Silhouette-Based Variational Methods for Single View Reconstruction

Eno Töppe[1,2], Martin R. Oswald[1], Daniel Cremers[1], and Carsten Rother[2]

[1] Technische Universität München, Germany
[2] Mircosoft Research, Cambridge, UK

Abstract. We explore the 3D reconstruction of objects from a single view within an interactive framework by using silhouette information. In order to deal with the highly ill-posed nature of the problem we propose two different reconstruction priors: a shape and a volume prior and cast them into a variational problem formulation. For both priors we show that the corresponding relaxed optimization problem is convex. This leads to unique solutions which are independent of initialization and which are either globally optimal (shape prior) or can be shown to lie within bounds from the optimal solution (volume prior). We analyze properties of the proposed priors with regard to the reconstruction results as well as their impact on the minimization problem. By employing an implicit volumetric representation our reconstructions enjoy complete topological freedom. Being parameter-based, our interactive reconstruction tool allows for intuitive and easy to use modeling of the reconstruction result.

Keywords: Single View Reconstruction, Image-Based Modeling, Convex Optimization.

1   Introduction

1.1   Single View Reconstruction

The general problem of 3D reconstruction has been considered in a plethora of works in computer vision - at least in the case of given multiple views and in stereo vision. With the help of multi-view concepts like point correspondences and photo-consistency it has been shown that high-quality reconstructions can be inferred from a set of photographs of a single object. However, there are relatively few works on single view reconstruction, although its underlying problem may be considered one of the most fundamental in vision. This may to a great extent be due to the high ill-posedness of the corresponding mathematical problem, but is nevertheless astonishing, as humans excel in solving the task in every-day life.

The main difficulty in inferring 3D geometry from a single image lies in the fact that it is inherently ill-posed. The process of image formation is not invertible and it is impossible to retrieve exact depth values from a single image. Thus, we have to make use of a strong prior. Such priors can either be obtained by

statistical learning of shape or by restraining the solution space, e.g. by making the assumption of smoothness and compactness. Additionally to a prior, user input can be incorporated into the reconstruction process which can be realized as a modeling tool.

While the growing amount of image data on the Internet increases the availability of multiple views for certain scenes, this does only apply to few places of strong public interest, like e.g. touristic hot spots. Single view reconstruction becomes particularly important in situations where a rough estimate of object geometry is desired rather than an exact reconstruction. This is the case when generating an alternate view of a single photograph or changing the illumination of the depicted scene.

In this work we follow the idea of modeling an object from a single view gradually by user input, but with the ultimate goal of keeping the process simple for the user. Instead of an involved modeling stage that amounts to the specification of absolute vertex positions and normal directions, we rather rely on user provided global and local constraints that, together with a strong prior, lead to a reconstruction estimate.

This work recapitulates two different priors for single view reconstruction which were proposed in papers [1] and [2]. One is based on a shape prior formulation, the other one amounts to a global constraint on the volume of the reconstruction. We evaluate both approaches and compare them to each other.

1.2   Issues and Related Work

Existing work on single view reconstruction and on interactive 3D modeling can be roughly classified into the categories *planar* versus *curved* and *implicit* versus *parametric approaches*.

Many approaches such as that of Horry et al. [3] aim to reconstruct *planar* surfaces by evaluating user defined vanishing points and lines. This has been extended by Liebowitz [4] and Criminisi [5]. This process has been completely automated by Hoiem et al. [6], yielding appealing results on a limited number of input images. Sturm et al. [7] make use of user-specified constraints such as coplanarity, parallelism and perpendicularity in order to reconstruct piecewise planar surfaces.

An early work for the reconstruction of *curved* objects is Terzopoulos et al. [8] in which symmetry seeking models are reconstructed from a user defined silhouette and symmetry axis using snakes. However, this approach is restricted to the class of tube-like shapes. Moreover, reconstructions are merely *locally* optimal. The work of Zhang et al. [9] addresses this problem and proposes a model which globally optimizes a smoothness criterion. However, it concentrates on estimating height fields rather than reconstructing real 3D representations. Moreover, it requires a huge amount of user interaction in order to obtain appealing reconstructions. Also related to the field are easy-to-use tools like *Teddy* [10] and *FiberMesh* [11] that have pioneered sketch based modeling but are not image-based.

All of the cited works are using *explicit* surface representation – while surface manipulation is often straightforward and a variety of cues are easily integrated leading to respective forces or constraints on the surface, there are two major limitations: Firstly numerical solutions are generally not independent of the choice of parameterization. And secondly, parametric representations are not easily extended to objects of varying topology. While Prasad et al. [12] were able to extend their approach to surfaces with one or two holes, the generalization to objects of arbitrary topology is by no means straightforward and quite involved for the user. Similarly, topology-changing interaction in the FiberMesh system requires a complex remeshing of the modeled object leading to computationally challenging numerical optimization schemes. For the given reasons in this work we pursue an implicit representation of the reconstructed object. Joshi et al. [13] also suggest a silhouette-based surface inflation method and minimize a similar energy as [9] or [12] in order to obtain a smooth surface. However, like Zhang et al. [9], Joshi et al. aim to reconstruct depth maps rather than full 3D objects.

Another problem of all existing works is the fact that they revert to inflation heuristics in order to avoid surface collapsing. These techniques boil down to fixing absolute depth values which undesirably restrict the solution space. We show that a prior on the volume of the reconstruction solves this problem. A precursor to volume constraints are the volume inflation terms pioneered for deformable models by Cohen and Cohen [14]. However, no constant volume constraints were considered and no implicit representations were used.

1.3   Contribution

In this paper, we focus on the reconstruction of curved objects of arbitrary topology with a minimum of user input in an interactive and intuitive framework. We propose a convex variational method which generates a 3D object in a matter of seconds using silhouette information only.

To this end, we revert to an implicit representation of the surface given by the indicator function of its interior (sometimes referred to as voxel-occupancy). In this representation, the weighted minimal surface problem is a convex functional and relaxation of the binary function leads to an overall convex problem.

Two approaches are presented to overcome the ambiguity in the reconstruction process: In the first one we formulate a shape prior which determines the basic shape and at the same time inflates the reconstruction geometry. In the second approach we introduce a constraint on the volume of the reconstruction. We discuss advantages and shortcomings of both approaches. In both cases we detail how to solve the resulting optimization problem by means of relaxation. This leads to a solution to the unrelaxed problem that is globally optimal in the case of a shape prior and that we show to be within in a bound of the optimum in the case of a volume prior.

User Strokes for Segmentation	Silhouette	First Estimate	Final Result

Fig. 1. The basic workflow of the single view reconstruction process: The user marks the input image with scribbles (*left*) from which a silhouette is generated by segmentation (*second from left*). A first reconstruction estimate is generated automatically from the silhouette (*third from left*). The user can then iteratively adapt the model in an interactive manner (*right*).

2   Reconstruction Workflow

A good silhouette is the main prerequisite for a reasonable reconstruction result with the algorithms proposed in Sections 4 and 5. The number of holes in the segmentation of the target object determines the topology of the reconstructed surface. Notably, the proposed reconstruction methods can also cope with disconnected regions of the object silhouette.

The segmentation is obtained by utilizing an interactive graph cuts scheme similar to the ones described by [15] and [16]. The algorithm calculates two distinct regions based on respective color histograms which are defined by representational pen strokes given by the user (see Fig. 1).

From the input image and silhouette a first reconstruction is generated automatically, which - depending on the complexity and the class of the object - can already be satisfactory. However, for some object classes and due to the general over-smoothing of the resulting mesh (see Section 3), the user can subsequently adapt the reconstruction by specifying intuitive and simple global and local constraints. These editing tools are completely parameter-based. The editing stage can be reiterated by the user until the desired result is obtained.

3   Implicit Variational Surfaces

Assume we are given the silhouette of an object in an image as returned by an interactive segmentation tool. The goal is then to obtain a smooth $3D$ model of the object which is consistent with the silhouette. How should we select the correct $3D$ model among the infinitely many that match the silhouette? Clearly, we need to impose additional information, at the same time we want to keep this information at a minimum since user interaction is always tedious and slow.

Formally, we are given an image plane Ω which contains the input image and lies in $\mathbb{R}^3$. As part of the image we also have an object silhouette $\Sigma \subset \Omega$. Now, we are seeking to compute reconstructions as minimal weighted surfaces $S \subset \mathbb{R}^3$ that are compliant with the object silhouette Σ:

$$min \int_S g(s)ds \tag{1}$$

$$\text{subject to} \qquad \pi(S) = \Sigma$$

where $\pi : \mathbb{R}^3 \rightarrow \Omega$ is the orthographic projection onto the image plane Ω, $g : \mathbb{R}^3 \rightarrow \mathbb{R}^+$ is a smoothness weighting function and $s \in S$ is an element of the surface S.

We now introduce an implicit representation by replacing the surface S with its implicit binary indicator function $u \in BV(\mathbb{R}^3; \{0,1\})$ representing the voxel occupancy (0 =exterior, 1 =interior), where BV denotes the functions of bounded variation [17]. The desired minimal weighted surface area is then given by minimizing the total variation (TV) over a suitable set U_Σ of feasible functions u:

$$\min_{u \in U_\Sigma} \int g(x)|\nabla u(x)|d^3x \tag{2}$$

where ∇u denotes the derivative in the distributional sense. Eq. (2) favors smooth solutions. However, smoothness is locally affected by the function $g(x) :$ $\mathbb{R}^3 \rightarrow \mathbb{R}^+$ which will be used later for modeling. Without any modeling g is the identity mapping by default, i.e. $g(x) \equiv 1$.

How does the set U_Σ of feasible functions look like? For simplicity, we assume the silhouette to be enclosed by the surface. Then all surface functions that are consistent with the silhouette Σ must be in the set

$$U_\Sigma = \left\{ u \in BV(\mathbb{R}^3; \{0,1\}) \,\Big|\, u(x) = \begin{cases} 0, & \pi(x) \notin \Sigma \\ 1, & x \in \Sigma \\ \text{arbitrary,} & \text{otherwise} \end{cases} \right\} \tag{3}$$

Obviously, solving problem (1) / (2) results in the silhouette itself. Therefore we need further assumptions in order to rule out trivial solutions. In the subsequent sections we propose two different approaches to the problem.

Using the Weighting Function for Modeling

The weight $g(x)$ of the TV-norm in Eq. (2) can be used to locally control the smoothness of the reconstruction: With a low value $0 \leq g(x) < 1$, the smoothness condition on the surface is locally relaxed, allowing for creases and sharp edges to form. Conversely, higher values for $g(x) > 1$ locally enforces surface smoothness. For controlling the weighting function we employ a user scribble interface. The parameter associated to each scribble marks the local smoothness $g(x)$ within the respective scribble area and is propagated through the volume along projection direction.

This approach of parametric local smoothness adaptation can be applied in the case of a data term (Section 4) as well as in case of a constant volume constraint (Section 5).

4   Inflation via Shape Prior

By introducing a data term, we realize two objectives: volume inflation and determination of the basic reconstructed shape. Since there is no inherent data term in the single view setting we have to define one heuristically. We choose a term of the following form:

$$\int u(x)\ \phi(x)d^3x \tag{4}$$

$\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ can be adopted to achieve the desired object shape and may also be adopted by user-interaction later on. Adding this term to the energy in Equation (2) amounts to the following problem:

$$\min_{u \in U_\Sigma} \int u(x)\ \phi(x)d^3x + \lambda \int g(x)|\nabla u(x)|d^3x \tag{5}$$

where λ is a weighting parameter that determines the relative smoothness of the solution.

In order to fix a definition for ϕ we make the simple assumption that the thickness of the observed object increases as we move inward from its silhouette. For any point $p \in V$ let

$$dist(p, \partial S) = \min_{s \in \partial S} \|p - s\|\ \ , \tag{6}$$

denote its distance to the silhouette contour $\partial S \subset \Omega$. Then we set:

$$\phi(x) = \begin{cases} -1 & \text{if } dist(x, \Omega) \leq h(\pi(x)) \\ +1 & \text{otherwise}\ , \end{cases} \tag{7}$$

where the height map $h : \Omega \mapsto \mathbb{R}$ depends on the distance of the projected 3D point to the silhouette according to the function

$$h(p) = \min \left\{ \mu_{\text{cutoff}}, \mu_{\text{offset}} + \mu_{\text{factor}} * dist(p, \partial S)^k \right\} \tag{8}$$

with four parameters $k, \mu_{\text{offset}}, \mu_{\text{factor}}, \mu_{\text{cutoff}} \in \mathbb{R}^+$ affecting the shape of the function ϕ. How the user can employ these parameters to modify the computed 3D shape will be discussed in the following paragraph.

Note that this choice of ϕ implies symmetry of the resulting model with respect to the image plane. Since the backside of the object is unobservable, it will be reconstructed properly for plane-symmetric objects.

Data Term Parameters. By altering the parameters $\mu_{\text{offset}}, \mu_{\text{factor}}, \mu_{\text{cutoff}}$ and the exponent k of the height map function (8), users can intuitively change the data term (4) and thus the overall shape of the reconstruction. Note that the impact of these parameters is attenuated with increasing importance of the smoothness term. The effects of the *offset*, *factor* and *cutoff* parameters on the height map are shown in Fig. 2 and are quite intuitive to grasp. The exponent k of the distance function in (8) mainly influences the objects curvature in the proximity of the silhouette contour. This can be observed in Fig. 2 showing an evolution from a cone to a cylinder just by decreasing k.
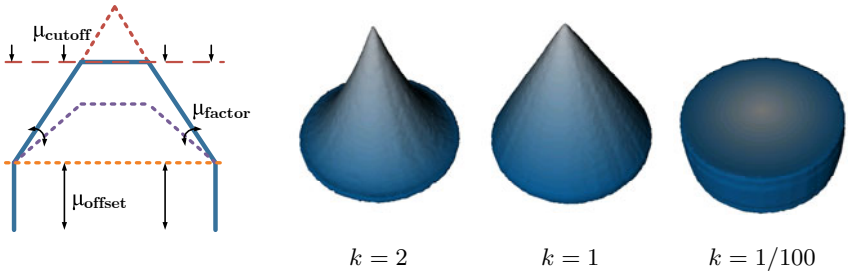
Fig. 2. Effect of $\mu_{\text{offset}}, \mu_{\text{factor}}, \mu_{\text{cutoff}}$ (left) and various values of parameter k and resulting (scaled) height map plots for a circular silhouette

Altering the Data Term Locally. Due to the incorporation of a distance transform in the data term, the reconstruction will always become flat at the silhouette border. However this is not always desired like for instance for the bottom and top of the vase in Fig. 3.

A simple remedy to this problem is to ignore parts of the contour during the calculation of the distance function. The user indicates the sections of the silhouette contour he wants to have ignored. To keep user interaction simple, we approximate the object contour by a polygon which is laid over the input image. By clicking on the edge, the user indicates to ignore the corresponding contour pixels during distance map calculation (see Fig. 3 top right).

4.1   Optimization via Convex Relaxation

To minimize energy (2) plus the data term (4) we follow the framework developed in [18]. To this end, we relax the binary problem, looking for functions $u : V \to [0, 1]$ instead.

We can globally minimize the resulting convex functional by solving the corresponding Euler-Lagrange equation

$$0 = \phi - \lambda \operatorname{div}\left(g\frac{\nabla u}{|\nabla u|}\right) \ ,\tag{9}$$

using a fixed-point iteration in combination with Successive Over-Relaxation (SOR). A global optimum of the original binary labeling problem is then obtained by simple thresholding of the solution of the relaxed problem – see [18] for details.

In [19] it was shown that such relaxation techniques have several advantages over graph cut methods. In this work, the two main advantages are the lack of metrication errors and the parallelization potential. These two aspects allow to compute smooth single view reconstructions with no grid bias within a few seconds using standard graphics hardware.

Fig. 3. Top row: height maps and corresponding reconstructions with and without marked sharp contour edges. Bottom row: input image with marked contour edges (blue) and line strokes (red) for local discontinuities which are shown right.
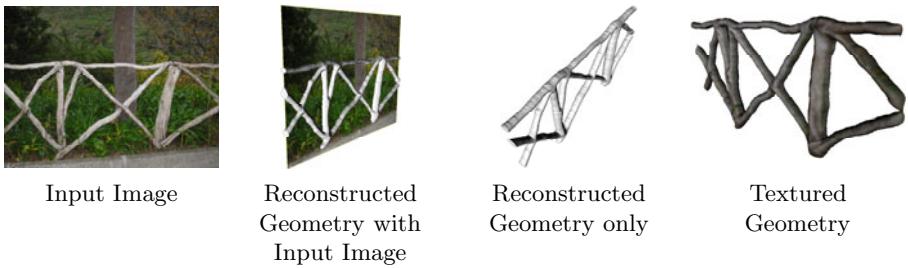

| Input Image | Reconstructed Geometry with Input Image | Reconstructed Geometry only | Textured Geometry |

Fig. 4. An example with intricate topology. Due to the implicit representation of the reconstruction surface, the algorithm of Sections 4 and 5 can handle any genus

4.2   Experiments

In the following we apply our reconstruction method with data term to several input images. We show different aspects of the reconstruction process for typical classes of target objects and mention limitations of the approach.

The experimental results are shown in Figures 4, 5, 6 and 7. Default values for the data term parameters (8) are $k = 1$, $\mu_{\text{offset}} = 0$, $\mu_{\text{factor}} = 1$, $\mu_{\text{cutoff}} = \infty$. Each row depicts several views of a single object reconstruction starting with the input image.
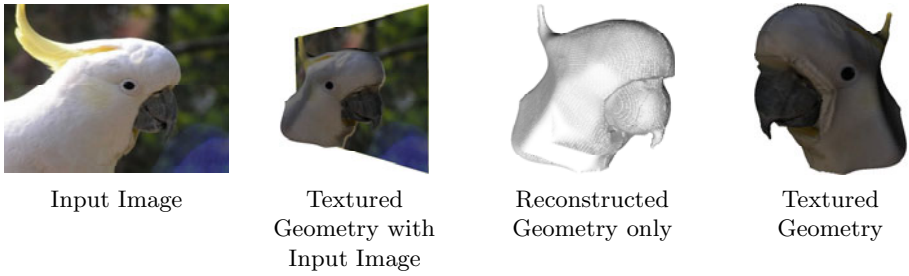
| Input Image | Textured Geometry with Input Image | Reconstructed Geometry only | Textured Geometry |

Fig. 5. For the cockatoo very little additional user input was necessary. The smoothness was reduced locally by a single user scribble (see Section 3).
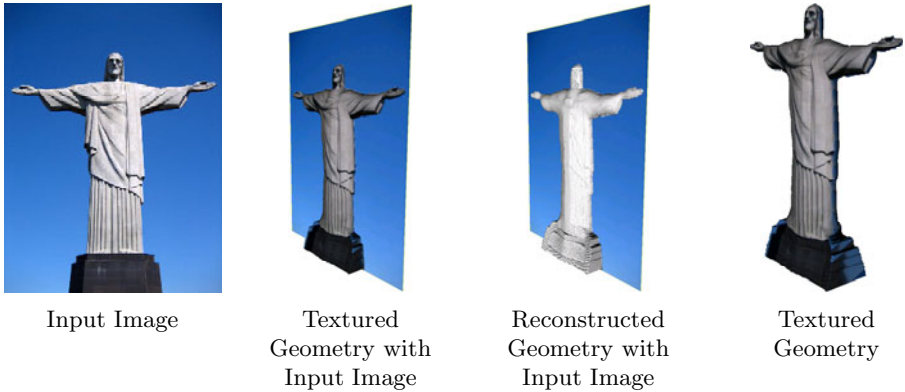

| Input Image | Textured Geometry with Input Image | Reconstructed Geometry with Input Image | Textured Geometry |

Fig. 6. The Cristo statue composes of smooth and non-smooth parts. The socket part was marked as non-smooth by a user scribble adapting the weight of the minimal surface locally.

The fence in Fig. 4 is an example of the complex topology the algorithm can handle. The reconstruction was automatically generated by the method right after the segmentation stage, i.e. without changing the surface smoothness.

Figures 5, 6 and 7 demonstrate the potential of user editing as described in Sections 3 and 4. The reconstructions were edited by adapting the local smoothness and locally editing the data term. It can be seen, that elaborate modeling effects can be achieved by these simple operations. Especially for the cockatoo a single curve suffices in order to add the characteristic indentation to the beak. No expert knowledge is necessary. For the socket of the Cristo statue, creases help to attain sharp edges, while keeping the rest of the statue smooth. It should be stressed, that no other post-processing operations were used.

The experiments in Figure 7 stand for a more complex series of target objects. A closer look reveals that the algorithm clearly attains its limit. The structure of the opera building (third row) as well as the elaborate geometry of the bike and its drivers cannot be correctly reconstructed with the proposed method due to a lack of information and more sophisticated tools. Yet the results are appealing and could be spiced up with the given tools.
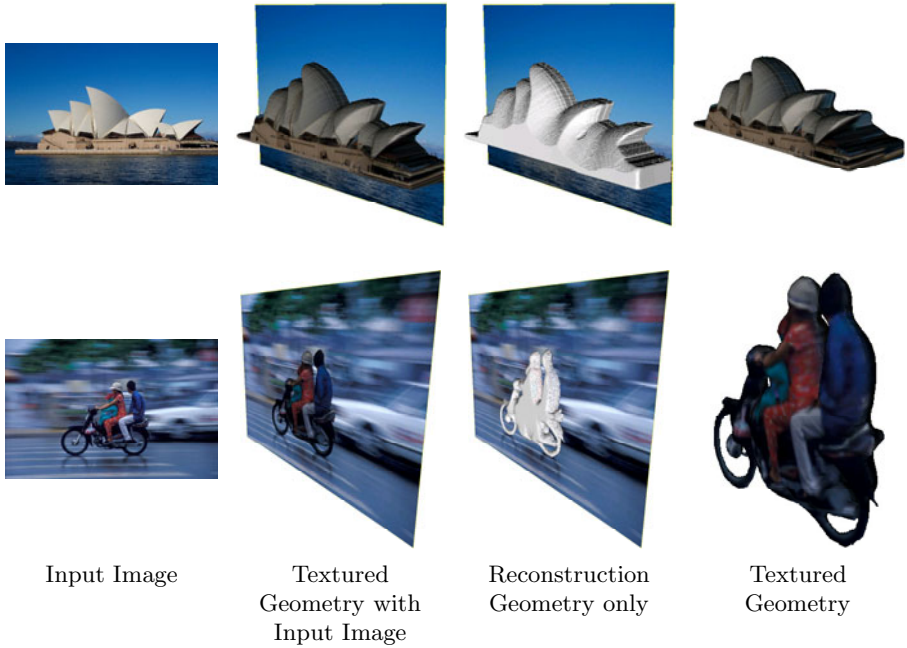
| Input Image | Textured Geometry with Input Image | Reconstruction Geometry only | Textured Geometry |

Fig. 7. Reconstruction examples where the algorithm attains its limit. Nevertheless the results are pleasing and could be used for tasks like new view synthesis.

5 Inflation via Volume Prior

Adding a data term to the variational problem (2) delivers reasonable results for single view reconstruction, as shown in the last section. However, we have also seen that a data term imposes a strong bias on the shape. Ideally we would like to have a non-heuristic inflation approach that does not restrict the shape variety while at the same time exhibits the same natural compactness as seen in the experiments above.

As an alternative inflation strategy we propose to use a constraint on the size of the volume enclosed by the minimal surface. We formulate this as a hard constraint by further constraining the feasible set of problem (2):

$$\min_{u \in U_\Sigma \cap U_V} E(u) \qquad \text{where} \quad E(u) = \int g(x)|\nabla u(x)|d^3x \qquad (10)$$

$$\text{and} \quad U_V = \left\{ u \in BV(\mathbb{R}^3; \{0,1\}) \,\middle|\, \int u(x)d^3x = V_t \right\} \qquad (11)$$

where U_V denominates all reconstructions with bounded variation that have the specific target volume V_t.

Different approaches to finding V_t can be considered. Since in the implementation the optimization domain is naturally bounded, we choose V_t to be a fixed

fraction of the volume of this domain. In a fast interactive framework the user can then adapt the target volume with the help of instant visual feedback. Most importantly, as opposed to a data term driven model volume constraints do not dictate where inflation takes place.

5.1 Optimization via Convex Relaxation

As in Section 4 we choose to relax the binary problem. This amounts to replacing U_V and U_Σ with their respective convex hulls U_V^r and U_Σ^r. The corresponding optimization problem is then convex:

Proposition 1. *The relaxed set $U^r := U_\Sigma^r \cap U_V^r$ is convex.*

Proof. The constraint in the definition of U_V is clearly linear in u and therefore U_V^r is convex. The same argument holds for U_Σ. Being an intersection of two convex sets U^r is convex as well.

One standard way of finding the globally optimal solution to this problem is gradient descent, which is known to converge very slowly. Since optimization speed is an integral part of an interactive reconstruction framework, we convert our problem to a form for which a *primal-dual* scheme [20] can be applied.

We start by replacing the TV-norm in our minimal surface problem by its weak equivalent:

$$\min_{u \in U^r} \int g(x)|\nabla u| d^3x = \min_{u \in U^r} \max_{|\xi(x)|_2 \leq g(x)} \left\{ \int -u \mathrm{div}\xi \; d^3x \right\} \qquad (12)$$

where $\xi \in \mathcal{C}_c^1(\mathbb{R}^3, \mathbb{R}^3)$.

The main problem is that we are dealing with an optimization problem over a constrained set. u needs to fulfill *three* constraints: Silhouette consistency, constant volume and $u \in [0, 1]$.

In order to maintain silhouette consistency (3) of the solution we simply restrict updates to those voxels which project onto the silhouette interior excluding the silhouette itself. Furthermore we reformulate the volume constraint as a Lagrange multiplier λ, which together with Equation (12) leads to the following Lagrangian dual problem [21]:

$$\max_{\substack{|\xi(x)|_2 \leq g(x) \\ \lambda}} \min_{u \in U_\Sigma^r} \left\{ \int -u \mathrm{div}\xi \; d^3x + \lambda \left(\int u \; d^3x - V_t \right) \right\} \qquad (13)$$

Equation (13) is a saddle point problem. In [20] it was presented how to solve such problems of this special form with a primal-dual scheme. We employ this scheme which is fast and provably convergent. It consists of alternating a gradient descent with respect to the function u and a gradient ascent for the dual variables ξ and λ interlaced with an over-relaxation step on the primal variable:

$$\begin{cases} \xi^{k+1} = \Pi_{|\xi(x)|_2 \leq g(x)}(\xi^k + \tau \cdot \nabla \bar{u}^k) \\ \lambda^{k+1} = \lambda^k + \tau \cdot (\int \bar{u} \; dx - V_t) \\ u^{k+1} = \Pi_{u \in [0,1]}(u^k - \sigma \cdot (\mathrm{div}\xi^{k+1} + \lambda)) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \end{cases} \qquad (14)$$

where Π_A denotes the projection onto the set A (see [20] for details). Note that the projection for the primal variable u now reduces to a clipping operation. Projection of ξ is done by simple clipping as well.

The scheme (14) is numerically attractive since it avoids division by the potentially zero-valued gradient-norm which appears in the Euler-Lagrange equation of the TV-norm. Moreover, it is parallelizable and we therefore implemented it on the GPU.

5.2 Optimality Bounds

Having computed a global optimal solution u_{opt} of Equation (12), the question remains how we obtain a binary solution and how the two solutions relate to one another energetically. Unfortunately no thresholding theorem holds, which would imply energetic equivalence of the relaxed optimum and its thresholded version for arbitrary thresholds.

Nevertheless we can construct a binary solution u_{bin} as follows:

Proposition 2. *The relaxed solution can be projected to the set of binary functions in such a way that the resulting binary function preserves the user-specified volume* V_t.

Proof. It suffices to order the voxels x by decreasing values $u(x)$. Subsequently, one sets the value of the first V_t voxels to 1 and the value of the remaining voxels to 0.

Concerning an optimality bound the following holds:

Proposition 3. *Let* u_{opt}^r *be the global optimal solution of the relaxed energy and* u_{opt} *the global optimal solution of the binary problem. Then*

$$E(u_{bin}) - E(u_{opt}) \leq E(u_{bin}) - E(u_{opt}^r) \ . \tag{15}$$

5.3 Theoretical Analysis of Material Concentration

As we have seen above, the proposed convex relaxation technique does not guarantee global optimality of the binary solution. The thresholding theorem [22] – applicable in the unconstrained problem – no longer applies to the volume-constrained problem. While the relaxation naturally gives rise to posterior optimality bounds, one may take a closer look at the given problem and ask why the relaxed volume labeling u should favor the emergence of solid objects rather than distribute the prescribed volume equally over all voxels.

In the following, we prove analytically that the proposed functional has an energetic preference for material concentration. For simplicity, we consider the case that the object silhouette in the image is a disk. And we compare the two extreme cases of all volume being concentrated in a ball (a known solution of the Cheeger problem) compared to the case that the same volume is distributed equally over the feasible space (namely a cylinder) – see Figure 8.
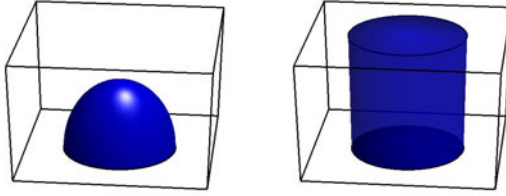
Fig. 8. The two cases considered in the analysis of the material concentration for the approach in Section 5. On the *left* hand side we assume a hemi-spherical condensation of the material. On the *right* hand side the material is distributed evenly over the volume.

Proposition 4. *Let u_{sphere} denote the binary solution which is 1 inside the sphere and 0 outside – Fig. 8, left side – and let u_{cyl} denote the solution which is uniformly distributed (i.e. constant) over the entire cylinder – Fig. 8, right side. Then we have*

$$E(u_{sphere}) < E(u_{cyl}), \tag{16}$$

independent of the height of the cylinder.

Proof. Let R denote the radius of the disk. Then the energy of u_{sphere} is simply given by the area of the half-sphere:

$$E(u_{sphere}) = \int |\nabla u_{sphere}| d^2x = 2\pi R^2. \tag{17}$$

If instead of concentrated to the half-sphere, the same volume, i.e. $V = \frac{2\pi}{3}R^3$, is distributed uniformly over the cylinder of height $h \in (0, \infty)$, we have

$$u_{cyl}(x) = \frac{V}{\pi R^2 h} = \frac{2\pi R^3}{3\pi R^2 h} = \frac{2}{3}\frac{R}{h}. \tag{18}$$

inside the entire cylinder, and $u_{cyl}(x) = 0$ outside the cylinder. The respective surface energy of u_{cyl} is given by the area of the cylinder weighted by the respective jump size:

$$\begin{aligned} E(u_{cyl}) &= \int |\nabla u_{cyl}| d^2x \\ &= \left(1 - \frac{2R}{3h}\right)\pi R^2 \ + \ \frac{2R}{3h}(\pi R^2 + 2\pi Rh) \\ &= \frac{7}{3}\pi R^2 > E(u_{sphere}). \end{aligned} \tag{19}$$

5.4   Experiments

In this section we study the properties of constant volume weighted minimal surfaces again within an interactive reconstruction environment. We show that appealing and realistic 3D models can be generated with minimal user input.
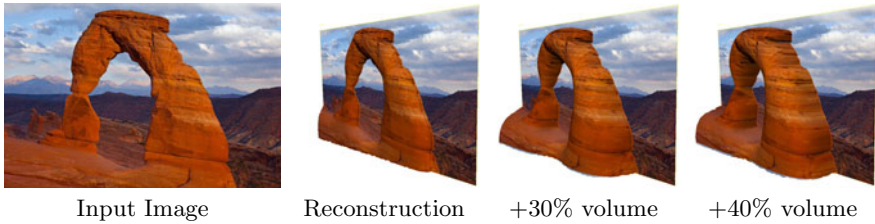
Input Image   Reconstruction   +30% volume   +40% volume

Fig. 9. By increasing the target volume with the help of a slider, the reconstruction is intuitively inflated. In this example the intial rendering of the volume with 175x135x80 voxels took 3.3 seconds. Starting from there each subsequent volume adaptation took only about 1 second.

Input Image   Reconstructed Geometry   Textured Geometry
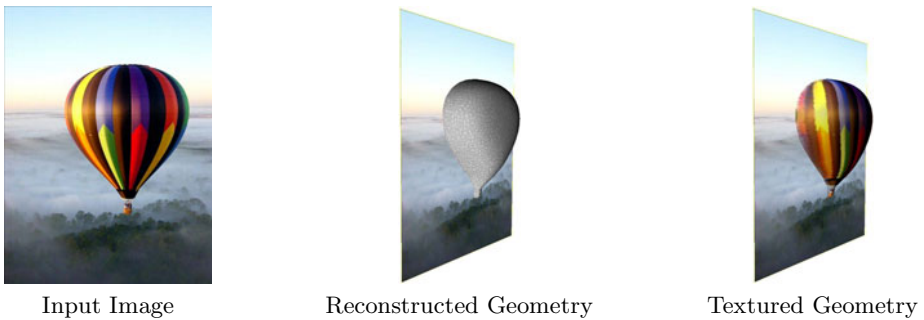
Fig. 10. The constant volume approach favors minimal surfaces for a user-specified volume. This amounts to solving a Cheeger set problem.

Cheeger Sets and Single View Reconstruction. Solutions to the problem in Eq. (10) are so called Cheeger sets, i.e. minimal surfaces for a fixed volume. In the simplest case of a circle-shaped silhouette the corresponding Cheeger set is a ball. Fig. 10 demonstrates that in fact round silhouette boundaries (in the unweighted case $g(x) \equiv 1$) result in round shapes. In the example of the balloon it also becomes apparent that thinner structures in the silhouette are inflated less than compact parts: Coming from the top of the balloon toward the basket on the bottom the inflation gradually degrades along with the silhouette width.

Varying the Volume. In the constant volume formalism presented in this section the only parameter we have to determine for our reconstruction is the target volume V_t (apart from the weighting function $g(x)$ of the TV-norm in Eq. (12)).

The effect of changing this scalar parameter on the appearance of the reconstruction surface can be witnessed in Fig. 9. One can see that the adaptation of the target volume has an intuitive effect on the resulting shape. This is important for an interactive user driven reconstruction process that is made possible by the small computation times that we gain through a parallel implementation of algorithm in Eq. (14).
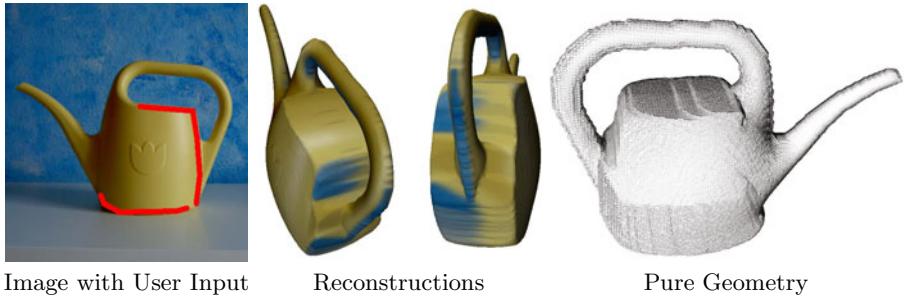
Image with User Input   Reconstructions   Pure Geometry

Fig. 11. The proposed approach allows to generate $3D$ models with sharp edges, marked by the user as locations of low smoothness (see Section 3). Along the red user strokes (*second from left*) the local smoothness weighting is decreased.
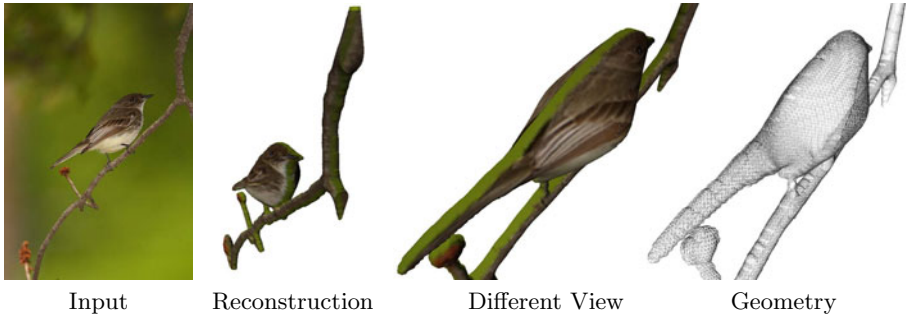

Input   Reconstruction   Different View   Geometry

Fig. 12. Volume inflation dominates where the silhouette area is large (bird) whereas thin structures (twigs) are inflated less

Sharp Edges. Similar to Section 4 we examine the effects of adapting the smoothness locally, but now using the volume prior instead of the shape prior.

Fig. 11 shows that by adapting the weighting function $g(x)$ of Eq. (12) not only round, but other very characteristic shapes can be modeled with minimal user interaction. The 2D user input is shown alongside with the reconstruction results. More experiments with smoothness adaptation for the constant volume case are presented in the following Section 6.

6   Comparison

In this section we compare the two proposed priors with respect to their reconstruction results, usability and runtime.

Comparison of Experimental Results. We have already indicated that the data term acts as a strong prior on the resulting shape of the reconstruction. This can be verified in Fig. 14: The left reconstruction was done with the shape prior as described in Section 4. Clearly the silhouette distance transform dominates
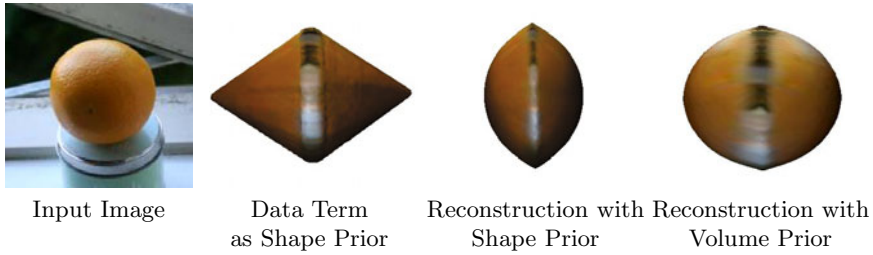
| Input Image | Data Term as Shape Prior | Reconstruction with Shape Prior | Reconstruction with Volume Prior |

Fig. 13. Using a silhouette distance transform as shape prior the relation between data term (*second from left*) and reconstruction (*third from left*) is not easy to assess for a user. The Cheeger set approach of Section 5 behaves more naturally in this respect (*right*).

the shape in the resulting reconstruction. This might of course be advantageous for particular shapes like the examples shown in Fig. 2. Still, often a Cheeger set (right of Fig. 14) is a better guess for natural shapes.

Increasing the smoothness parameter in the data term approach will mitigate the influence of the distance transform. However, with higher smoothness the result tends to be less voluminous making it hard to achieve ball-like shapes (see Fig. 13). In both approaches thin structures in the silhouette are less inflated than more compact parts. This is a basic property of minimal surfaces. Nevertheless in the data term approach thin structures tend to be too flat, especially in the presence of a high smoothness parameter (see Fig. 14).

In principle a data term inhibits the flexibility of the reconstructions. The air plane in Fig. 15 represents an example in which a parametric shape prior - just as the proposed data term of Section 4 - would fail to offer the necessary flexibility required for modeling protrusions. Since our fixed-volume approach does not impose points of inflation user input can influence the reconstruction more freely: Marking the wings as highly non-smooth (i.e. low weights $0 < g(x) < 0.3$) effectively makes them pop out. From a user perspective the shape prior approach is much more involved in terms of the amount of user input. The shape prior consists of four parameters to offer reasonable but still limited flexibility to adapt its shape. On the other hand, for the volume prior approach only one parameter needs to be specified by the user.

In Fig. 14 and 13 we make use of the same input images that were used in [12]. Comparing both results with the ones in their work reveals that we get comparable results with a significantly lower amount of user input. As opposed to their work, our method exhibits complete topological freedom. Figure 16 depicts a direct comparison of the two proposed priors on several reconstruction results. Again one can observe that the volume prior generally yields more roundish shapes, while the distance function dominates the results with the shape prior.

In sum, both priors yield comparable reconstruction results. Although for the shape prior approach the user has slightly more possibilities to adapt the final shape of the reconstruction, this freedom is paid off by a significant amount of additional user input as it is not always simple to find the right combination of
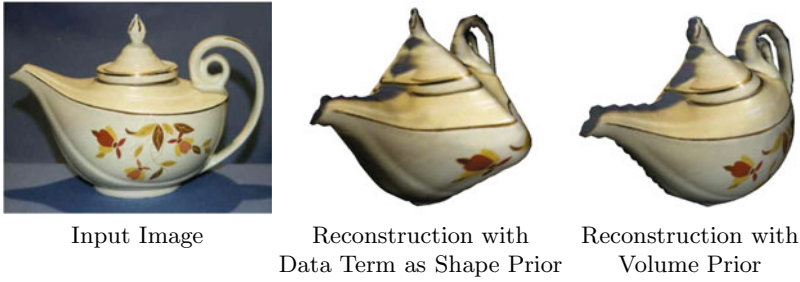
Input Image   Reconstruction with   Reconstruction with
Data Term as Shape Prior   Volume Prior

Fig. 14. In contrast to a solution with shape prior (Eq. 5) (*center*), the solution with volume prior (*right*) does not favor a specific shape and generates more natural looking results. Although in the center reconstruction the dominating shape prior can be mitigated by a higher smoothness (λ in Eq. (5)), this ultimately leads to the flattening of thin structures like the handle.

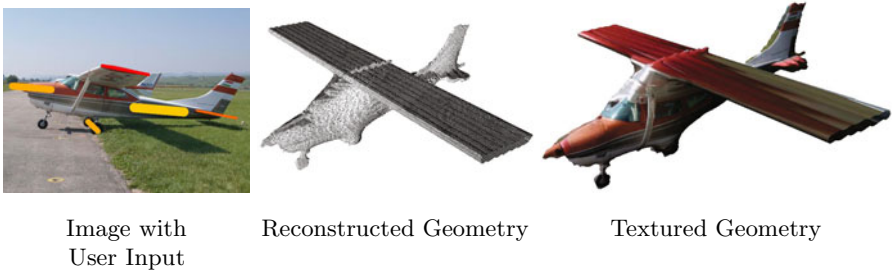Image with   Reconstructed Geometry   Textured Geometry
User Input

Fig. 15. An example for a minimal surface with prescribed volume and local smoothness adaptation. Colored lines in the input image mark user input, which locally alters the surface smoothness. Red marks low, yellow marks high smoothness (see Section 3 for details).

parameters. In contrast, the volume prior has only a single parameter making it simpler to adapt the shape of the reconstruction. A limitation of both methods is the implicit assumption that the plane of object symmetry should be approximately parallel to the image plane since the topology of the reconstructed object is directly inferred from the topology of the object's silhouette.

Runtime Comparison. The two priors lead to different optimization problems and we solved them with different optimization schemes. Both have been implemented in a parallel manner using the NVIDIA CUDA Framework. All computation times refer to a PC with a 2.27GHz Intel Xeon CPU and an NVIDIA GeForce GTX580 graphic device running a recent Linux distribution.

The computation times for the shape prior approach are slightly lower than the ones of the volume prior approach. For instance, the teapot example (Fig. 14) with a resolution of $189 \times 139 \times 83$ the method with shape prior needs 2 seconds while the method with volume prior needs 4.7 seconds. However, the computation times depend mainly on the volume resolution and also on the
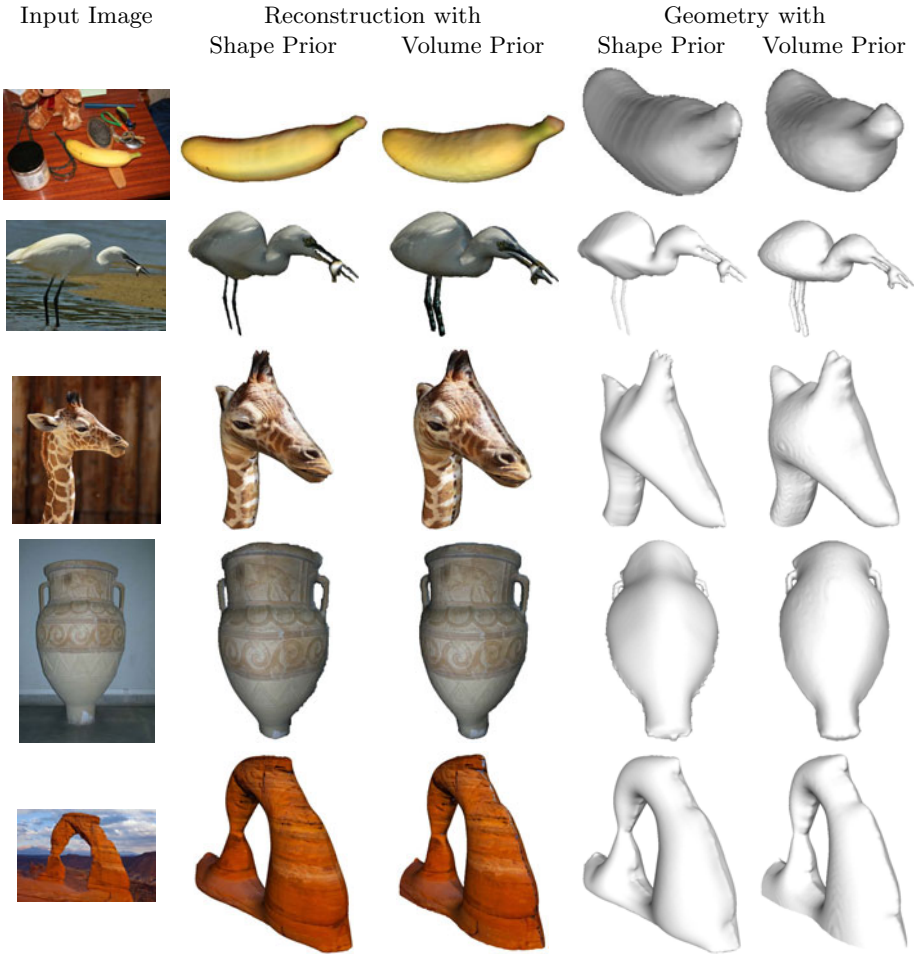
| Input Image | Reconstruction with | | Geometry with | |
| | Shape Prior | Volume Prior | Shape Prior | Volume Prior |

Fig. 16. Direct comparison of the methods with shape and volume prior for several examples

object to be reconstructed. For a reasonable quality of the reconstruction also lower volume resolutions may be sufficient. When using e.g. $63 \times 47 \times 32$ the computation times drop down to $0.03s$ and $0.13s$ for the methods with shape and volume prior, respectively.

7   Conclusion

In this work we considered a variational approach to the problem of 3D reconstruction from a single view by searching for a weighted minimal surface that is consistent with the given silhouette. A major part of our contribution is to show

that this can be done in an interactive way by providing a tool that is intuitive and computes solutions within seconds.

Two paradigms were proposed in order to deal with the highly ill-posed task. In the first one we introduce a shape prior that is incorporated as a data term in order to avoid flat solutions. This approach is along the lines of other works as it boils down to fixing depth values of the reconstruction in order to inflate it. In the other proposed method we search for a weighted minimal surface that complies with a fixed user given volume. The resulting Cheeger set problem goes without specifying expected depth of any sort thus providing more geometric flexibility of the result. In the former case we compute globally optimal solutions to the variational problem. In the latter case we showed that the solution lies within a bound of the optimum and exactly fulfills the prescribed volume. We compared both priors and found that the volume prior is more flexible and thus better suited for the task of single view reconstruction. On a variety of challenging real world images, we showed that the proposed method compares favorably over existing approaches, that volume variations lead to families of realistic reconstructions and that additional user scribbles allow to locally reduce smoothness so as to easily create protrusions.

References

1. Oswald, M.R., Töppe, E., Kolev, K., Cremers, D.: Non-parametric single view reconstruction of curved objects using convex optimization. In: Denzler, J., Notni, G., Süße, H. (eds.) Pattern Recognition. LNCS, vol. 5748, pp. 171–180. Springer, Heidelberg (2009)
2. Toeppe, E., Oswald, M.R., Cremers, D., Rother, C.: Image-based 3D modeling via cheeger sets. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part I. LNCS, vol. 6492, pp. 53–64. Springer, Heidelberg (2011)
3. Horry, Y., Anjyo, K.I., Arai, K.: Tour into the picture: using a spidery mesh interface to make animation from a single image. In: SIGGRAPH 1997: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 225–232. ACM Press, Addison-Wesley Publishing Co., New York, USA (1997)
4. Liebowitz, D., Criminisi, A., Zisserman, A.: Creating architectural models from images. In: Proc. EuroGraphics, vol. 18, pp. 39–50 (1999)
5. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Comput. Vision 40, 123–148 (2000)
6. Hoiem, D., Efros, A.A., Hebert, M.: Automatic photo pop-up. ACM Trans. Graph 24, 577–584 (2005)
7. Sturm, P.F., Maybank, S.J.: A method for interactive 3d reconstruction of piecewise planar objects from single images. In: Proc. BMVC, pp. 265–274 (1999)
8. Terzopoulos, D., Witkin, A., Kass, M.: Symmetry-seeking models and 3d object reconstruction. IJCV 1, 211–221 (1987)
9. Zhang, L., Dugas-Phocion, G., Samson, J.S., Seitz, S.M.: Single view modeling of free-form scenes. In: Proc. of CVPR, pp. 990–997 (2001)
10. Igarashi, T., Matsuoka, S., Tanaka, H.: Teddy: a sketching interface for 3d freeform design. In: SIGGRAPH 1999, pp. 409–416. ACM Press, Addison-Wesley Publishing Co., New York, USA (1999)

11. Nealen, A., Igarashi, T., Sorkine, O., Alexa, M.: Fibermesh: designing freeform surfaces with 3d curves. ACM Trans. Graph. 26, 41 (2007)
12. Prasad, M., Zisserman, A., Fitzgibbon, A.W.: Single view reconstruction of curved surfaces. In: CVPR, pp. 1345–1354 (2006)
13. Joshi, P., Carr, N.: Repoussé: Automatic inflation of 2d art. In: Eurographics Workshop on Sketch-Based Modeling (2008)
14. Cohen, L.D., Cohen, I.: Finite-element methods for active contour models and balloons for 2-d and 3-d images. IEEE Trans. on Patt. Anal. and Mach. Intell. 15, 1131–1147 (1993)
15. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary region segmentation of objects in n-d images. In: ICCV, vol. 1, pp. 105–112 (2001)
16. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph 23, 309–314 (2004)
17. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. The Clarendon Press Oxford University Press, New York (2000)
18. Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multview 3d reconstruction. International Journal of Computer Vision (2009)
19. Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., Cremers, D.: An experimental comparison of discrete and continuous shape optimization methods. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 332–345. Springer, Heidelberg (2008)
20. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth mumford-shah functional. In: IEEE Int. Conf. on Computer Vision, Kyoto, Japan (2009)
21. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
22. Chan, T., Esedoḡlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics 66, 1632–1648 (2006)