Non-Parametric Single View Reconstruction of Curved Objects using Convex Optimization

M. R. Oswald, E. Töppe[1], K. Kolev and D. Cremers

Computer Science Department, University of Bonn, Germany.

Abstract. We propose a convex optimization framework delivering intuitive and reasonable 3D meshes from a single photograph. For a given input image, the user can quickly obtain a segmentation of the object in question. Our algorithm then automatically generates an admissible closed surface of arbitrary topology without the requirement of tedious user input. Moreover we provide a tool by which the user is able to interactively modify the result afterwards through parameters and simple operations in a 2D image space. The algorithm targets a limited but relevant class of real world objects. The object silhouette and the additional user input enter a functional which can be optimized globally in a few seconds using recently developed convex relaxation techniques parallelized on state-of-the-art graphics hardware.

1 Introduction

One of the most impressive abilities of human vision is the extraction of three-dimensional information from a single image. From the mathematical point of view, depth information is lost due to the projection. In contrast to multiview methods, this operation cannot be simply inverted. Hence, depth information can only be guessed by image features like object contours, edges and texture patterns. Especially for images of textured objects under complex lighting conditions, shape from shading methods usually fail to work and further assumptions or user interactions are required. In computer vision, this fundamental problem has recently attracted a large amount of attention.

[1] This work was supported in part by Microsoft Research Cambridge through its PhD Scholarship Programme.

Fig. 1. Input images and textured reconstruction results from the proposed method.

1.1 Existing Approaches to Single View Reconstruction

Many approaches such as that of Horry et al. [1] aim to reconstruct planar surfaces by evaluating user defined vanishing points and lines. This has been extended by Liebowitz [2] and Criminisi [3]. Recently, this process has been completely automated by Hoiem et al. [4], yielding appealing results on a limited number of input images. Sturm et al. [5] 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. [6] 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. [7] addresses this problem and proposes a model which globally optimizes a smoothness criterion. However, it concentrates on estimating 2.5D image features rather than reconstructing real 3D representations. Moreover, it requires a huge amount of user interaction in order to obtain appealing reconstructions. In "Teddy", Igarashi et al. [8] make use of a contour based distance function in order to inflate a volume. The method performs modifications of a triangle mesh in multiple steps and is rather heuristic. This leads to problems with the maintenance of mesh consistency and suboptimal silhouette fitting results. Moreover, the object's topology is restricted to that of a sphere.

Closely related to our work, Prasad et al. [9] have studied the reconstruction of smooth and curved 3D surfaces from single photographs. They also calculate a globally optimal 3D surface satisfying user specified constraints. The main drawback of this method is the vast amount of necessary user input which is mainly due to the use of parametric surfaces. For a reconstruction, the user needs to select several contour edges and place them appropriately in the parameter space grid which requires explicit consideration of the topology of the object in question. Moreover, parameter space boundaries need to be considered and connected by the user requiring expert knowledge even for simple object topologies. For topologies of higher genus the required user placement of contour edges or creases in the parameter space may easily lead to over-oscillations of the surface and incorrect surface distortions.

To our knowledge, all existing approaches to single view reconstructions are based on *parametric* representations. Consequently, solutions will be affected by the choice of parametrization and extensions to different topology are by no means straight-forward.

1.2 Contributions

In this paper, we focus on the reconstruction of curved objects of arbitrary topology with a minimum of user input. We propose a convex variational method which generates a 3D object in the matter of a second using silhouette information only. Moreover, the proposed reconstruction framework provides the user with a simple but powerful post-editing toolbox which does not require expert knowledge at all. Post-editing can be done interactively due to the short computation times obtained by massively parallelized implementation of the underlying nonlinear diffusion process. Compared to previous works, the proposed method allows to compute globally optimal reconstructions of arbitrary topology due to the use of implicit surfaces and respective convex relaxation techniques.

In the following section, we will introduce a variational framework for single view reconstruction and show how it can be solved by convex relaxation techniques. In Sect. 3, we give a complete overview of the proposed reconstruction framework and explain how users can provide silhouette and additional information with minimal user interaction. The viability of our approach is tested on several examples in Sect. 4, followed by concluding remarks in Sect. 5.

2 Variational Framework for Single View Reconstruction

In the following, we introduce a variational framework for single view reconstruction. The proposed functional can subsequently be optimized using convex relaxation techniques recently developed for segmentation [10] and multiview reconstruction [11].

2.1 Variational Formulation

Let $V \subset \mathbb{R}^3$ be a volume surrounding the input image $I : \Omega \mapsto \mathbb{R}^3$ with image plane $\Omega \subset V$. We are looking for a closed surface $\Sigma \subset V$ which inflates the object in the image I and is consistent with its silhouette S. For simplicity, an orthographic projection is assumed and defined by $\pi : V \mapsto \Omega$. In order to handle arbitrary topologies, the surface Σ is represented implicitly by the indicator function $u : V \mapsto \{0, 1\}$ denoting the exterior ($u = 0$) or interior ($u = 1$) of the surface.

A smooth surface with the desired properties is obtained by minimizing the following energy functional:

$$E(u) = E_{\text{data}}(u) + \nu E_{\text{smooth}}(u) \ , \tag{1}$$

where $\nu \geq 0$ is a parameter controlling the smoothness the surface. The smoothness term is imposed via the weighted total variation (TV) norm

$$E_{\text{smooth}}(u) = \int_V g(x) \, |\nabla u(x)| \, d^3x \ , \tag{2}$$

where the diffusivity $g : V \mapsto \mathbb{R}_+$ can be used to adaptively adjust smoothness properties of the surface in different locations. The range of g needs to be nonnegative to maintain the convexity of the model. The data term

$$E_{\text{data}}(u) = \int_V u(x) \ \phi_{\text{vol}}(x) d^3x + \int_V u(x) \ \phi_{\text{sil}}(x) d^3x \tag{3}$$

realizes two objectives: volume inflation and compliance with silhouette constraints.

2.2 Silhouette Consistency

The function $\phi_{\mathrm{sil}}(x)$ merely imposes silhouette consistency. It assures that all points projecting outside the silhouette will be assigned to the background ($u=0$) and that all points which are on the image plane and inside the object will be assigned object ($u=1$):

$$\phi_{\mathrm{sil}}(x) = \begin{cases} -\infty & \text{if } \chi_S(\pi(x)) = 1 \text{ and } x \in \Omega \\ +\infty & \text{if } \chi_S(\pi(x)) = 0 \\ 0 & \text{otherwise} \end{cases} , \tag{4}$$

where characteristic function $\chi_S : \Omega \mapsto \{0, 1\}$ indicates exterior or interior of the silhouette, respectively.

2.3 Volume Inflation

The volume inflation function ϕ_{vol} allows to impose some guess of the shape of the object. The function can be adopted to achieve any desired object shape and may also be changed by user-interaction later on. In this work, 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

$$dist(p, \partial S) = \min_{s \in \partial S} \|p - s\| \quad , \tag{5}$$

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

$$\phi_{\mathrm{vol}}(x) = \begin{cases} -1 & \text{if } dist(x, \Omega) \leq h(\pi(x)) \\ +1 & \text{otherwise} \end{cases} , \tag{6}$$

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

$$h(p) = \min \left\{ \lambda_{\mathrm{cutoff}}, \lambda_{\mathrm{offset}} + \lambda_{\mathrm{factor}} * dist(p, \partial S)^k \right\} \tag{7}$$

with four parameters $k, \lambda_{\mathrm{offset}}, \lambda_{\mathrm{factor}}, \lambda_{\mathrm{cutoff}} \in \mathbb{R}_{>0}$ affecting the shape of the reconstructed object. How the user can employ these parameters to modify the computed 3D shape will be discussed in Sect. 3.

Note that this choice of ϕ_{vol} 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.

2.4 Optimization via Convex Relaxation

To minimize energy (1) we follow the framework developed in [11]. To this end, we relax the binary assumption by allowing u to take on intermediate values,

Fig. 2. Workflow of the proposed method: Input image with user provided seeds (foreground: blue, background: red), segmentation, reconstruction with default parameter settings, reconstruction with user-adapted parameter settings.

i.e. $u : V \to [0, 1]$. Subsequently, we can globally minimize the convex functional (1) by solving the corresponding Euler-Lagrange equation

$$0 = \phi_{\text{vol}} + \phi_{\text{sil}} - \nu \ \text{div} \left(g \frac{\nabla u}{|\nabla u|} \right) \quad , \tag{8}$$

using a fixed-point iteration. A global optimum of the original binary labeling problem is then obtained by simple thresholding of the solution of the relaxed problem – see [11] for details.

In [12] 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 parallelizability. These two aspects allow to compute smooth single view reconstructions with no grid bias within a few seconds using standard graphics hardware.

3 Interactive Single View Reconstruction

To make optimal use of the proposed reconstruction method we explain its integration within an interactive tool and which methods can be used to obtain good reconstructions with only a few mouse clicks. The typical workflow of our method is depicted in Fig. 2 and single stages are further explained in the following subsections.

3.1 Image Segmentation

The main prerequisite for a good result with the algorithm proposed in Sect. 2 is a reasonable silhouette. The number of holes in the segmentation of the target object determines the topology of the reconstructed surface. Notably, the proposed reconstruction method 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 [13] and [14]. The algorithm calculates two distinct regions based on respective color histograms which are defined by representational pen strokes given by the user (see Fig. 2). The output of the segmentation defines the silhouette indicator function χ_S.

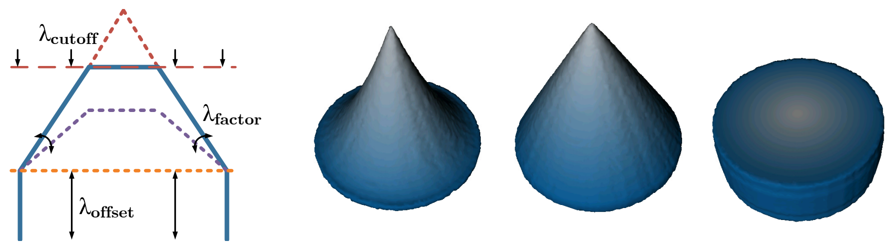

Fig. 3. Effect of $\lambda_{\text{offset}}, \lambda_{\text{factor}}, \lambda_{\text{cutoff}}$ (left) and various values of parameter k and resulting (scaled) height map plots for a circular silhouette.

3.2 Interactive Editing

From the input image and silhouette a first reconstruction is generated, 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, we propose several editing techniques on a 1D (parameter) and a 2D (image space) level. The goal is to have easy-to-use editing tools which cover important cases of object features.

In this paper we present three different kinds of editing tools: parameter-based, contour-based and curve-based tools. The first two classes operate directly on the data-term of (1), whereas the third one alters the diffusivity of the TV-norm (2).

Data Term Parameters By altering the parameters $\lambda_{\text{offset}}, \lambda_{\text{factor}}, \lambda_{\text{cutoff}}$ and the exponent k of the height map function (7), users can intuitively change the data term (3) 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. 3 and are quite intuitive to grasp. The exponent k of the distance function in (7) mainly influences the objects curvature in the proximity of the silhouette contour. This can be observed in Fig. 3 showing an evolution from a cone to a cylinder just by decreasing k.

Local Data Term Editing Due to the use of a distance function for the volume inflation, depth values of the data term will always increase for larger distances to the silhouette contour. Thus, large depth values will never occur near the silhouette contour. However this can become necessary for an important class of object shapes like for instance the bottom and top of the vase in Fig. 4. A simple remedy to this problem is to ignore user specified contour parts during the calculation of the distance function. We therefore approximate the object contour by a polygon which is laid over the input image. The edges of the polygon are points of high curvature and each edge represents the contour pixels

Fig. 4. 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.

between the endpoints. By clicking on the edge, the user indicates to ignore the corresponding contour pixels during distance map calculation (see Fig. 4 top right).

Local Discontinuities Creases on the surface often add critically to the characteristic shape of an object. With the diffusivity function of the smoothness term (2) we are given a natural way of integrating discontinuities into the surface reconstruction. By setting the values of g to less than one for certain subsets of the domain, the smoothness constraint is relaxed for these regions. Accordingly for values greater than one smoothness is locally fortified. To keep things simple, we let the user specify curves of discontinuities by drawing them directly into the input image space. In the reconstruction space, the corresponding pre-images are uniquely defined hyperplanes (remembering that we make use of parallel projection). For the points lying on these planes or surrounding them, the diffusivity is reduced resulting in a surface crease at the end of the reconstruction process.

3.3 Implementational Issues

In order to efficiently solve the Euler-Lagrange equation (8) and allow fast interactive modeling the choice of the solving method and its appropriate implementation is crucial to achieve short calculation times.

Instead of minimizing (1) with a gradient descent scheme, we solve the approximated system of linear equations with successive over-relaxation (SOR) as proposed in [11]. On the one hand, this increases the convergence speed drastically and on the other the solution method can be parallelized to further increase computational speed. Therefore, we make use of the CUDA framework to implement SOR with a Red-Black scheme which speeds up calculations by factor 6 compared to the sequential method. Moreover, the computational effort for the surface evolution during interactive modeling can be further reduced by initializing the calculations with the previous reconstruction result. For small parameter changes this initialization is usually close to the next optimal solution. In sum, this allows single view reconstruction close to realtime.

4 Experiments

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

The experimental results are shown in Fig. 5. Default values for the data term parameters (7) are $k = 1$, $\lambda_{\text{offset}} = 0$, $\lambda_{\text{factor}} = 1$, $\lambda_{\text{cutoff}} = \infty$. Each row depicts several views of a single object reconstruction starting with the input image.

The following main advantages are showcased in the examples. The fence (top row) is an example of an object with complex topology, the algorithm can handle. Obviously reconstructions of the shown type are nearly impossible to achieve with the help of parametrized representations. The same example is also a proof for how little user interaction is necessary in some cases to obtain a good reconstruction result. In fact, the fence was automatically generated by the method right after the user segmentation stage. The rest of the examples demonstrate the power of the editing tools described in Sect. 3. The reconstructions were edited by adding creases and selecting sharp edges. It can be seen, that elaborate modeling effects can be readily achieved with these 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 the lower three rows 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. To keep the runtime and memory demand within convenient limits, we work on 256^2-input images. These result in a very detailed mesh. On a GeForce GTX card an update step of the geometry takes about 2-15 seconds, dependent on the applied operation.

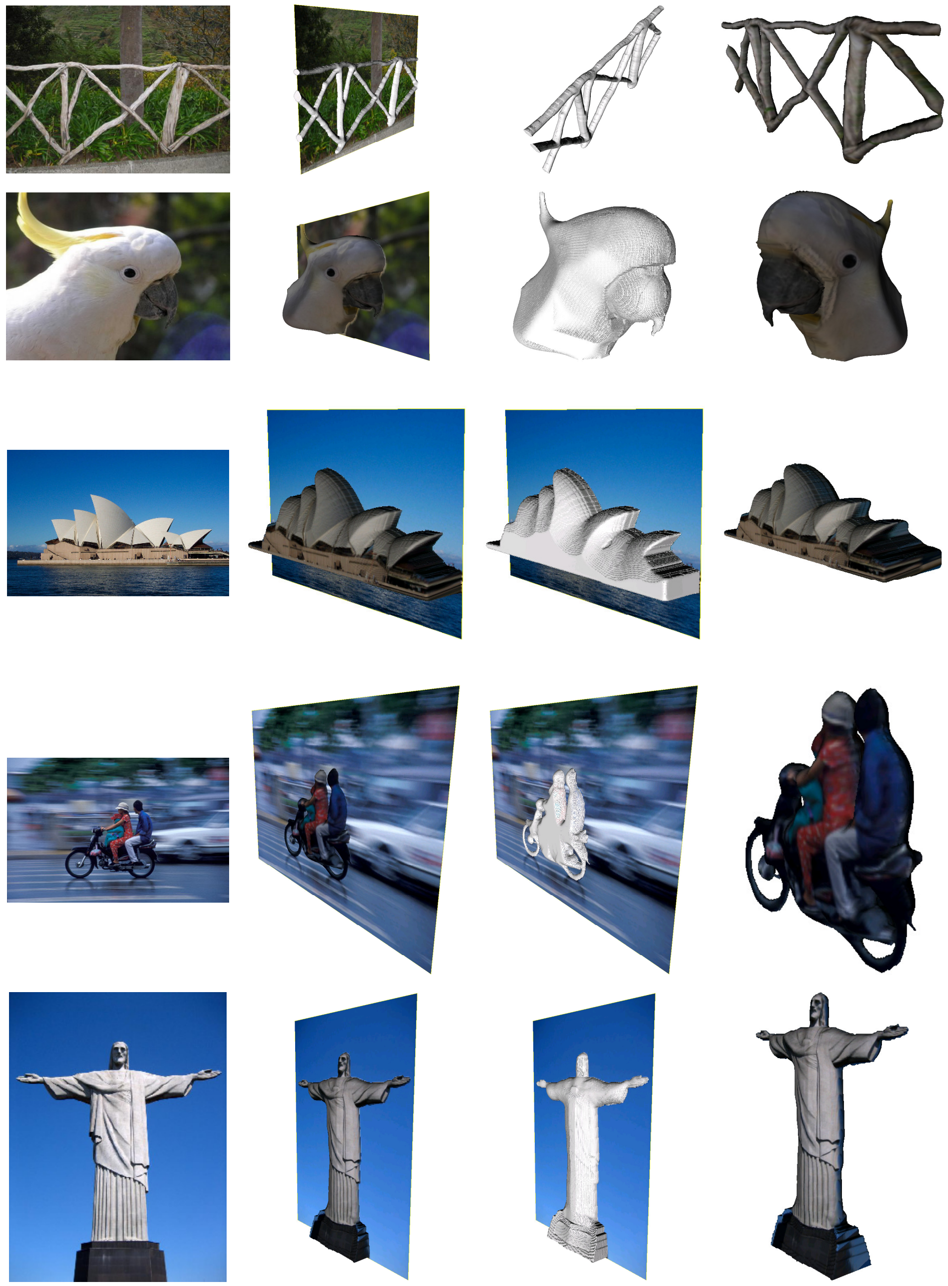

Fig. 5. Input images (1st column) and corresponding reconstruction results (2nd-4th column): textured model, untextured geometry, textured model without image plane.

5 Conclusion

We presented the first variational approach for single view reconstruction of curved objects with arbitrary topology. It allows to compute a plausible 3D model for a limited but reasonable class of single images. By using an implicit surface representation we eliminate the dependency on a choice of surface parameterization and the subsequent difficulty with objects of varying topology. The proposed functional integrates silhouette information and additional user input. Globally optimal reconstructions are obtained via convex relaxation. The algorithm can be used interactively, since the parallel implementation of the underlying nonlinear diffusion process on standard graphics cards only requires a few seconds. Compared to other works, the amount of user input is small and intuitive, post-editing is kept simple and does not require expert knowledge. Future work is focused on incorporating information from edges, pattern or shading to further improve the quality of reconstructions.

Acknowledgements. We thank Mukta Prasad, Carsten Rother and Andrew Fitzgibbon for helpful discussions, suggestions and for providing their images.

References

1. 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) 225–232
2. Liebowitz, D., Criminisi, A., Zisserman, A.: Creating architectural models from images. In: Proc. EuroGraphics. Volume 18. (1999) 39–50
3. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Comput. Vision **40**(2) (2000) 123–148
4. Hoiem, D., Efros, A.A., Hebert, M.: Automatic photo pop-up. ACM Trans. Graph. **24**(3) (2005) 577–584
5. Sturm, P.F., Maybank, S.J.: A method for interactive 3d reconstruction of piecewise planar objects from single images. In: In Proc. BMVC. (1999) 265–274
6. Terzopoulos, D., Witkin, A., Kass, M.: Symmetry-seeking models and 3d object reconstruction. IJCV **1** (1987) 211–221
7. Zhang, L., Dugas-Phocion, G., Samson, J.S., Seitz, S.M.: Single view modeling of free-form scenes. In: Proc. of CVPR. (2001) 990–997
8. Igarashi, T., Matsuoka, S., Tanaka, H.: Teddy: a sketching interface for 3d freeform design. In: SIGGRAPH '99. (1999) 409–416
9. Prasad, M., Zisserman, A., Fitzgibbon, A.W.: Single view reconstruction of curved surfaces. In: CVPR. (2006) 1345–1354
10. Chan, T., Esedoḡlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM **66**(5) (2006) 1632–1648
11. Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. Int.J. of Comp.Vision **84**(1) (August 2009) 80–96
12. Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., Cremers, D.: An experimental comparison of discrete and continuous shape optimization methods. In: Proc. ECCV. (2008)
13. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary region segmentation of objects in n-d images. In: ICCV 2001. 105–112 vol.1
14. Rother, C., Kolmogorov, V., Blake, A.: "grabcut": interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. (3) (2004) 309–314