

Multiview Stereo and Silhouette Consistency via Convex Functionals over Convex Domains

Daniel Cremers and Kalin Kolev

Abstract—We propose a convex formulation for silhouette and stereo fusion in 3D reconstruction from multiple images. The key idea is to show that the reconstruction problem can be cast as one of minimizing a *convex* functional, where the exact silhouette consistency is imposed as convex constraints that restrict the domain of feasible functions. As a consequence, we can retain the original stereo-weighted surface area as a cost functional without heuristic modifications of this energy by balloon terms or other strategies, yet still obtain meaningful (nonempty) reconstructions which are guaranteed to be silhouette-consistent. We prove that the proposed convex relaxation approach provides solutions that lie within a bound of the optimal solution. Compared to existing alternatives, the proposed method does not depend on initialization and leads to a simpler and more robust numerical scheme for imposing silhouette consistency obtained by projection onto convex sets. We show that this projection can be solved exactly using an efficient algorithm. We propose a parallel implementation of the resulting convex optimization problem on a graphics card. Given a photoconsistency map and a set of image silhouettes, we are able to compute highly accurate and silhouette-consistent reconstructions for challenging real-world data sets. In particular, experimental results demonstrate that the proposed silhouette constraints help to preserve fine-scale details of the reconstructed shape. Computation times depend on the resolution of the input imagery and vary between a few seconds and a couple of minutes for all experiments in this paper.

Index Terms—Image-based modeling, silhouette and stereo fusion, convex optimization.

1 INTRODUCTION

1.1 Multiview Reconstruction Using Stereo and Silhouettes

RECOVERING 3D geometrical structure from a series of calibrated images is among the fundamental problems in computer vision, with numerous applications in computer graphics, augmented reality, robot navigation, and tracking. Among a variety of existing methods for multiview reconstruction, one can identify two major classes of approaches: shape from silhouettes and shape from stereo.

Historically, the first strategy for multiview 3D shape retrieval, dating back to the 1970s, has been to use the outlines of the imaged objects [2]. Most of these *shape from silhouettes* approaches aim at approximating the visual hull [19] of the observed solid. The visual hull is an outer approximation, constructed as the intersection of the visual cones associated with all image silhouettes. In the course of research, different shape representations have been proposed: volumetric [21], surface-based [5], and polyhedral [11]. Apart from shape representation, research has also been focused on the development of methods operating on raw image data instead of predetermined silhouettes. Most of them are based on an energy minimization framework allowing us to impose regularization in the labeling process

[29], [33]. The segmentation of each image is obtained through the evolution of a single surface in 3D rather than separate contours in 2D. As a result, such approaches exhibit considerable robustness to image noise and erroneous camera calibration.

The main drawback of silhouette-based approaches is their inability to reconstruct concavities since these do not affect the silhouettes. *Stereo-based methods* capture such indentations by measuring the photoconsistency of surface patches in space. The fundamental idea is that under the Lambertian assumption, only points on the object's surface have a consistent appearance in the input images, while all other points project to incompatible image patches. The earliest algorithms use carving techniques to obtain a volumetric representation of the scene by repeatedly eroding inconsistent voxels [25]. Yet, such greedy methods do not explicitly enforce the smoothness of the surface, which often results in rather noisy reconstructions. Subsequently, proposed energy minimization techniques [10], [7], [20] overcome this drawback. They typically aim at computing a weighted minimal surface, where the weights reflect the local photoconsistency.

Such minimal surface approaches have two fundamental limitations: First, the inherent coupling of data term and regularization often produces oversmoothing at locations of thin structure [30], [14]. Although this bias toward surfaces with small area can be alleviated by utilizing regional terms [3], [14], [18], it still persists when a smoothness prior is applied. Second, the global minimizer of minimal surface functionals is the empty set, a solution that clearly violates the silhouette constraints.

Ideally, one would like to combine the advantages of silhouette and stereo-based techniques so as to compute reconstructions which do allow concavities and yet are

- The authors are with the Department of Computer Science, Technical University of Munich, Boltzmanstrasse 3, 85748 Garching bei München, Germany. E-Mail: {daniel.cremers, kalin.kolev}@in.tum.de.

Manuscript received 18 Feb. 2010; revised 17 June 2010; accepted 25 July 2010; published online 1 Sept. 2010.

Recommended for acceptance by F. Kahl.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2010-02-0107.

Digital Object Identifier no. 10.1109/TPAMI.2010.174.

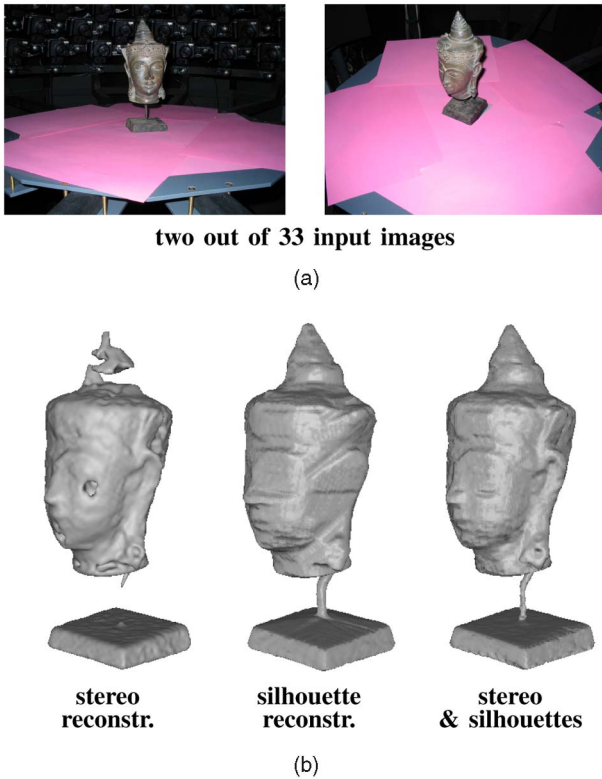


Fig. 1. Silhouette and stereo integration. While stereo-based reconstruction methods tend to remove thin structures and are heavily affected by specular reflections such as those of metal objects (a), silhouette-based methods cannot reconstruct indentations of the object surface since these do not appear in the silhouette. In this paper, we propose fusing stereo and silhouette information in a convex optimization framework in order to obtain stereo-based reconstructions which are guaranteed to be silhouette-consistent. As a consequence, the resulting algorithm allows us to restore concave areas (around the ears) as well as fine geometric details such as the pedestal (b).

consistent with all observed silhouettes—see Fig. 1. A simple strategy to fuse these complementary features is to use a visual hull (computed from silhouettes) as initialization for a stereo-based approach [22], [32]. First, this requires constraining the solution space to avoid the empty set. Second, the resulting reconstruction will generally no longer fulfill the silhouette constraints. Alternatively, one can unify both information sources in a single formulation. Two different techniques have been proposed to achieve this goal: One can integrate in the evolution silhouette-aligning forces [6], [9], [12] or one can use predetermined surface points [15], [30], [26] to impose exact silhouette constraints. Both strategies have their shortcomings. The first one may lead to a numerically unstable behavior and introduces a bias near the visual hull boundary, while the second one requires premature decisions about voxel occupancy. A mathematically elegant alternative to fuse silhouette and stereo information is stated by the stereoscopic segmentation model [13], [33]. Appropriate evolution terms enforcing photoconsistency and silhouette consistency criteria are thereby derived by means of the derivative of the reprojection error of the estimated shape. However, due to the complexity of the model, it is difficult to overcome local optimization,

which is prone to undesired local minima.¹ To address this difficulty and move apart from local evolution schemes, Sinha and Pollefy [27] proposed a graph cut framework for silhouette and stereo fusion. Unfortunately, the practical applicability of this method is limited due to its high memory requirements. This poses a severe restriction on the volume resolution at which reconstructions can be computed. Finally, we can conclude that the development of robust and efficient schemes for silhouette and stereo integration remains an open challenge.

1.2 Contribution

In this paper, we propose a mathematically transparent framework for silhouette and stereo fusion in 3D reconstruction. The idea is to cast multiview stereovision as a *convex* variational problem where exact silhouette consistency is imposed by means of convex constraints that restrict the domain of feasible functions. Silhouette-consistent reconstructions are computed by convex relaxation, finding global minimizers of the relaxed problem, and subsequent projection onto the original nonconvex set. Compared to existing fusion techniques, we thus compute guaranteed silhouette-consistent reconstructions without constraining the search space and without extending the original stereo-weighted cost functional by heuristic ballooning terms or more sophisticated balancing terms. Compared to classical local fusion techniques, the proposed formulation does not depend on initialization and leads to a more tractable numerical scheme by removing the bias near the visual hull boundary. In experiments on several challenging real data sets (including metallic objects, low-texture objects, and complex objects with fine-scale details), we show the advantages of imposing silhouette consistency in the reconstruction of small-scale structures which cannot be restored by state-of-the-art stereo algorithms.

A preliminary version of this paper was presented at the European Conference on Computer Vision [17]. In addition to the conference version, in this paper, we provide an exact algorithm that allows us to efficiently compute the projection onto the set of silhouette-consistent solutions. Moreover, we explore the convergence properties of the employed numerical optimization scheme in more detail.

The paper is organized as follows: In the next section, we review the formulation of stereo-based multiview reconstruction as a weighted minimal surface problem. In Section 3, we show that the integration of stereo and silhouette constraints can be formulated as a problem of minimizing a convex functional over the convex set of silhouette-consistent functions. In Section 5, we provide details on the numerical implementation of the constrained optimization problem, including two alternative methods to impose silhouette consistency. In Section 6, we show experimental results on several real data sets which demonstrate the advantages of silhouette consistency for the reconstruction of fine-scale structures and emphasize the superiority of the proposed approach over traditional techniques.

1. An example of an undesired local minimum is the case of a torus where the initialization consists of a disk enclosing it. Since local deformations (indentation) of the disk only increase the overall surface area without decreasing the reprojection error, the local minimization procedure will not give rise to a torus but, rather, remain stuck with the disk—see [16] for an experimental validation.

2 MULTIVIEW 3D RECONSTRUCTION AS A WEIGHTED MINIMAL SURFACE PROBLEM

Let $V \subset \mathbb{R}^3$ be a volume which contains the scene of interest and

$$I_1, \dots, I_n : \Omega \rightarrow \mathbb{R}^3 \quad (1)$$

be a collection of calibrated color images with perspective projections π_1, \dots, π_n . Let

$$S_1, \dots, S_n \subset \Omega \quad (2)$$

be the observed silhouettes of the 3D object and let

$$\rho : V \rightarrow (0, 1] \quad (3)$$

be a photoconsistency map measuring the discrepancy among various image projections. In particular, low values of $\rho(x)$ indicate strong agreement from different cameras on the observed image patches, indicating a high likelihood that the surface passes through the given point. It should be noted that the photoconsistency values generally depend on the surface S to be estimated, which determines the projective warping of corresponding image patches. To this end, researchers proposed explicitly modeling this dependence by incorporating the surface normal as an additional argument [10], [13]. However, this leads to higher order terms and significantly exacerbates the optimization. More details on the computation of photoconsistency are given in Section 5.1.

With the above definitions, multiview reconstruction can be done by minimizing the classical energy [10]

$$E(S) = \int_S \rho(x) dA(x), \quad (4)$$

where $dA(x)$ denotes the infinitesimal surface element at x . The reconstruction is therefore given by a minimal surface measured in a Riemannian metric that favors boundaries along photoconsistent locations. While *local* optimization techniques (using coarse-to-fine strategies) provide useful reconstructions, there is little guarantee regarding the optimality of the solutions. In fact, the question of optimality is somewhat meaningless as the global minimum of (4) is obviously the empty set. This indicates that the minimization problem (4) in itself is not a meaningful approach to multiview reconstruction since surfaces of lower energy do not necessarily correspond to better reconstructions. A remedy to the above problem is to either to constrain the search space around the visual hull [32] or add regional balancing terms to the cost functional using balloon forces [20] or heuristically constructed regional terms [14], [18]. Global optima of respective cost functionals can then be computed either in a spatially discrete setting using graph cuts [32], [20] or in a spatially continuous setting using convex relaxation techniques [18].

Nevertheless, there are two grave limitations of such methods. First, the balancing regional terms are typically based on a number of somewhat heuristic assumptions and can often introduce a bias in the resulting reconstruction—favoring surfaces of larger volume, for example. Second, the resulting reconstructions are not guaranteed to be silhouette-consistent in the sense that the projections of the estimated surface do not generally coincide with the observed silhouettes.

3 CONVEX INTEGRATION OF SILHOUETTES AND STEREO

3.1 Imposing Silhouette Consistency

We saw that the classical approach to multiview reconstruction via minimization of a stereo-weighted minimal surface energy (4) by itself is not meaningful since the lowest energy solution is the empty set. Yet, obviously the empty set is not a desired solution. How can we avoid this trivial solution while retaining a required grade of smoothness? If we are given the silhouettes of the observed object in each of the images, then we can impose silhouette alignment of the computed shape during the energy minimization process. This naturally prevents the collapse of the surface since the empty set is clearly not a silhouette-consistent solution.

More specifically, we propose to solve the following constrained optimization problem:

$$\begin{aligned} \min_S \int_S \rho(x) dA(x), \\ \text{s.t. } \pi_i(S) = S_i \quad \forall i = 1, \dots, n. \end{aligned} \quad (5)$$

Unfortunately, the above optimization is highly non-convex: Depending on the values of the photoconsistency function ρ , variations in S may give rise to an arbitrary increase or decrease in the cost functional. Furthermore, imposing silhouette consistency in the above representation is not a trivial task.

3.2 Convex Relaxation

Surprisingly, these difficulties can be removed by reverting to an implicit representation of the surface S . In order to cast (5) as a convex optimization problem, the surface S is represented by the characteristic function $u : V \rightarrow \{0, 1\}$ of its interior S_{int} . Hence, changes in the topology of S are handled automatically without reparameterization.

With this implicit surface representation, the optimization problem (5) in S is equivalent to the following optimization problem in the binary labeling u :

$$\min_{u \in \mathcal{D}'} \int_V \rho(x) |\nabla u(x)| d^3x, \quad (6)$$

where

$$\mathcal{D}' = \left\{ u : V \rightarrow \{0, 1\} \left| \begin{array}{l} \int_{R_{ij}} u(x) dx \geq \delta, \quad \text{if } j \in S_i \forall i, j, \\ \int_{R_{ij}} u(x) dx = 0, \quad \text{if } j \notin S_i \forall i, j, \end{array} \right. \right\}, \quad (7)$$

where R_{ij} denotes the visual ray through pixel j of image i . In this implicit formulation, the silhouette consistency constraint in (5) gives rise to equality and inequality constraints. These constraints are based on the following reasoning: For all visual rays from the camera center passing through a pixel inside a silhouette, *at least one* of the voxels along this ray should be occupied, whereas for a visual ray passing through a pixel outside a silhouette, *all* voxels along the respective ray should be empty (see Fig. 2). In the spatially continuous formulation adopted in this paper (where there exists no notion of discrete *voxels*), the constant δ in (7) denotes the thickness of material below

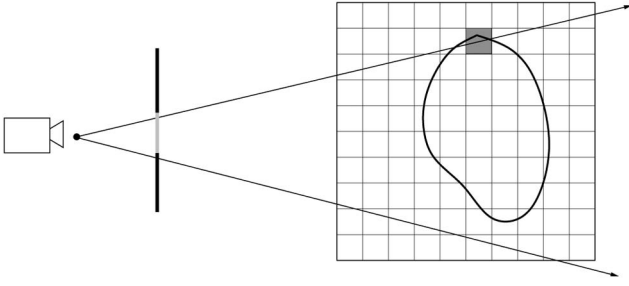


Fig. 2. Schematic views of silhouette consistency. A 2D visualization of the object surface and its projection onto the image shows that for a silhouette-consistent shape, at least one voxel along each visual ray through a silhouette pixel is occupied, whereas all voxels along rays through nonsilhouette pixels are empty. The bright area on the image plane indicates the outlines of the observed object and the shaded voxel is an occupied one along the given viewing ray.

which the considered object becomes translucent. It should be noted that this material-dependent parameter δ gives rise to a more realistic physical model than imposing

$$\int_{R_{ij}} u(x) dx > 0 \quad \text{if } j \in S_i \forall i, j.$$

Moreover, while the latter also leads to a convex set, this set is open such that projections are not defined and the existence of solutions cannot be assured. In discrete implementations, we simply choose $\delta = 1$, which we also do in the following to simplify the notation.

There are a few important remarks regarding the formulation in (6) that should be made. In the case of imperfect silhouettes, the above constraints can be applied only to those pixels for which we have certain confidence about being inside or outside a silhouette. Note that this property is in contrast to most of the existing approaches, which require apparent contours of the imaged object to be provided. In a discrete setting, the volume resolution should be determined according to the given image resolution. When the respective volume sampling is too coarse, some discretization artifacts could appear in the reconstruction. Unfortunately, increasing the voxel resolution entails a considerable computational and memory burden and is not always possible in practice. In such cases, subsampling the input silhouettes could be helpful. Note that the constraints in (6) could give an empty set. This can happen when all voxels along a visual ray passing through a silhouette pixel in one of the images project to background in another image. Although this is unlikely to happen in a real scenario, it can occur in the case of noisy silhouette input. Such difficulties can be circumvented by restricting the computations to the visual hull of the object and holding the values of u for all other voxels fixed.

Now, let us concentrate on the optimization of the energy model in (6). Interestingly, the functional is convex. Yet, due to the constraint that u is a binary-valued function, the overall minimization problem (6) is nonconvex (because the space of binary functions is nonconvex). This can be resolved by relaxing the binary constraint and allowing the function u to take on values in the interval $[0, 1]$. The relaxed

problem therefore becomes that of minimizing a convex functional over a convex set:

$$\min_{u \in D} \int_V \rho(x) |\nabla u(x)| d^3x, \quad (8)$$

where

$$D = \left\{ u : V \rightarrow [0, 1] \left| \begin{array}{l} \int_{R_{ij}} u(x) dx \geq 1, \quad \text{if } j \in S_i \forall i, j, \\ \int_{R_{ij}} u(x) dx = 0, \quad \text{if } j \notin S_i \forall i, j, \end{array} \right. \right\} \quad (9)$$

is the set of continuous-valued functions u which are silhouette-consistent with respect to all images i and all rays j . Again, the corresponding constraints follow the formulation in (7).

Surprisingly, in this implicit representation of the shape S by a relaxed labeling function $u : V \rightarrow [0, 1]$, the set D of silhouette-consistent configurations has the following favorable property:

Proposition 1. *The set D of all silhouette-consistent functions defined in (9) forms a convex set.*

Proof. In order to show the convexity, let $u_1, u_2 \in D$ be two elements of D . Then, any convex combination $u = \alpha u_1 + (1 - \alpha)u_2$ with $\alpha \in [0, 1]$ is also an element in D . In particular, $u(x) \in [0, 1]$ for all x . Moreover,

$$\int_{R_{ij}} u dx = \alpha \int_{R_{ij}} u_1 dx + (1 - \alpha) \int_{R_{ij}} u_2 dx \geq 1 \text{ if } j \in S_i,$$

and similarly,

$$\int_{R_{ij}} u dx = \alpha \int_{R_{ij}} u_1 dx + (1 - \alpha) \int_{R_{ij}} u_2 dx = 0 \text{ if } j \notin S_i.$$

Thus, $u \in D$. \square

The above statement implies that a global minimum u^* of the relaxed problem (8) exists and can be computed, for example, by a simple gradient descent procedure or by more efficient numerical schemes. A necessary condition for a minimum of (8) is stated by the associated Euler-Lagrange equation:

$$0 = \operatorname{div} \left(\rho \frac{\nabla u}{|\nabla u|} \right) = \rho \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) + \left\langle \nabla \rho, \frac{\nabla u}{|\nabla u|} \right\rangle. \quad (10)$$

A numerical solution to this partial differential equation within the domain of admissible functions specified by the convex constraints in (8) is detailed in Section 5.

3.3 Binary Solution via Thresholding

Since we are interested in minimizers of the nonconvex binary labeling problem (6), a straightforward methodology is to threshold the solution of the convex problem appropriately. Although this does not guarantee finding the global minimum of (6), the proposed strategy entails a series of advantages compared to classical local optimization techniques. Intuitively, extending the set of feasible functions, computing the global minimum over this domain, and subsequently projecting to the nearest point within the original set is expected to give a more accurate estimate than a simple gradient descent procedure for

smooth functionals. In particular, the proposed methodology has the following advantages:

- It allows us to incorporate *exact* silhouette constraints without making premature hard decisions about voxel occupancy along each viewing ray passing through a silhouette pixel.
- It does not depend on initialization since the relaxed functional is optimized globally.
- It leads to a simple and tractable numerical scheme which does not rely on a locally estimated surface orientation and thus does not introduce a bias near the visual hull boundary.
- The algorithm provides a solution to the binary optimization problem (6) which lies within an energetic bound of the optimal solution. This is stated in the following proposition:

Proposition 2. *Let E denote the functional in (8) and u^* be a minimizer. Let $D' \subset D$ be the set of binary silhouette-consistent functions defined in (7). Furthermore, let $u' \in D'$ be the (global) minimum of (6) and \tilde{u} the solution obtained with the above procedure. Then, the computed solution lies within an energetic bound of the optimum.*

Proof. Since $D' \subset D$, we have $E(u^*) \leq E(u')$, and hence,

$$E(\tilde{u}) - E(u') \leq E(\tilde{u}) - E(u^*),$$

where the right-hand side can easily be computed. \square

The projection $\tilde{u} \in D'$ of a minimizer u^* onto D' can be computed by simple thresholding

$$\tilde{u}(x) = \begin{cases} 1, & \text{if } u^*(x) \geq \mu, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where

$$\mu = \min \left\{ \left(\min_{i \in \{1, \dots, n\}, j \in S_i} \max_{x \in R_{ij}} u^*(x) \right), 0.5 \right\}. \quad (12)$$

This threshold μ provides the closest (in any L^p -norm) silhouette-consistent binary function to the solution of the relaxed problem.²

Proposition 3. *The reconstructed surface exactly fulfills all silhouette constraints, i.e., $\tilde{u} \in D'$.*

Proof. Let R_{pq} be a given ray. For $q \notin S_p$, the silhouette constraint is fulfilled for any threshold $\mu \in (0, 1)$ since the labels $\tilde{u}(x)$ of all voxels x along the respective ray are 0. For $q \in S_p$, we have

$$\mu \leq \min_{i \in \{1, \dots, n\}, j \in S_i} \max_{x \in R_{ij}} u^*(x) \leq \max_{x \in R_{pq}} u^*(x).$$

This implies $\exists x \in R_{pq} : u^*(x) \geq \mu$, and hence, $\exists x \in R_{pq} : \tilde{u}(x) = 1$. \square

2. An anonymous reviewer pointed out that the above thresholding could alternatively be realized by computing the binary labeling function \tilde{u} of the lowest energy instead of the closest one. We tested this procedure and found out that it typically gives a threshold very close to the one specified in (12). As a result, both thresholds give rise to visually indistinguishable reconstructions.

In practice, thresholding by μ can be replaced by subtracting μ from the values of u^* and extracting the zero-level. This methodology has the advantage that it produces smoother surfaces than the thresholding procedure without affecting the energy which is defined on a voxel basis.

4 IMPOSING SILHOUETTE CONSISTENCY

4.1 A Set of Linear Constraints

The minimization of (8) should be performed within the specified domain of admissible functions. A straightforward way to achieve this is to project the current estimate after each iteration to the next point in the convex domain of permissible functions. This convex set is formed by three types of linear equality or inequality constraints.

1. At every location x , the function u must take on values within the domain $[0, 1]$.
2. For all pixels j of image i that lie outside the silhouette $S_i \subset \Omega$ for $i = 1, \dots, n$, the integral of u along that ray must vanish:

$$\int_{R_{ij}} u(x) dx = 0 \quad \text{if } j \notin S_i. \quad (13)$$

3. For all pixels j that lie inside the silhouette $S_i \subset \Omega$ for $i = 1, \dots, n$, the integral of u along that ray must be at least 1:

$$\int_{R_{ij}} u(x) dx \geq 1 \quad \text{if } j \in S_i. \quad (14)$$

4.2 A Simple Iterative Projection Scheme

Individually, each type of constraint can be easily imposed as follows: The first constraint is imposed by simply clipping the values of u to the interval $[0, 1]$ for all voxels x . The second type of constraint in (13) is imposed by simply setting all values of u along the respective ray S_i to zero. This concerns all voxels outside the visual hull of the imaged object. The third type of constraint in (14) requires more effort. Assume that it is violated for some pixel j in silhouette S_i , i.e., there exists some residual $\epsilon > 0$ such that

$$\int_{R_{ij}} u dx = 1 - \epsilon. \quad (15)$$

Then, the projection onto the space of functions where the respective constraint is fulfilled is obtained by simply adding that residual (in equal amounts) to the values of u as follows:

$$u_{new}(x) = u(x) + \frac{\epsilon}{\int_{R_{ij}} dx} \quad \forall x \in R_{ij}. \quad (16)$$

Iterating these individual projections can be done quite efficiently. It leads to a feasible configuration.

Enforcing these constraints in a different order will generally produce a different result. Since the constraints defined above are generally not orthogonal, sequentially projecting onto each respective hyperplane will not give rise to the euclidean projection onto the convex domain. It will

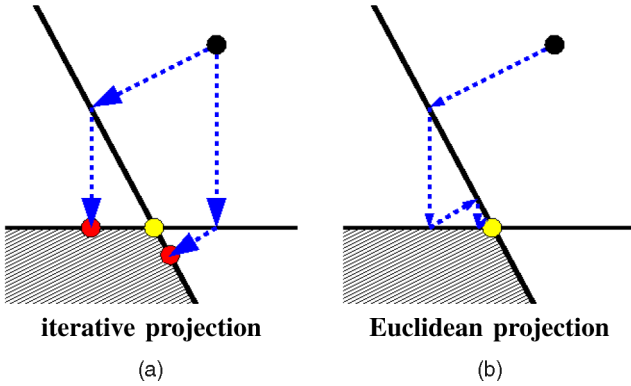


Fig. 3. Schematic view of projections of an initial configuration (black dot) onto two nonorthogonal linear constraints. While the sequential projection onto each individual constraint (a) leads to configurations (red dots) within the set of feasible solutions (shaded area), these configurations will generally not coincide with the euclidean projection onto the set (yellow dot). In addition, the final configuration depends on the order in which the projections are performed. In contrast, the recursive projection algorithm described in Section 4.3 allows us to compute the euclidean projection (b).

lead to a feasible solution but typically not the closest one among all. This can be seen in the schematic drawing of Fig. 3a: By sequentially projecting from an initial configuration (black dot) onto two nonorthogonal hyperplanes, one ends up with a solution (red dot) in the feasible set but not the closest one to the input configuration (yellow dot). In our scenario, this will happen if the same voxel lies on two rays that both violate the constraint. Depending on which of the two constraints is imposed first, we may obtain a different solution.

In practice, we find that when the evolution step is small enough, this issue is not crucial and the simple sequential projection often produces acceptable results. In order to avoid computations of ray-volume intersections any time the constraint is checked, one can compute the set of relevant voxels to each viewing ray in a preprocessing step and store them in lists. However, the size of this data structure could grow significantly when the resolution of input images is high. For this reason, in our implementation, we stored only the first and last voxels along each ray. Another important issue when using constraints is the frequency of enforcing them. In our implementation, we achieved stable behavior when applying the first constraint after each optimization iteration and the silhouette constraints after each 10 iterations.

Nevertheless, one may ask if there is a more elegant solution to impose the constraints since imposing each constraint sequentially as above not only depends on the order of projections, but is also likely to shrink the current solution more than necessary.

4.3 Euclidean Projection onto the Convex Set

Ideally, one would like to compute the euclidean projection onto the convex set of feasible solutions such that all constraints are taken into account simultaneously. Since there are a huge number of these constraints, a central challenge is to find an efficient algorithm to do this. Fortunately, such algorithms exist in the literature. In this work, we make use of a method that was first published by

Dijkstra and Boyle [8], [4]. By alternating projections onto each convex set in a recursive manner, this approach allows us to compute the euclidean projection onto the intersection of finite number of closed convex sets.

In our scenario, we first perform the clipping to the values $u(x) \in [0, 1]$ in order to satisfy constraint (1) above. As mentioned above, constraints of type (2) do not have to be handled explicitly and can be realized by restricting the computations to the visual hull of the imaged object. Let us now denote the projections associated with violated constraints of type (3) by Π_i for $i = 1, \dots, p$. More concretely, Π_i denotes the projection onto the hyperplane corresponding to constraint i . Furthermore, let \vec{u} denote a vector obtained by stacking the values of u at all voxels and let \vec{u}_{cur} be our current estimate corresponding to a function $u_{cur} \notin D$. We then iteratively compute a series of projections $\{\vec{u}_i^k\}$ and increments $\{\vec{v}_i^k\}$ as follows:

$$\begin{aligned} \vec{u}_0^k &= \vec{u}_p^{k-1}, \\ \vec{u}_i^k &= \Pi_i(\vec{u}_{i-1}^k - \vec{v}_i^{k-1}), \quad i = 1, 2, \dots, p, \\ \vec{v}_i^k &= \vec{u}_i^k - (\vec{u}_{i-1}^k - \vec{v}_i^{k-1}), \quad i = 1, 2, \dots, p, \end{aligned} \quad (17)$$

for $k = 1, 2, \dots$ with initial values $\vec{u}_p^0 = \vec{u}_{cur}$ and $\vec{v}_i^0 = \vec{0}$ for $i = 1, 2, \dots, p$.

Similarly to the simple procedure described in Section 4.2, this recursive algorithm also applies the individual projections Π_i sequentially. Yet, it systematically separates the residuals associated with various projection directions so as to make sure that the computed estimates \vec{u}_p^k gradually approach the euclidean projection onto the intersection of half-spaces.

Proposition 4. *The algorithm defined in (17) converges to the euclidean projection onto the intersection of convex sets.*

Proof. For the proof, we refer to [8], [4]. \square

Fig. 3b shows a visualization of this recursive projection method for the case of two nonorthogonal constraints.

It can be noted that the dimensionality of all involved variables in (17) is equal to the number of voxels and is usually in the order of multiple million. Hence, some care should be taken when implementing the above procedure to reduce the memory requirements. First, one can observe that the updates in \vec{u}_i^k only involve the previously projected values \vec{u}_{i-1}^k and thus can be evaluated on the fly. Moreover, the auxiliary vectors \vec{v}_i^k always have the direction of the normals to the corresponding hyperplanes. This implies that only their lengths can be stored instead of the full configurations.

5 IMPLEMENTATION

This section gives more details on the implementation of the proposed approach.

5.1 Photoconsistency Estimation

In this paper, we propose a novel strategy for integration of silhouette and stereo information. The presented method operates on a precomputed photoconsistency map $\rho: V \rightarrow [0, 1]$ and is independent of its particular implementation. To validate its concept, we used the voting scheme proposed in

[9] for photoconsistency computation. The choice of this technique was motivated by its robustness, even without explicit visibility estimation and increased accuracy compared to traditional methods. See [9] for more details.

5.2 Linearization and Fixed-Point Iteration

In order to solve (10), we suggest using a fixed-point iteration scheme that transforms the nonlinear system into a sequence of linear systems. These can be efficiently solved with an iterative solver like successive overrelaxation (SOR).

It can be observed that the only source of nonlinearity in (10) is the diffusivity $g := \frac{\rho}{|\nabla u|}$. Starting with an initialization u^0 , we can compute g and keep it constant. For constant g , (10) is linear and discretization yields a sparse linear system of equations of the form $A\vec{u} = \vec{0}$, which we solve with the SOR method. This means we iteratively compute an update of u at voxel i by

$$u_i^{l,k+1} = (1 - \omega)u_i^{l,k} + \omega \frac{\sum_{j \in \mathcal{N}(i), j < i} g_{i \sim j}^l u_j^{l,k+1} + \sum_{j \in \mathcal{N}(i), j > i} g_{i \sim j}^l u_j^{l,k}}{\sum_{j \in \mathcal{N}(i)} g_{i \sim j}^l}, \quad (18)$$

where $\mathcal{N}(i)$ denotes the six-neighborhood of i . Finally, $g_{i \sim j}$ denotes the diffusivity between voxel i and its neighbor j . It is defined as

$$g_{i \sim j}^l := \frac{g_i^l + g_j^l}{2}, \quad g_i^l := \frac{\rho_i}{\sqrt{|\nabla u_i^l|^2 + \epsilon^2}}, \quad (19)$$

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 specific problem at hand, we obtained a stable behavior for $\omega = 1.3$. After the linear solver yields a sufficiently good approximation (we iterated for $k = 1, \dots, 10$), 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.

The described optimization procedure shares similarities with previously introduced fixed-point iteration methods [31]. However, while [31] suggests completely solving the arising linear systems before updating the nonlinear diffusivity part, we argue for making the updates on the fly after a fixed number of iterations, which is a better strategy to address constrained minimization problems. In fact, the proposed numerical scheme can be interpreted as a specific quasi-Newton method (see the Appendix). Experimental evaluations showed that compared to established, provably convergent alternatives like gradient descent or primal-dual techniques [23], for the weighted minimal surface model at hand the proposed approach is substantially faster and achieves a higher degree of numerical stability even though it does not involve any line search procedures and does not require updating a data-dependent time-step parameter. These observations coincide with the conclusions in [31] and motivated our particular choice.

6 EXPERIMENTAL RESULTS

In the previous sections, we developed a convex optimization method to combine stereo and silhouette information in multiview reconstruction. In the following, we provide a series of experiments to assess the properties of the proposed method. Recently, impressive reconstructions of sufficiently textured objects acquired under controlled conditions with accuracy competing laser scanning techniques have been reported [24]. Although the presented approach achieves state-of-the-art accuracy, we believe that its main advantage is its robustness, i.e., the ability to operate in many practical situations where traditional methods fail. To emphasize this, we focus on four types of objects which exhibit important challenges for image-based modeling techniques, namely:

- We consider objects that are not Lambertian, such as metallic and shiny objects. For such objects, the key assumption underlying the stereo approach is that points on the surface have the same color when seen from different views is violated such that the stereo information alone is likely to provide suboptimal reconstructions.
- We consider objects with complex and fine-scale details. In traditional weighted minimal-surface approaches, elongated and fine-scale details are typically suppressed as this drastically reduces the surface area.
- We consider objects that exhibit little prominent texture. Such objects are known to be difficult for stereo approaches since the matching of similar structures no longer provides a reliable discrimination of good and bad matches.
- We consider image sequences acquired with a hand-held camera lacking perfect color and camera calibration. Analogously, the matching process is quite challenging under such conditions.

In addition, we compare the proposed method to other state-of-the-art multiview stereo methods (based on ballooning and propagated photoconsistency) and also to alternative techniques simultaneously imposing photometric and silhouette consistency.

6.1 Comparison to Two Alternative Multiview Stereo Methods

We validate the proposed approach on a scene of a head statue with complex reflection properties containing thin structures (the pedestal); see Fig. 4.³ Scenes of this type are a known challenge for variational stereo-based methods due to the violation of the Lambertian assumption and the presence of a regularizer, which introduces a bias toward shapes with small area. In particular, we implemented two classical paradigms in multiview stereo-vision: a weighted minimal surface formulation with a ballooning constraint [32] and a stereo propagating scheme [14], [18]. The first method produces clear over-smoothing effects at concavities and small-scale structures. The second approach retrieves shape indentations but also leads to erroneous carving at thin parts and specularities. In contrast, the introduced technique produces accurate

3. The image sequences used in Figs. 4 and 9 can be downloaded from <http://www.cremers.in.tum.de/index.php?nav=data>.

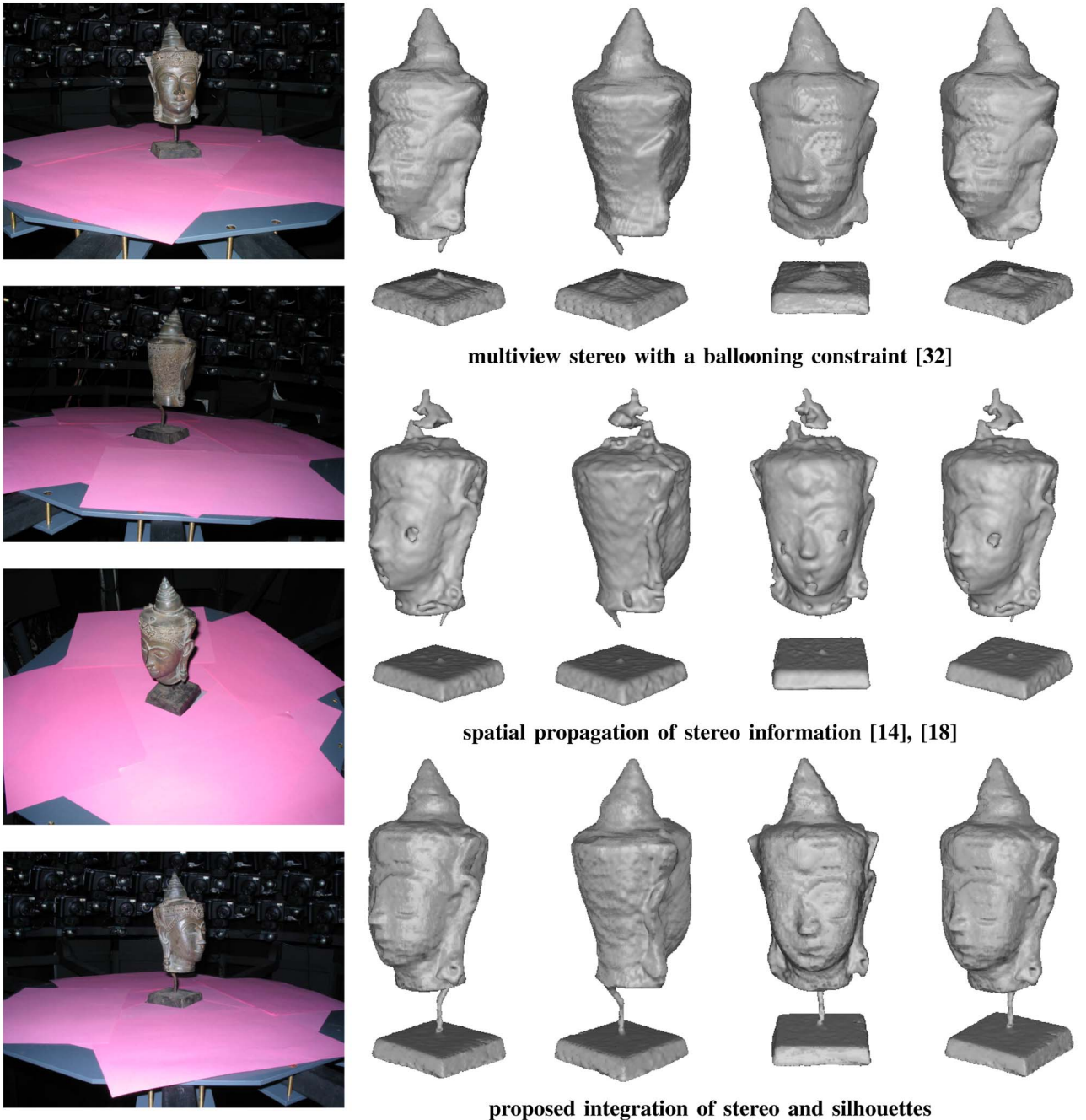


Fig. 4. Experimental comparison with alternative state-of-the-art methods on the reconstruction of a metallic head. While both the ballooning approach (upper row) and the more elaborate model based on volumetric photoconsistency (middle row) fail to recover the pedestal of the statue and produce oversmoothing effects or erroneous carving, the proposed approach (bottom row) accurately recovers all relevant details.

reconstructions of thin structures (the pedestal) as well as concave areas by incorporating silhouette constraints in the optimization process. Note that all three models are based on a classical minimal surface formulation but use different methodologies to avoid the empty surface as a solution. Note also that all three methods use silhouette information to restrict the ballooning, for initialization or to constrain the domain of admissible functions.

6.2 Absence of Bias toward the Visual Hull

The proposed method is based on minimizing a convex energy under convex constraints. Although the computed

solution does not depend on the initialization, in practice, we can accelerate computation by initializing the solution with the visual hull. In Fig. 5, we show intermediate steps in the evolution process of the proposed approach. Usually, local minimization techniques use the surface orientation to identify locations to deform the current shape in order to minimize the resulting reprojection error. However, this could lead to instabilities and introduce a bias near the visual hull boundary by involving surface points beyond the contour generator. In contrast, the introduced method recovers shape indentations effortlessly while retaining silhouette alignment during the optimization process.

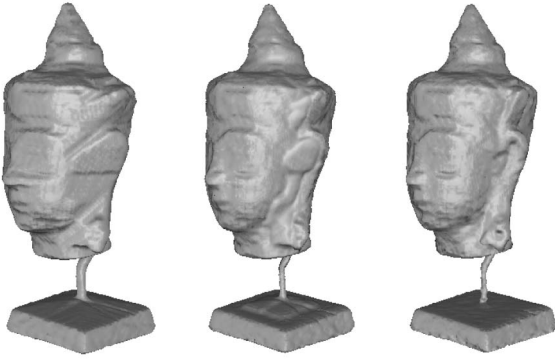


Fig. 5. Minimization process. Surface evolution starting from the visual hull, obtained by projecting the current estimate onto the original domain. Note that the presented method is able to generate accurate shapes starting from this initialization since it does not take the local surface orientation into account.

6.3 Comparison to the Multiview Stereo and Silhouette Fusion Approach of Sinha et al. [26]

Fig. 6 shows a comparison between the proposed convex relaxation method and the approach of [26] on an image sequence of a statue of a Greek goddess. The experiments show that both approaches give rise to a high-quality silhouette-consistent reconstruction. The proposed method, however, offers visible improvements in the area of the face and the creases of the cloth. Overall, it provides a smoother reconstruction that nevertheless preserves the relevant structures. On the other hand, the recovered surface seems to be slightly overflattened at some locations (e.g., on the nose or the back), which is due to the discrepancy between image and volume sampling, discussed previously. Note that the input photographs are of relatively high resolution.

6.4 Reconstruction of Complex and Fine-Scale Objects

While the above objects were topologically rather simple, Fig. 7 shows experimental results obtained on a considerably more complex geometry of a warrior figurine that has many small-scale structures, such as the hammer and the sword.⁴ Again, we observe that the proposed reconstruction exhibits a relatively high degree of smoothness while preserving all fine-scale geometric details. Note that the image sequence consists of only 24 images from camera positions arranged on a hemisphere around the object, which implies large baselines and thus restricted accuracy of the estimated photoconsistency.

Fig. 8 shows a comparison between the reconstructions obtained by applying the simple iterative projection method and the more accurate euclidean projection in the course of optimization (see Section 4). While the simple iterative approach produces artifacts in the area below the fur at the back (marked in red), the euclidean one overcomes these numerical difficulties and gives a more accurate reconstruction. Generally, the simple iterative procedure performs poorly in the presence of occluded concave areas. However,

4. The data to this reconstruction are courtesy of Yasutaka Furukawa—see <http://www.cs.washington.edu/homes/furukawa/research/mview/index.html> for the image data set and a reconstruction obtained with the method in [12].

in most practical scenarios, such cases do not occur. In fact, in our experiments, the “warrior” sequence was the only one where we could observe considerable differences between the results obtained by both techniques.

6.5 Reconstruction of Low-Textured Objects

The next experiment shown in Fig. 9 illustrates that the proposed method allows us to compute good quality reconstructions even for objects that exhibit rather little prominent texture. For such experiments where the stereo-matching provides almost no relevant information regarding good or bad matches, the proposed method essentially computes a euclidean minimal surface that exactly fulfills all silhouette constraints. The results show that even in the absence of reliable stereo information, we are able to obtain highly detailed reconstructions.

6.6 Robustness to Missing Silhouette Information

In many practical scenarios, perfect silhouette information is hard to obtain due to background clutter, image noise, etc. This raises the question about the practical relevance of the proposed approach. In contrast to most existing methods for silhouette-based reconstruction, our method can easily exploit partial silhouette information. While alternative techniques require given apparent contours, the presented approach exploits the pixel occupancy of the object projections. Obviously, we can impose silhouette constraints only for those pixels (and corresponding visual rays) for which we have a minimal confidence about being inside or outside the silhouette—for example, based on some color likelihood criterion. How will such limited silhouette information affect the final reconstruction?

In order to explore the behavior of the proposed method in the case of incomplete silhouette information, we assume that only a certain percentage $p \in [0, 1]$ of all constraints of type (14) are taken into account for reconstruction. Note that omitting constraints of type (13) will have less impact on the reconstruction due to the nature of the silhouette fusion process. In addition, we assume for objectivity that the subset of ignored constraints has been selected randomly, i.e., a certain proportion of it is randomly switched off.

Let $u_{full} : V \rightarrow \{0, 1\}$ denote the solution computed with complete silhouette information and $u_p : V \rightarrow \{0, 1\}$ the solution obtained using only a fraction $p \in [0, 1]$ of it. To quantify the decay in performance with fewer and fewer constraints, we define the reconstruction error with respect to the solution with complete information as the relative deviation ϵ_p :

$$\epsilon_p = \frac{\int_V |u_{full}(x) - u_p(x)| d^3x}{\int_V u_{full}(x) d^3x + \int_V u_p(x) d^3x}. \quad (20)$$

In particular, we have $\epsilon_p \in [0, 1]$, with $\epsilon = 0$ if and only if both reconstructions are identical and $\epsilon = 1$ if u_p is the empty set.

Fig. 10 shows the average reconstruction error computed as a function of the percentage of missing silhouette constraints. It shows that the reconstruction accuracy gradually degrades with decreasing value of p . Yet, for values $p \leq 0.96$, the reconstruction error is below 0.02. The figure

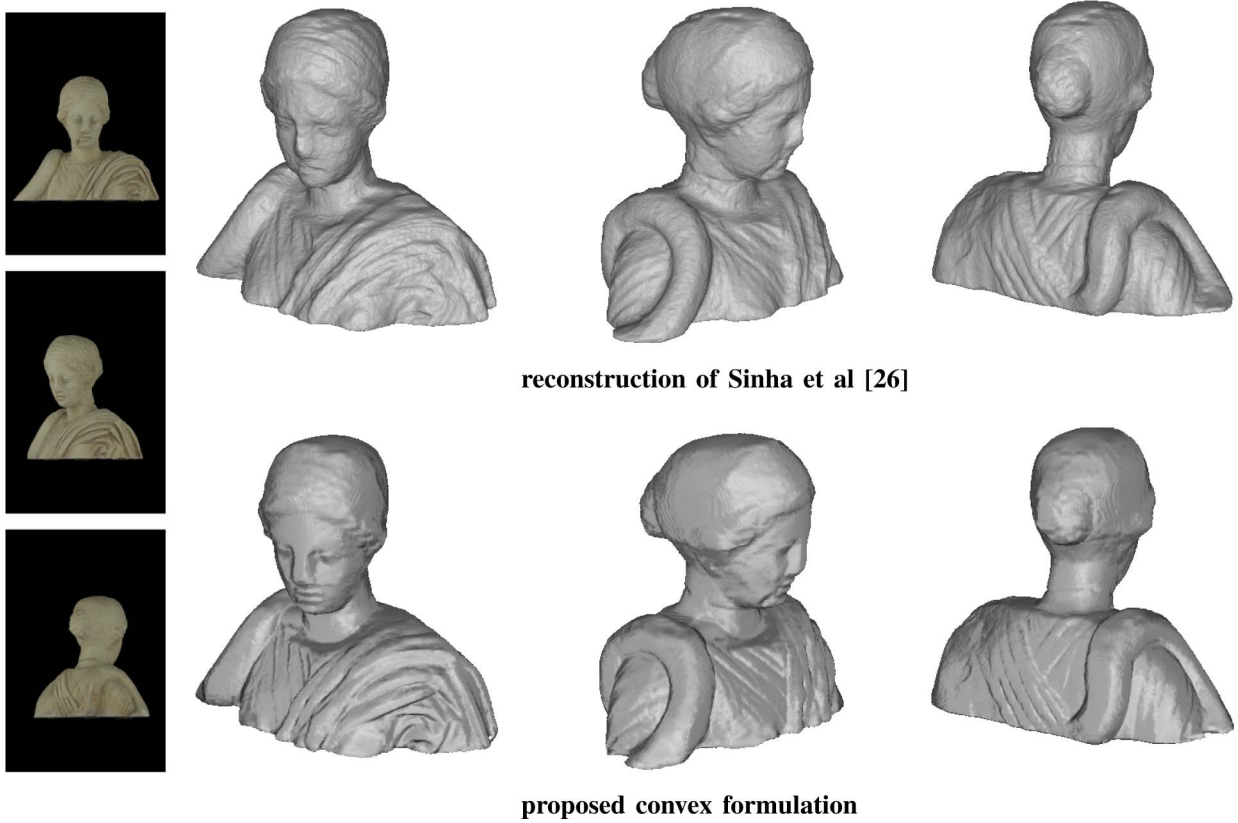


Fig. 6. Hygeia sequence. Three out of 36 input images of resolution $2,008 \times 3,040$ and multiple views of the reconstructed surface compared to the reconstruction of Sinha et al. [26]. Our result exhibits a higher grade of smoothness while recovering surface details more accurately (for example, the face and the creases of the cloth).

also depicts the reconstruction with only 4 percent of the silhouette constraints (i.e., $p = 0.96$). A direct comparison to Fig. 9 reveals that many of the fine-scale details are over-smoothed, while all relevant large-scale parts of the object are preserved. We can conclude that due to the utilized silhouette fusion scheme, the proposed approach enjoys considerable

resilience to missing silhouette information. This assures high practical applicability of the proposed method.

6.7 Reconstruction from a Hand-Held Camera

The calibration of the input images is a crucial ingredient of the reconstruction pipeline. While the above reconstructions

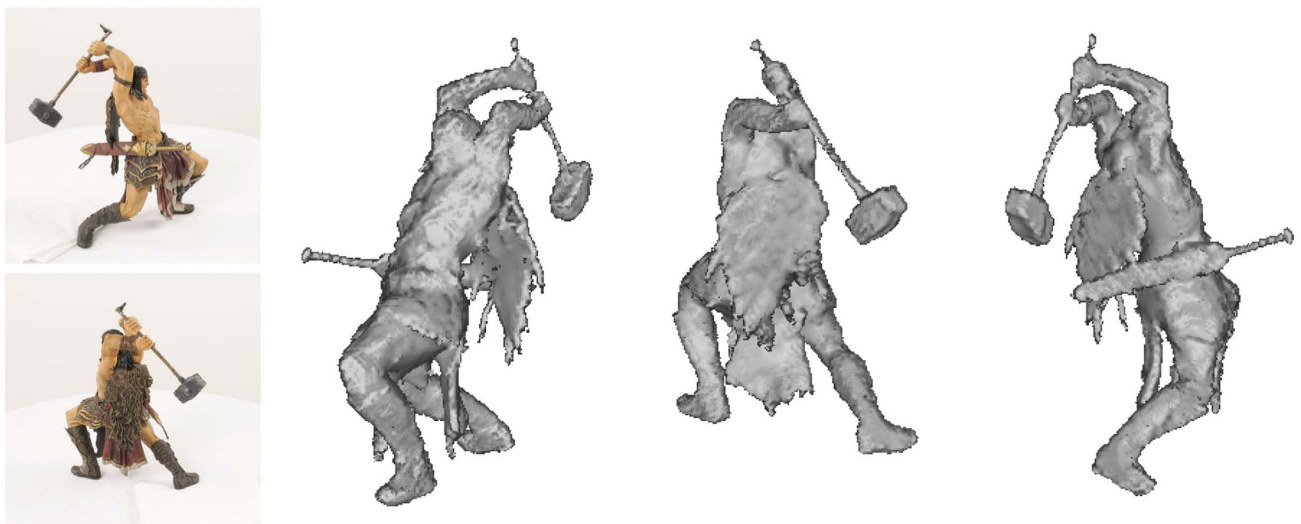


Fig. 7. Warrior sequence. Two out of 24 input images of resolution $1,600 \times 1,600$ and multiple views of the reconstructed surface. Note that thin structures (for example, the handle of the hammer) as well as concavities (for example, at the chest) are reconstructed accurately.

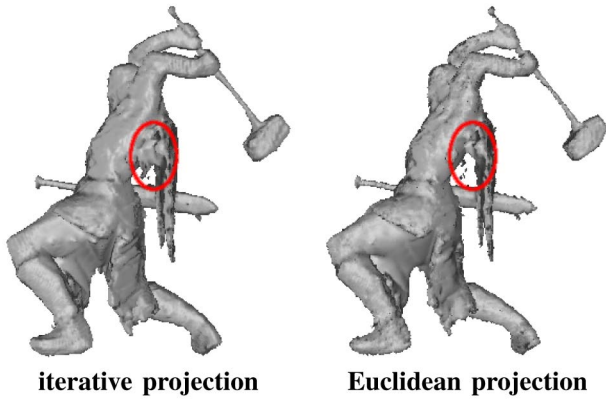


Fig. 8. Iterative versus euclidean projection. Imposing silhouette consistency by means of euclidean projection onto the feasible set of surfaces is particularly important to correctly handle occluded concave parts (see the marked area). While the iterative projection described in Section 4.2 often produces artifacts in such cases, the more accurate euclidean projection overcomes the numerical difficulties and gives better reconstructions.

were all conducted with imagery obtained either from the Internet or in controlled environments with precalibrated cameras, one may ask if the proposed method is sufficiently robust to also deal with outdoor sequences generated with a hand-held camera. In many real-world applications, one cannot precalibrate the camera possibly because the images were acquired by another person or because it is not possible to include calibration patterns in the scene.

In the following experiment, we demonstrate the robustness of the proposed approach to operate with input images lacking perfect color and camera calibration, even though the imaged object is well-textured and does not feature complex reflectance properties. We acquired 28 photographs of an ancient statue with a hand-held camera (see Fig. 11). We estimated calibration parameters directly from the images using the Bundler software [28], [1] (which performs feature

point matching, five-point algorithm, and RANSAC, followed by bundle adjustment refinement). Fig. 11 depicts multiple views of the recovered surface. Generally, the reconstruction exhibits a high grade of accuracy despite the severe intensity changes in the images and the imprecisions of the calibration. Note, in particular, the detailed recovery of the creases of the cloth. Note also the correct handling of the nontrivial topology of the reconstructed surface.

6.8 Energy Bound and Runtimes

As detailed in Section 3, the proposed approach does not give a global minimum of the underlying energy functional (5). Yet, the computed solution lies within an energy bound, dependent on the particular input data, around the global minimum (see Proposition 2). Here, we present the perimeter of this upper bound for all demonstrated experiments.

Apart from robustness, another crucial issue for a silhouette and stereo integration approach is the computational time needed. To this end, we used a GPU implementation of the proposed approach where the SOR optimization in a red-black strategy runs on the GPU. For comparison, we implemented the simple iterative projection method (see Section 4.2) as well as the more accurate euclidean projection method (see Section 4.3). Note, however, that due to their sequential nature, a GPU implementation is not possible. As a result, the backprojection dominated the overall computational time of the optimization.

Table 1 summarizes the energy bounds and runtimes for all presented experiments. In order to deliver insight into how tight the bounds are, we estimated the energy gap as the ratio of the energy at the computed solution (i.e., after thresholding) and the solution of the relaxed problem (i.e., before thresholding). The variations seem to correlate with the accuracy of the photoconsistency map and the geometry of the imaged object (e.g., the presence of concavities). The computational times were measured on a PC with 2.83 GHz and 8 GB of main memory, equipped with an NVIDIA Tesla

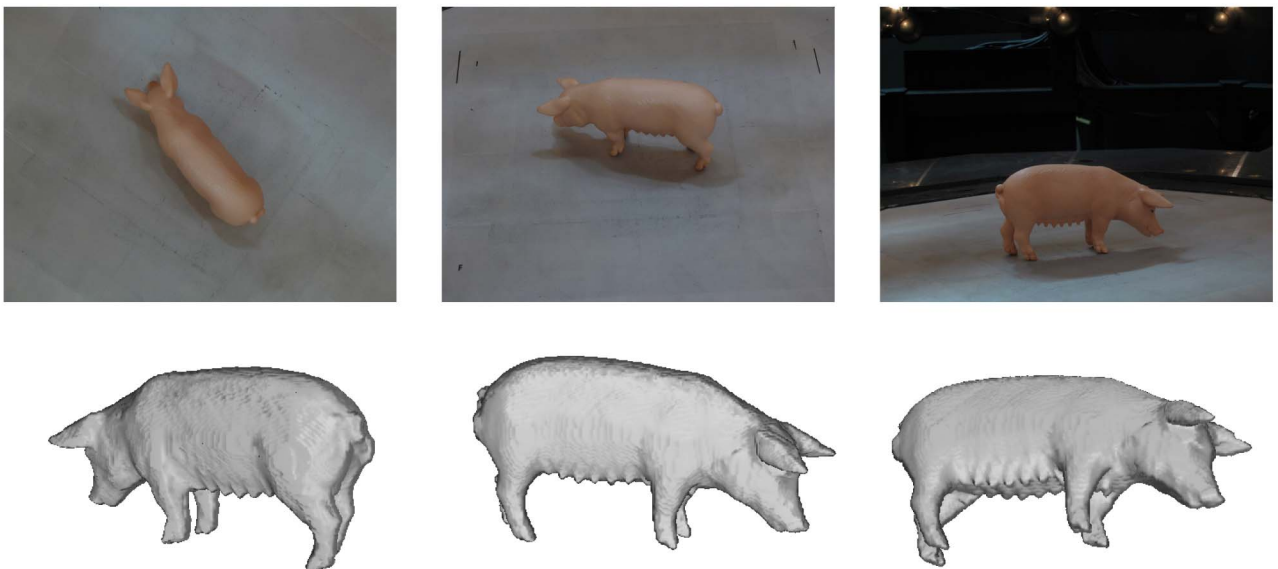


Fig. 9. Sow sequence. *First row*: Three out of 27 input images of resolution $1,024 \times 768$. *Second row*: Multiple views of the reconstructed surface. Despite a lack of prominent texture, the proposed method allows an accurate reconstruction even of fine-scale details of the object.

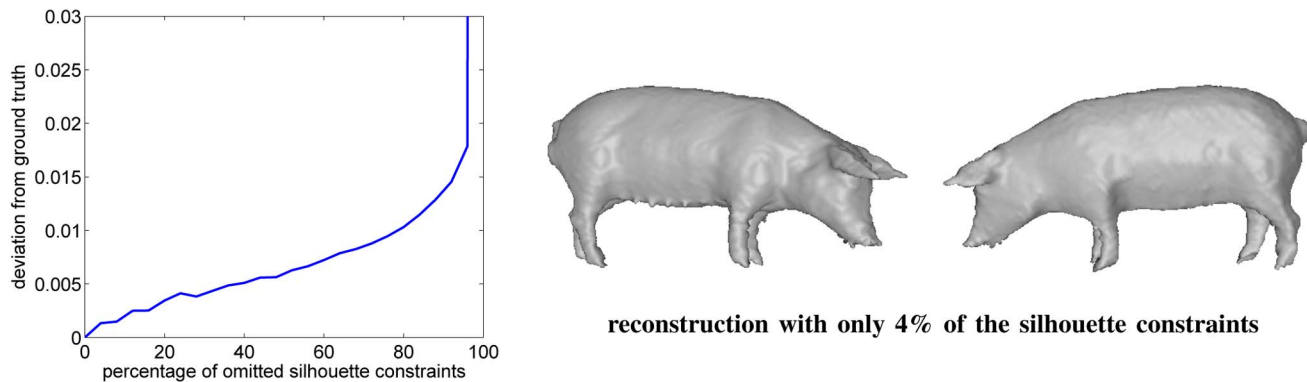


Fig. 10. Robustness to missing silhouette information. *Left*: Completeness of the reconstruction obtained with the proposed approach applied on the image sequence in Fig. 9 for the reduced number of silhouette constraints. The omitted constraints of type (14) were chosen randomly. As a ground truth served the reconstruction obtained with complete information (see Fig. 9), the deviation from it was estimated by evaluating (20). *Right*: Reconstruction obtained with only 4 percent of the constraints of type (14). The corresponding deviation value is about 0.018. Although the accuracy of fine-scale details degrades by reducing the number of silhouette constraints, the overall large-scale accuracy remains high, even when ignoring 96 percent of them.



Fig. 11. Reconstruction from the hand-held camera. Two out of 28 input images of resolution $2,048 \times 3,072$ and multiple views of the reconstructed surface. Despite severe illumination variations in the input images, the creases of the cloth are recovered in high detail.

TABLE 1
Energy Bounds and Runtimes for All Experiments

	# images	image resolution	energy before thresholding	energy after thresholding	energy gap	runtime (iterative proj)	runtime (Euclidean proj)
head	33	1024×768	86315.2	107146	1.24	14 s	51 s
hygeia	36	2008×3040	127877	165292	1.29	74 s	492 s
warrior	24	1600×1600	68160.8	109829	1.61	26 s	83 s
sow	27	1024×768	59661.2	70429.6	1.18	7 s	20 s
statue	28	2048×3072	112417	146437	1.3	125 s	280 s

The optimization runs partially on the GPU (SOR scheme) and partially on the CPU (constraint enforcement). For all experiments, the volumetric resolution was between 15 and 20 million voxels.

C1060 graphics card. Note that the optimization runs partially on the GPU (SOR scheme) and partially on the CPU (constraint enforcement). Note also that the runtime of the proposed approach strongly depends on the resolution of the input images since it gives the number of constraints to be considered. The runtime also depends on the particular choice of the projection method. However, the dependence on the volume resolution is less relevant since the evolution scheme runs on the GPU. For all experiments,

the volumetric resolution was within the range of 15-20 million voxels. The reported times do not include the photoconsistency estimation as the latter is very sensitive to the choice of methodology.

7 CONCLUSION

We proposed a novel framework for integrating silhouette and stereo information in 3D reconstruction from multiple

images. The key idea is to cast multiview stereovision as a convex variational problem and to impose exact silhouette constraints by restricting the domain of feasible functions. Relaxation of the resulting formulation leads to the minimization of a convex functional over the convex set of silhouette-consistent functions, which can be performed in a globally optimal manner using classical techniques. A solution of the original problem is obtained by projecting the computed minimizer onto the corresponding restricted domain. We presented an algorithm which allows us to efficiently compute the projection of the current solution onto the convex set of silhouette-consistent configurations. We also proved that the final reconstruction is within an energetic bound of the optimum.

In contrast to classical techniques for silhouette and stereo integration, the proposed approach of minimizing convex energies over convex sets leads to a more robust and tractable numerical scheme. The method makes no hard decisions about voxel occupancy and does not exhibit any bias near the visual hull boundary. It allows us to efficiently compute highly accurate silhouette-consistent reconstructions for challenging real-world problems (shiny metallic objects, low-texture objects, multiply connected objects with fine-scale structure, and outdoor reconstructions from a hand-held camera).

APPENDIX

Here, we will discuss some properties of the optimization procedure proposed in Section 5.2.

Proposition 5. *The linearized fixed-point iteration scheme in (18) can be interpreted as a specific quasi-Newton method.*

Proof. We will prove the claim in a constructive way. For simplicity, we will consider the case when the diffusivity terms are updated after each iteration and the implicit Gauss-Seidel step is replaced by an explicit Jacobi one. A generalization to the procedure in (18) is straightforward.

We start with some notations. Without loss of generality, we can assume $V = [0, 1]^3 \subset \mathbb{R}^3$. Let

$$G := \{ (i_1 h, i_2 h, i_3 h) \mid i_1, i_2, i_3 = 0, \dots, N \} \subset V,$$

with $h = \frac{1}{N}$ denoting a discretization of V , where $N \in \mathbb{N}$ is the sampling density along each dimension. Again, let $\vec{u} \in \mathbb{R}^d$, where $d = N^3$ summarizes the values of u on G . Furthermore, let $g(u) := \frac{\rho}{|\nabla u|}$ denote the nonlinear diffusivity term in (10).

First, we observe that if E denotes the functional in (8), its gradient in a continuous setting reads

$$\nabla E(u) = -\operatorname{div} \left(\rho \frac{\nabla u}{|\nabla u|} \right),$$

which, after discretization, leads to $\nabla E(\vec{u}) \in \mathbb{R}^d$. A quasi-Newton step is given by

$$\vec{u}^{t+1} = \vec{u}^t - \omega (B^t)^{-1} \nabla E(\vec{u}^t),$$

where $B^t \in \mathbb{R}^{d \times d}$ and $\omega \in \mathbb{R}$ is a time-step parameter. Now, we can focus on the components of $\nabla E(\vec{u}^t)$:

$$\begin{aligned} (\nabla E(\vec{u}^t))_{l,m,n} &= -\operatorname{div} \left(\rho_{l,m,n} \frac{\nabla u_{l,m,n}^t}{|\nabla u_{l,m,n}^t|} \right) \\ &= -\operatorname{div} (g_{l,m,n}^t \nabla u_{l,m,n}^t) \\ &= -\frac{\partial}{\partial x} (g_{l,m,n}^t u_x^t) - \frac{\partial}{\partial y} (g_{l,m,n}^t u_y^t) - \frac{\partial}{\partial z} (g_{l,m,n}^t u_z^t), \end{aligned}$$

where $l, m, n \in \{0, \dots, N\}$. Note that the diffusivity g^t depends on u^t , which is omitted in the above expressions for simplicity. For an interior point on G , we can discretize the arising partial derivatives by evaluating the corresponding terms between the respective neighboring grid points, i.e.,

$$\frac{\partial}{\partial x} (g_{l,m,n}^t u_x^t) = (g_{l+1/2,m,n}^t u_x^t) - (g_{l-1/2,m,n}^t u_x^t),$$

and analogously for the other terms. Next, we can discretize the partial derivatives of u^t by means of forward or backward differences, respectively. For example, the expressions for u_x^t read

$$\begin{aligned} (u_x^t)_{l+1/2,m,n} &= u_{l+1,m,n}^t - u_{l,m,n}^t, \\ (u_x^t)_{l-1/2,m,n} &= u_{l,m,n}^t - u_{l-1,m,n}^t. \end{aligned}$$

Finally, we specify $B^t = (b_{ik}^t)_{i,k=1,\dots,d}$ as

$$b_{ik}^t = \begin{cases} \sum_{j \in \mathcal{N}(i)} g_{i \sim j}^t, & \text{if } i = k, \\ 0, & \text{otherwise,} \end{cases}$$

where $g_{i \sim j}^t$ denotes again the diffusivity between voxels i and j , i.e., $g_{i \sim j}^t := (g_i^t + g_j^t)/2$. Summarizing all computations and reverting to the notations in Section 5.2 leads to the following evolution scheme:

$$u_i^{t+1} = (1 - \omega) u_i^t + \omega \frac{\sum_{j \in \mathcal{N}(i)} g_{i \sim j}^t u_j^t}{\sum_{j \in \mathcal{N}(i)} g_{i \sim j}^t}.$$

□

Remarks.

1. All matrices B^t and $(B^t)^{-1}$ arising in the optimization are diagonal matrices with positive diagonal elements. Hence, they are symmetric and positive-definite, which guarantees that each update direction is a descent direction.
2. A closer look at the evolution scheme reveals that the matrices $(B^t)^{-1}$ are designed in a way that some small gradient components are encouraged while too large components are damped. For example, in areas where u is approximately constant, the corresponding gradient components will be small in magnitude. Yet, the respective matrix value is expected to be large due to $1/|\nabla u| \gg 1$. Analogously, at locations of high variation of u , the diffusivity terms $g = \rho/|\nabla u|$ may become smaller than 1 and damp the corresponding gradient components. This results in favorable properties of the proposed fixed-point iteration scheme, which is both fast and numerically stable.

ACKNOWLEDGMENTS

The authors thank Reinhard Klein and Dirk Koch for support with indoor image acquisition. They thank Svetlana Matiouk for helping with outdoor image acquisition and camera calibration, Carlos Hernandez and Yasutaka Furukawa for providing the data sets in Figs. 6 and 7, Sudipta Sinha for sharing his results for Fig. 6, and Antonin Chambolle and Thomas Pock for fruitful discussions on projection methods and convex optimization. This research was supported by the German Research Foundation, grant # CR250/1-2.

REFERENCES

- [1] <http://phototour.cs.washington.edu/bundler/>, 2010.
- [2] B. Baumgart, "Geometric Modeling for Computer Vision," PhD thesis, Dept. of Computer Science, Stanford Univ., 1974.
- [3] Y. Boykov and V. Lempitsky, "From Photohulls to Photoflux Optimization," *Proc. British Machine Vision Conf.*, vol. 3, pp. 1149-1158, 2006.
- [4] J.P. Boyle and R.L. Dykstra, "A Method for Finding Projections onto the Intersection of Convex Sets in Hilbert Spaces," *Lecture Notes in Statistics*, pp. 28-47, Springer, 1986.
- [5] R. Cipolla and P. Giblin, *Visual Motion of Curves and Surfaces*. Cambridge Univ. Press, 2000.
- [6] G. Cross and A. Zisserman, "Surface Reconstruction from Multiple Views Using Apparent Contours and Surface Texture," *Confluence of Computer Vision and Computer Graphics*, pp. 25-47, Kluwer Academic Publishers, 2000.
- [7] Y. Duan, L. Yang, H. Qin, and D. Samaras, "Shape Reconstruction from 3D and 2D Data Using PDE-Based Deformable Surfaces," *Proc. European Conf. Computer Vision*, pp. 238-251, 2004.
- [8] R.L. Dykstra, "An Algorithm for Restricted Least-Squares Regression," *J. Am. Statistical Assoc.*, vol. 78, pp. 837-842, 1983.
- [9] C.H. Esteban and F. Schmitt, "Silhouette and Stereo Fusion for 3D Object Modeling," *Computer Vision and Image Understanding*, vol. 96, no. 3, pp. 367-392, 2004.
- [10] O. Faugeras and R. Keriven, "Variational Principles, Surface Evolution, PDE's, Level Set Methods, and the Stereo Problem," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 336-344, Mar. 1998.
- [11] J.S. Franco and E. Boyer, "Exact Polyhedral Visual Hulls," *Proc. 14th British Machine Vision Conf.*, pp. 329-338, Sept. 2003.
- [12] Y. Furukawa and J. Ponce, "Carved Visual Hulls for Image-Based Modeling," *Proc. European Conf. Computer Vision*, pp. 564-577, 2006.
- [13] P. Gargallo, E. Prados, and P. Sturm, "Minimizing the Reprojection Error in Surface Reconstruction from Images," *Proc. IEEE Int'l Conf. Computer Vision*, 2007.
- [14] C. Hernández, G. Vogiatzis, and R. Cipolla, "Probabilistic Visibility for Multi-View Stereo," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2007.
- [15] J. Isidoro and S. Sclaroff, "Stochastic Refinement of the Visual Hull to Satisfy Photometric and Silhouette Consistency Constraints," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1335-1342, 2003.
- [16] K. Kolev, T. Brox, and D. Cremers, "Robust Variational Segmentation of 3D Objects from Multiple Views," *Pattern Recognition*, K. Franke, K.R. Müller, B. Nickolay, and R. Schäfer, eds., pp. 688-697, Springer-Verlag, Sept. 2006.
- [17] K. Kolev and D. Cremers, "Integration of Multiview Stereo and Silhouettes via Convex Functionals on Convex Domains," *Proc. European Conf. Computer Vision*, Oct. 2008.
- [18] K. Kolev, M. Klodt, T. Brox, and D. Cremers, "Propagated Photoconsistency and Convexity in Variational Multiview 3D Reconstruction," *Proc. Workshop Photometric Analysis for Computer Vision*, Oct. 2007.
- [19] A. Laurentini, "The Visual Hull Concept for Visual-Based Image Understanding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150-162, Feb. 1994.
- [20] V. Lempitsky, Y. Boykov, and D. Ivanov, "Oriented Visibility for Multiview Reconstruction," *Proc. European Conf. Computer Vision*, pp. 226-238, 2006.
- [21] W.N. Martin and J.K. Aggarwal, "Volumetric Descriptions of Objects from Multiple Views," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 150-158, Mar. 1983.
- [22] Y. Matsumoto, K. Fujimura, and T. Kitamura, "Shape-from-Silhouette/Stereo and Its Application to 3D Digitizer," *Discrete Geometry for Computing Imagery*, pp. 177-190, Springer, 1999.
- [23] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "An Algorithm for Minimizing the Piecewise Smooth Mumford-Shah Functional," *Proc. IEEE Int'l Conf. Computer Vision*, 2009.
- [24] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *Proc. IEEE CS Int'l Conf. Computer Vision and Pattern Recognition*, pp. 519-528, 2006.
- [25] S. Seitz and C. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 1067-1073, June 1997.
- [26] S. Sinha, P. Mordohai, and M. Pollefeys, "Multiview Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh," *Proc. IEEE Int'l Conf. Computer Vision*, Oct. 2007.
- [27] S. Sinha and M. Pollefeys, "Multi-View Reconstruction Using Photo-Consistency and Exact Silhouette Constraints: A Maximum-Flow Formulation," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 349-356, 2005.
- [28] N. Snavely, S.M. Seitz, and R. Szeliski, "Modeling the World from Internet Photo Collections," *Int'l. J. Computer Vision*, vol. 80, no. 2, pp. 189-210, 2008.
- [29] D. Snow, P. Viola, and R. Zabih, "Exact Voxel Occupancy with Graph Cuts," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 345-353, 2000.
- [30] S. Tran and L. Davis, "3D Surface Reconstruction Using Graph Cuts with Surface Constraints," *Proc. European Conf. Computer Vision*, pp. 219-231, 2006.
- [31] C.R. Vogel and M.E. Oman, "Iterative Methods for Total Variation Denoising," *SIAM J. Scientific Computing*, vol. 17, pp. 227-238, 1996.
- [32] G. Vogiatzis, P. Torr, and R. Cipolla, "Multi-View Stereo via Volumetric Graph-Cuts," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 391-399, 2005.
- [33] A. Yezzi and S. Soatto, "Stereoscopic Segmentation," *Int'l J. Computer Vision*, vol. 53, no. 1, pp. 31-43, 2003.



Daniel Cremers received the MS (diploma) degree in theoretical physics from the University of Heidelberg in 1997, and the PhD degree in computer science from the University of Mannheim, Germany, in 2002. Subsequently, he spent two years as a postdoctoral researcher at the University of California, Los Angeles, and one year as a permanent researcher at Siemens Corporate Research, Princeton, New Jersey. From 2005 to 2009, he headed the Computer Vision Group at the University of Bonn, Germany. Since 2009, he has been a full professor at TU München. He has received several awards, in particular, the Best Paper of the Year Award 2003 by the Pattern Recognition Society, the Olympus Award 2004, and the 2005 UCLA Chancellor's Award.



Kalin Kolev received the BS (vordiploma) in 2002 and the MS (diploma) degree in computer science in 2005 from the University of Bonn, Germany. Since January 2006, he has been working toward the PhD degree in the Computer Vision Group at the University of Bonn (until November 2009) and TU München (since December 2009). His main research interests include multiview 3D reconstruction, image-based modeling, and continuous optimization.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.