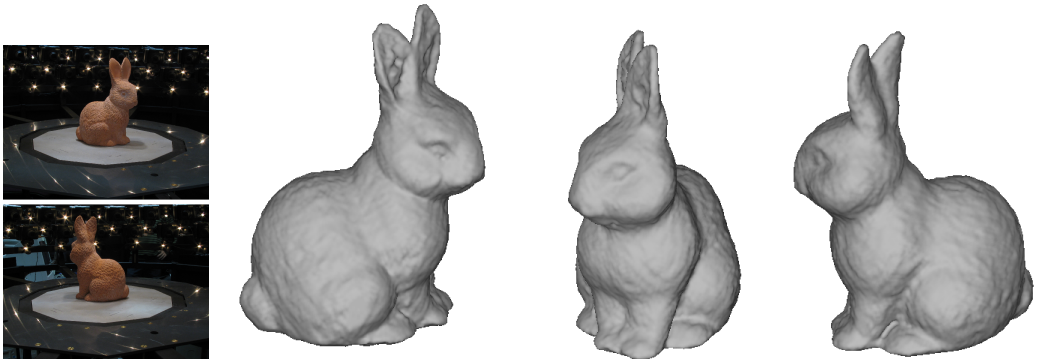

Figure 6. Bunny sequence. Two of 33 input images of resolution 1024×768 and three views of the reconstructed surface at volume resolution $216 \times 288 \times 324$.

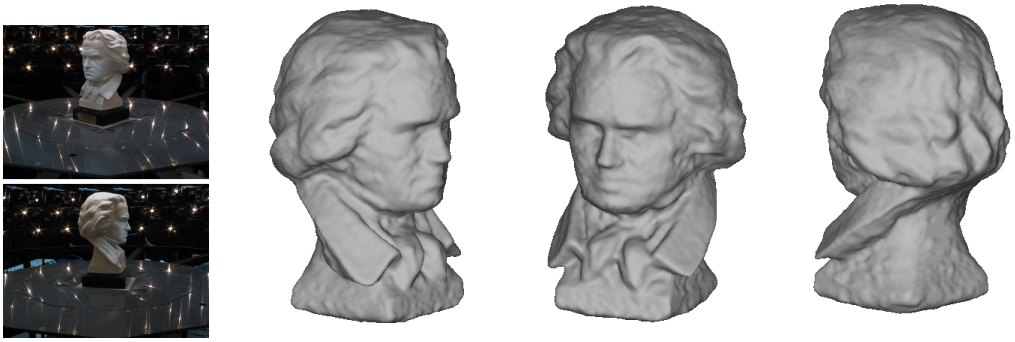

Figure 7. Beethoven sequence. Two of 33 input images of resolution 1024×768 and three views of the reconstructed surface at volume resolution $240 \times 288 \times 360$.

optimization does not suffer from similar discretization artifacts while computing a globally optimal solution.
This claim is emphasized on a synthetic experiment with missing data shown in Figure 8. We compared both minimization methods on a scenario with a known analytic solution - a bounded catenoid defined by

$$\begin{aligned} x &= 2\cosh\left(\tfrac{v}{2}\right)\cos u \\ y &= 2\cosh\left(\tfrac{v}{2}\right)\sin u \\ z &= v \end{aligned} \tag{26}$$

with $(u, v) \in [0, 2\pi] \times [-1, 1]$. To this end, the photoconsistency function ρ was set to a constant and the regional terms ρ_{obj}, ρ_{bck} used to fix the base circles ($v = \pm 1$) only. In effect, this formulation describes a minimal surface problem with given boundary constraints. The result of the proposed optimization technique at a volume resolution of $180 \times 180 \times 60$ is depicted in Fig. 8 (c) and the graph cut estimates are illustrated in Fig. 8 (a) for the 6-connectivity system and in Fig. 8 (b) for the 26-connectivity, respectively. The 6-neighborhood system completely fails to reconstruct the correct surface topology in contrast to the full 26-neighborhood, since the Euclidean metric is

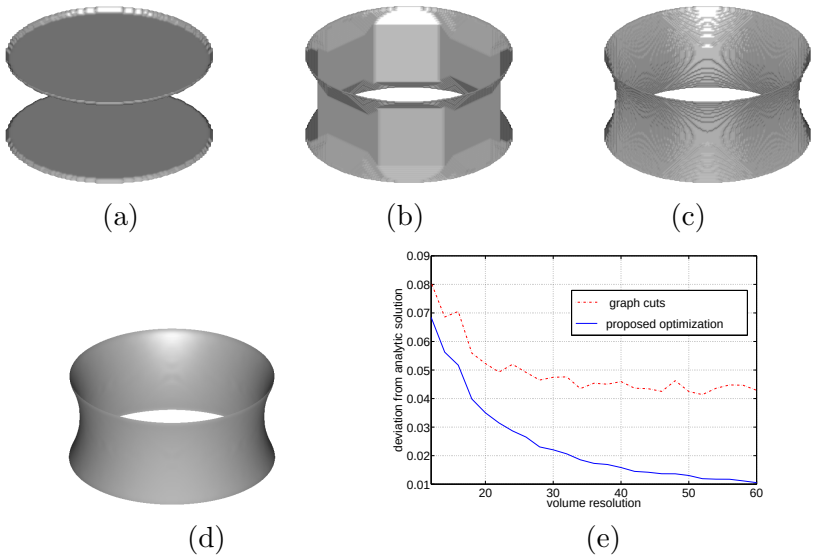

Figure 8. Continuous vs. discrete shape optimization (see text). (a) Graph cut reconstruction with a 6-neighborhood system at a volume resolution of $180 \times 180 \times 60$ (the highest in the plot). (b) Graph cut reconstruction with a 26-neighborhood system at the same volume resolution. (c) Surface produced by the proposed optimization technique. (d) Known analytic solution. (e) Deviation of the recovered surface from the analytic ground-truth for increasing volume resolution. The experiment demonstrates that graph cut solutions can indeed be improved by reverting to larger neighborhood connectivity (26 instead of 6 neighbors). Yet, for any connectivity there is a metrication error, which persists with increasing resolution. The proposed continuous global optimization, on the other hand, is consistent as the discretization error decays to zero.

better approximated in the latter case (Kolmogorov and Zabih, 2002). However, discretization artifacts are still visible in terms of polyhedral blocky structures. In fact, for a fixed connectivity structure the computed graph cut solution is not consistent with respect to the volume resolution in contrast to the solution of the proposed continuous minimization. This is demonstrated in Fig. 8 (e), where for both optimization models the deviation of the estimated surface from the analytic ground-truth is plotted for increasing spatial resolution. This measure was computed in terms of the Hausdorff metric

$$\epsilon = \int_{S_{true}} d(x, S_{num}) d^2 x, \tag{27}$$

where S_{true} and S_{num} denote the ground-truth and the computed numerical solution respectively, and $d(x, S)$ is the distance from a point x to the nearest point on S. As expected, the proposed continuous model produces shapes that converge to the analytic one. In contrast, the deviation of the graph cut generated surfaces contains a constant error that is independent of the spatial resolution. Although the reached value can be improved by increasing the graph connectivity, the discrete model will always exhibit an asymptotic behavior for a fixed graph structure.

To further demonstrate the practical applicability of the proposed convex optimization technique, we provide an additional comparison to graph cuts on the "dinoRing" data set, shown in Fig. 9. For the sake of a fair comparison, we ran both optimization

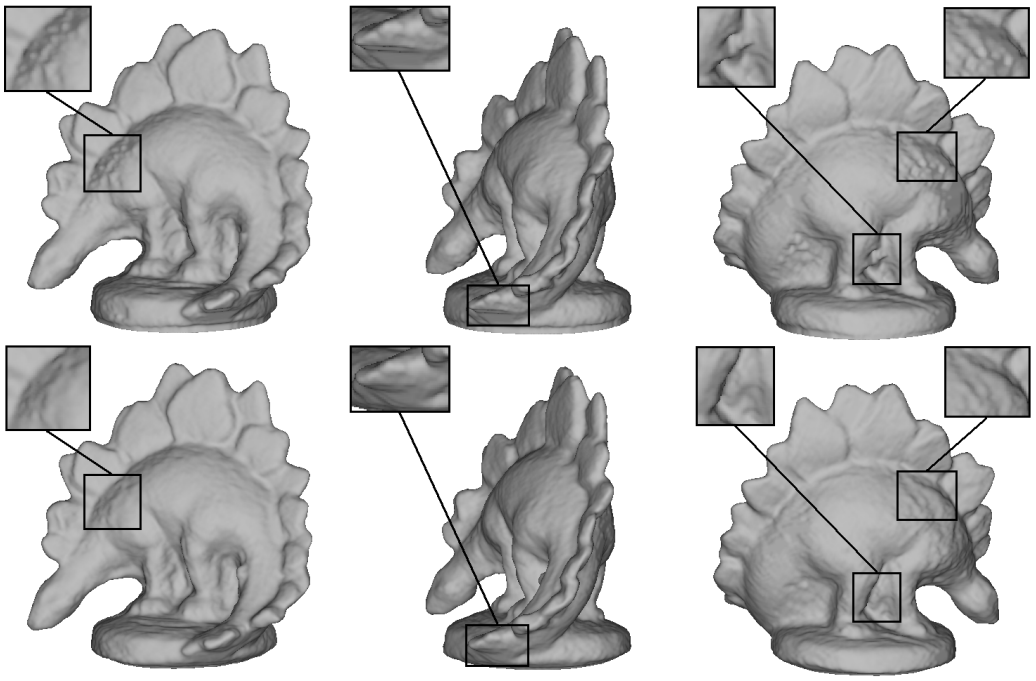

Figure 9. Comparison between graph cuts and the proposed continuous convex optimization on the "dinoRing" data set. *First row:* Graph cut reconstruction at volume resolution 256^3. *Second row:* Surface generated with the proposed optimization technique at the same resolution. Both reconstructions are computed on the same cost volumes, obtained with energy model III. Note that the presented continuous method leads to visual improvements at areas of noisy data due to the lack of texture or occlusions.

Table II. Quantitative evaluation of the reconstructions shown in Fig. 9.

optimization technique	**neighborhood system**	**completeness**	**accuracy**	**runtime**
graph cuts (CPU)	6	99.2 %	0.44 mm	41 s
convex TV (CPU)	6	99.4 %	0.43 mm	588 s
convex TV (GPU)	-	-	-	23 s

techniques on the same cost volumes at resolution 256^3, obtained with energy model III. A graph structure of 6-connectivity was used for the graph cuts due to memory restrictions. Note however that the continuous method also relies on a 6-neighborhood system to impose surface smoothness. At first glance, both reconstructions look similar. However, a closer look reveals that the presented continuous approach achieves more success in suppressing noise. This leads to visual improvements at areas of inaccurate data due to lacking texture or erroneous occlusion handling. A quantitative evaluation of both reconstructions is shown in Table II. As expected, the convex optimization registers minor improvements. Note that the difference between both techniques consists in the representation of surface regularization. Hence, the overall quality of the

reconstructions is determined by the utilized data terms. However, we argue that in case of noisy data, when surface smoothing becomes crucial, a continuous PDE-based approach should be preferred over a discrete one.
Apart from metrication errors, the proposed continuous optimization method entails additional practical advantages like parallelizability, which allows for a GPU implementation. Table II lists also the runtimes of both optimization techniques. For the graph cuts, only a CPU implementation is proposed, since a GPU implementation of classical graph cut algorithms is not straightforward. The continuous convex optimization was implemented for both CPU and GPU and carried out on a PC with a NVIDIA GeForce GTX 280 graphics card. The CPU runtimes were measured on a 2.66 GHz Intel Core2 architecture. Although discrete approaches are generally faster on the CPU due to their non-iterative nature, they do not make use of recent progress in parallel computing. Moreover, continuous methods involve a considerable reduction of memory requirements compared to graph cuts (in our implementation about a factor of 20), which allows to perform global minimization at higher volume resolutions. For that reasons, such techniques seem to have more potential in the long run in time- and memory-consuming applications like shape optimization. A detailed recent discussion on these issues can be found in (Klodt et al., 2008).

6. Conclusion

We cast multiview 3D reconstruction as a continuous convex optimization problem. As for graph cuts this allows to compute globally optimal shapes. However, in contrast to discrete techniques, the proposed continuous formulation does not suffer from metrication errors and requires considerably less memory (about a factor of 20), thereby allowing for optimal reconstructions at higher resolutions. In particular, we considered three different energy models, that can be optimized with the presented approach. While the first two models are based on established paradigms, the third one introduces the concept of propagated photoconsistency, thereby addressing some of the shortcomings of classical methodologies. In both qualitative and quantitative experiments we demonstrated that precise and spatially consistent reconstructions can be computed by minimizing continuous convex functionals.

Acknowledgements

We gratefully acknowledge funding by the German Research Foundation (DFG) under the projects CR-250/1-1 and CR-250/1-2. We thank Martin Oswald for helping us with the visualization.

References

Appleton, B. and H. Talbot: 2005, 'Globally Optimal Geodesic Active Contours'. *J. Math. Imaging Vis.* **23**(1), 67–86.

Appleton, B. and H. Talbot: 2006, 'Globally Minimal Surfaces by Continuous Maximal Flows'. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1), 106–118.

Boykov, Y. and V. Lempitsky: 2006, 'From Photohulls to Photoflux Optimization'. In: *Proc. British Machine Vision Conference*, Vol. 3. pp. 1149–1158.

Bresson, X., S. Esedoḡlu, P. Vandergheynst, J. P. Thiran, and S. Osher: 2005, 'Global minimizers of the active contour/snake model'. Technical Report CAM-05-04, Department of Mathematics, University of California at Los Angeles, CA, U.S.A.

Caselles, V., R. Kimmel, and G. Sapiro: 1995, 'Geodesic active contours'. In: *Proc. Fifth International Conference on Computer Vision*. Cambridge, MA, pp. 694–699, IEEE Computer Society Press.

Chambolle, A.: 2005, 'Total Variation Minimization and a Class of Binary MRF Models'. In: *EMMCVPR*. pp. 136–152.

Chan, T., S. Esedoḡlu, and M. Nikolova: 2006, 'Algorithms for finding global minimizers of image segmentation and denoising models'. *SIAM Journal on Applied Mathematics* **66**(5), 1632–1648.

Curless, B. and M. Levoy: 1996, 'A volumetric method for building complex models from range images'. In: *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, NY, USA, pp. 303–312, ACM Press.

Duan, Y., L. Yang, H. Qin, and D. Samaras: 2004, 'Shape Reconstruction from 3D and 2D Data Using PDE-Based Deformable Surfaces'. In: *Proc. European Conference on Computer Vision*. pp. 238–251.

Faugeras, O. and R. Keriven: 1998, 'Variational principles, surface evolution, PDE's, level set methods, and the stereo problem'. *IEEE Transactions on Image Processing* **7**(3), 336–344.

Greig, D., B. Porteous, and A. Seheult: 1989, 'Exact maximum a posteriori estimation for binary images'. *Journal of the Royal Statistical Society B* **51**(2), 271–279.

Hernandez, C. and F. Schmitt: 2004, 'Silhouette and stereo fusion for 3D object modeling'. *Computer Vision and Image Understanding* **96**(3), 367–392.

Hernández, C., G. Vogiatzis, and R. Cipolla: 2007, 'Probabilistic visibility for multi-view stereo'. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. Minneapolis, Minnesota, USA, IEEE Computer Society.

Hornung, A. and L. Kobbelt: 2006, 'Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding'. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. New York, NY, USA, pp. 503–510.

Hu, T. C.: 1969, *Integer Programming and Network Flows*. Reading, MA: Addison-Wesley.

Kass, M., A. Witkin, and D. Terzopoulos: 1988, 'Snakes: Active contour models'. *International Journal of Computer Vision* **1**, 321–331.

Kichenassamy, S., A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi: 1995, 'Gradient flows and geometric active contour models'. In: *Proc. Fifth International Conference on Computer Vision*. Cambridge, MA, pp. 810–815, IEEE Computer Society Press.

Kirsanov, D. and S. Gortler: 2004, 'A Discrete Global Minimization Algotithm for Continuous Variational Problems'. In: *Harvard Computer Science Technical Report: TR-14-04*.

Klodt, M., T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers: 2008, 'An Experimental Comparison of Discrete and Continuous Shape Optimization Methods'. In: *European Conference on Computer Vision (ECCV)*. Marseille, France.

Kolev, K., T. Brox, and D. Cremers: 2006, 'Robust variational segmentation of 3D objects from multiple views'. In: K. F. et al. (ed.): *Pattern Recognition (Proc. DAGM)*, Vol. 4174 of *LNCS*. Berlin, Germany, pp. 688–697, Springer.

Kolev, K., M. Klodt, T. Brox, and D. Cremers: 2007a, 'Propagated Photoconsistency and Convexity in Variational Multiview 3D Reconstruction'. In: *Workshop on Photometric Analysis for Computer Vision*. Rio de Janeiro, Brazil.

Kolev, K., M. Klodt, T. Brox, S. Esedoglu, and D. Cremers: 2007b, 'Continuous Global Optimization in Multiview 3D Reconstruction'. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, Vol. 4679 of *LNCS*. EZhou, China, pp. 441–452, Springer.

Kolmogorov, V. and Y. Boykov: 2004, 'An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(9), 1124–1137.

Kolmogorov, V. and R. Zabih: 2002, 'What Energy Functions Can Be Minimized via Graph Cuts?'. In: *Proc. European Conference on Computer Vision*. London, UK, pp. 65–81.

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

Labatut, P., J.-P. Pons, and R. Keriven: 2007, 'Efficient multi-view reconstruction of large-scale scenes using interest points, Delaunay triangulation and graph cuts'. In: *Proc. International Conference on Computer Vision*. Rio de Janeiro, Brazil.

Lempitsky, V., Y. Boykov, and D. Ivanov: 2006, 'Oriented visibility for multiview reconstruction'. In: *Proc. European Conference on Computer Vision*, Vol. 3953 of *LNCS*. pp. 226–238.

Mumford, D. and J. Shah: 1989, 'Optimal approximations by piecewise smooth functions and associated variational problems'. *Communications on Pure and Applied Mathematics* **42**, 577–685.

Pons, J.-P., R. Keriven, and O. Faugeras: 2007, 'Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score'. *International Journal of Computer Vision* **72**(2), 179–193.

Rudin, L. I., S. Osher, and E. Fatemi: 1992, 'Nonlinear total variation based noise removal algorithms'. *Physica D* **60**, 259–268.

Seitz, S., B. Curless, J. Diebel, D. Scharstein, and R. Szeliski: 2006, 'A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms'. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA, pp. 519–528, IEEE Computer Society.

Seitz, S. and C. Dyer: 1997, 'Photorealistic scene reconstruction by voxel coloring'. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. pp. 1067–1073.

Sethian, J. A.: 1996, *Level Set Methods*. Cambridge, UK: Cambridge University Press.

Sinha, S., P. Mordohai, and M. Pollefeys: 2007, 'Multiview Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh'. In: *Proc. International Conference on Computer Vision*. Rio de Janeiro, Brazil.

Soatto, S., A. J. Yezzi, and H. Jin: 2003, 'Tales of Shape and Radiance in Multi-view Stereo'. In: *Proc. International Conference on Computer Vision*. Washington, DC, USA, p. 974, IEEE Computer Society.

Strang, G.: 1983, 'Maximal flow through a domain'. *Mathematical Programming* **26**, 123–243.

Tran, S. and L. Davis: 2006, '3D surface reconstruction using graph cuts with surface constraints'. In: *Proc. European Conference on Computer Vision*, Vol. 3952 of *LNCS*. pp. 219–231.

Vogiatzis, G., C. H. Esteban, P. H. S. Torr, and R. Cipolla: 2007, 'Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency'. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2241–2246.

Vogiatzis, G., P. Torr, and R. Cippola: 2005, 'Multi-view stereo via volumetric graph-cuts'. In: *Proc. International Conference on Computer Vision and Pattern Recognition*. pp. 391–399.

Zach, C., T. Pock, and H. Bischof: 2007, 'A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration'. In: *Proc. International Conference on Computer Vision*. Rio de Janeiro, Brazil.