TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Computer Vision und Image Analysis

Convex Optimization Methods for Single View 3D Reconstruction

Eno Töppe

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Darius Burschka
Prüfer der Dissertation:	
	1. Univ.-Prof. Dr. Daniel Cremers
	2. Univ.-Prof. Dr. Dr. h.c. mult. Wolfgang Förstner Rheinische Friedrich-Wilhelms-Universität Bonn

Die Dissertation wurde am 14.06.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 02.09.2013 angenommen.

Abstract

Single view reconstruction denotes the task of estimating the 3D geometry of an object from only a single input image. The problem is inherently ill-posed as unique depth is only recoverable for very special cases and within controlled environments. Existing approaches, therefore, require demanding user input to solve the problem.

In this thesis we take a variational approach to the interactive 3D reconstruction of closed objects relying on simple and intuitive user input. The key idea is to compute minimal surfaces, a principle which is common for natural and man-made objects. Furthermore, the minimal surface must comply with the object's silhouette which is extracted from the input image by a simple user guided segmentation.

For improving the reconstruction quality we propose three constraints, which are included in the minimal surface approach. Firstly, a constraint on the object volume averts flat reconstructions and controls the compactness of the object. The second constraint imposes characteristic shape information which scales with the object volume. The last constraint relates the sizes of object parts and allows for self-occlusions, dents and protuberances. Moreover, we propose a new method that infers the shape information automatically from the shading in the input image.

To compute the reconstruction surface an energy functional is minimized via convex relaxation and fast primal-dual schemes. The solution can be shown to be either globally optimal or to lie within small bounds from the optimum. To accelerate the reconstruction process, we conceive, implement and compare several numerical optimization methods, among others the Fast Iterative Shrinkage Thresholding Algorithm (FISTA), the Alternating Direction Method of Multipliers (ADMM), parallelized Succesive Overrelaxation (SOR) and Primal-dual (PD).

The developed single view 3D reconstruction software keeps necessary user input simple and intuitive, it harnesses parallel computation hardware for an efficient reconstruction process and compares well to state-of-the-art methods.

Acknowledgements

I am indebted to Prof. Dr. Daniel Cremers for his valuable and professional guidance, for introducing me to scientific working, for the inspiration, his faith and also for his constant personal support over the last years. Especially, I am grateful for the very many opportunities that I was offered to evolve in research, to find my way and to realize my potential. My thanks also go to Carsten Rother for many interesting discussions and Microsoft for supporting this work and offering many opportunities for scientific exchange.

Furthermore, I would like to thank my colleagues for many fruitful discussions, a constant willingness to help and many fun activities in our spare time. Also, the majority of works in this thesis were developed in close collaboration with Martin Oswald, whom I thank for a fruitful and pleasant teamwork.

Last but definitely not least, my gratitude and deep respect go to Claudia Nieuwenhuis from whom I have learned so much professionally as well as personally and without whom this thesis would not have become what it is.

Contents

1 Introduction **3**

2 Basic Mathematical and Algorithmic Concepts **7**
2.1 Convex Optimization in Computer Vision 7
2.2 Convexity 8
2.3 Convex Relaxations 12
2.4 Variational Calculus 12
2.5 Algorithms for Convex Optimization 14
2.6 Total Variation 18

3 Literature Overview **21**
3.1 Image Cues and Priors 21
3.2 A Classification of Approaches to Single View 3D Reconstruction 25
3.3 Properties and Comparison of Related Works 31
3.4 Conclusion 34

4 Interactive Workflow and the Minimal Surface Assumption **35**
4.1 Interactive Work Flow 35
4.2 Basic Assumptions on the Reconstructions 37
4.3 Minimal Surfaces 37

5 A First Variational Method Using a Shape Prior **39**
5.1 Variational Formulation with Shape Prior 39
5.2 Efficient Optimization via Convex Relaxation 41
5.3 Experimental Results 44
5.4 Conclusion 46

6 Cheeger Set Formulation **49**
6.1 Variational Formulation with Volume Constraints 50
6.2 Efficient Optimization via Convex Relaxation 51
6.3 A Theoretical Analysis of the Compactness of Solutions 58
6.4 Experimental Results 59
6.5 Comparison to State-of-the-art Single View Approaches 66
6.6 Conclusion 72

7 A Globally Optimal Minimal Surface Approach Based on Height Fields 75
7.1 A Globally Optimal Minimal Surface Approach Based on Height Fields . 76
7.2 Efficient Optimization 79
7.3 Experimental Results 85
7.4 Conclusion 89

8 Single View 3D Reconstruction with Relative Volume Constraints 91
8.1 Absolute and Relative Volume Constraints 92
8.2 Efficient Optimization via Convex Relaxation 95
8.3 Interactive Modeling with Relative Volume Constraints 96
8.4 Estimating Relative Depth Profiles from Shading Information 98
8.5 Experimental Results 101
8.6 Conclusion 107

9 Conclusion 109

Own Publications 113

Bibliography 115

Appendix: Derivation of the Shrinkage Formula 125

1 Introduction

This thesis deals with the problem of inferring 3D geometry of one or several objects from just a single image. In particular, we will concentrate on estimating closed representations of objects, which can be viewed from all directions, instead of mere height fields.

The problem of *single view 3D reconstruction* - as we will denote this task throughout this thesis - is one of the fundamental tasks in computer vision, yet it is still far from being solved. It may surprise at a first glance, that compared to other popular fields there are few works that deal with this fundamental task. However, being one of the most *ill-posed* tasks in computer vision, single view 3D reconstruction is an extremely hard problem. Since it is mathematically impossible to invert the image formation process based on a single image, approaches to single view 3D reconstruction have to confine their reconstruction domain to very specific problems and make strong assumptions on the objects, their geometry or the original images.

This thesis will be no exception. We will have to focus on certain types of objects, feasible views and scene contexts in order to make the problem tractable. Yet, we will open up a new perspective on the task conceptionally, methodically, mathematically and computationally. In particular and to the best of our knowledge, we will present the first variational approach to single view 3D reconstruction based on energy minimization.

Why should continuous convex optimization methods be well suited for solving the very task of single view 3D reconstruction? Convex optimization has been around for longer in mathematical vision to solve problems such as denoising [90], inpainting [23], stereo [82], binary- and multi-label segmentation [22, 20, 110, 66], 3D reconstruction [62], super-resolution [40, 75], motion estimation [94] and optical flow [111, 96]. Furthermore, in [61] and [3, 2] it was shown that many convex relaxation methods have several advantages over graph cut methods, the two main ones being the avoidance of metrication errors (which occur due to the approximation of the L_2 norm in the total variation of discrete formulations) and the strong parallelizability. Still, the usefulness of convex optimization methods in connection with a strongly ill-posed problem such as ours is not obvious.

Therefore, this thesis demonstrates that the framework represents a new paradigm addressing several common shortcomings of existing approaches to the reconstruction of curved objects. More precisely it offers new solutions to non-heuristic surface inflation and topological freedom of reconstruction surfaces. A major goal of this thesis is to show that convex variational energy minimization methods lead to elegant, precise and simple

mathematical formulations for interactive 3D reconstruction from a single view that can be solved efficiently with strongly parallelizable algorithms within interactive runtimes. The proposed formulations can be optimized either globally optimally or within bounds from the global optimum. Especially in ill-posed settings it can be considered a strong benefit to have well defined global optima. Most importantly, the proposed approaches are sustainable by their high degree of extendability. Priors and constraints will be added to the basic framework as they help to sensibly restrict the solution space and allow for more reasonable reconstructions. In this context, simple and sparse user input, fundamental prior assumptions and shape information inferred from the input image will be exploited.

As is the case with many vision tasks, human perception will be a major guideline in the development of algorithms. Humans have a remarkably well developed ability to estimate relative 3D geometric information of never before seen objects from only a single image. Humans excel at this task, however studies have also shown that they, too, fail in assessing absolute depth information in a single view. Consequently, in this thesis, our goal is to infer reasonable reconstructions that comply with the human intuition of the shown object rather than exact solutions with respect to a ground truth.

Although *learning* certainly plays a significant role in human perception, for the proposed approaches we deliberately refrain from learning aspects. The reason is that such methods tend to strongly restrict the reconstruction domain. In addition, training data often has to cover a wide spectrum of different images and scene contexts in order to work well for arbitrary inputs. Instead, this thesis will show that a generic approach based on sparse user input and strong priors leads to compelling 3D estimates that can be used as 3D models for additional applications without the need for post-processing and further editing.

Applications of single view 3D reconstruction approaches very much depend on the proposed algorithms, their target images and the type of reconstructions they result in. In general, typical applications include high-level image editing tasks such as new view synthesis or relighting. The closed representations obtained by the approaches in this thesis can also be used as models for augmented reality applications, games and others.

The outline of this thesis is as follows. The first chapter will introduce mathematical concepts on convexity, optimization and minimal surfaces that will form the basis for the reconstruction approaches in this thesis. The second chapter will be a thorough introduction to the topic of single view 3D reconstruction. We will discuss the problem definition and provide a survey on related work that will include a classification and comparison. A closer insight will be given into algorithms that are strongly related to the approaches in this thesis. In Chapter 4 a new framework for single view reconstruction based on silhouette compliant minimal surfaces will be introduced that will lay the foundation for the approaches presented in the rest of the thesis. This includes Chapter 5, where we explore the formulation of parametric shape priors in a first approach to single view reconstruction, Chapter 6, which introduces a non-heuristic inflation strategy

and Chapter 8, which is dedicated to the design of dynamic shape priors that infer shape characteristics from shading information in the input image. An approach that also computes minimal surfaces but is based on a different representation and computes guaranteed global optimal solutions is proposed in Chapter 7.

2 Basic Mathematical and Algorithmic Concepts

2.1 Convex Optimization in Computer Vision

Computer vision tasks are often formulated as energy minimization problems, either in a finite dimensional discrete or in a variational continuous setting. In both cases, an energy functional is defined that often stems from a Bayesian interpretation and models a specific problem statement by assigning an energy to each possible configuration in a solution space. The task then is to find a feasible configuration with minimal energy. The computational complexity for this task quickly rises with the number of variables, so for most interesting problems a brute force search of the (discretized) solution space easily becomes infeasible. Extremality conditions constitute necessary circumstances for optimality, but they only indicate local instead of global optima. Therefore, often smarter optimization strategies are necessary that take advantage of the specific structure of the given problem formulation. Alternatively, an approximation of the problem with a simpler structure can be optimized.

One such desirable problem structure is *convexity*. **Convex functionals have the favorable property that each local minimum is a global minimum**. We will discuss formal definitions later on in this chapter. The analogon in the spatially discrete setting is *submodularity*.

Convex optimization deals with the challenge of approximating computational challenges as convex energy minimization problems and solving them efficiently. Since few reasonable energy models are convex by nature one has to revert to relaxation techniques that transform the models into convex counterparts which are easier to solve. Of course such an approach is only promising if some tight relation between the solutions of the relaxed and the original non-convex problem can be established. Devising optimization problems for which such relaxations exist that are as close to the original problem as possible is one of the biggest challenges in convex optimization. In many cases convex relaxation methods do not lead to globally optimal solutions of the original problem; however, they often result in provable optimality bounds. Furthermore, optimization algorithms based on convex relaxations are usually independent of their initialization as there are no local optima.

One may argue that convexity is a rather strong assumption for energy functions that model real world problems. However, many important and complex problems in early

and higher vision have already been solved over the last years either globally optimally or within provable optimality bounds by applying convex optimization methods. Pock et al. [82] showed that multi-label problems can be solved globally optimally by functional lifting assuming a linear label space, a discontinuity preserving TV regularizer and an arbitrary data fidelity term. An analogous result for the discrete case had been published before by Ishikawa et al. [55] for a convex data fidelity. These algorithms can be applied, e.g., to stereo problems. For the piece-wise smooth and piece-wise constant cases of the Mumford-Shah functional convex relaxation frameworks were proposed by Pock et al. [81] and others [20, 110, 66] leading to solutions that are within bounds from the global optimum. The latter case is also known as the minimal partition problem and can be used for segmentation with [110] and without [97] metric distance measures on the label space. Chan et al. [22] showed that the minimal partition problem for the binary labeling case can even be optimized globally by convex relaxation. Their result is relevant to some of the proposed approaches in this thesis and will, therefore, be reviewed later on. Furthermore, convex relaxation techniques have also been proposed for multi-label problems on product label spaces [41], which is suited for computing optical flow.

For the rest of this chapter we will give an introduction to the theoretical basics of convex functionals, convex optimization and relaxation techniques. This will lay the foundation for the chapters ahead.

2.2 Convexity

2.2.1 Convexity of Sets and Functionals

Let $\mathcal{V}$ be a vector space. A set $\mathcal{C} \subset \mathcal{V}$ is called *convex* if the linear convex combination of arbitrary points in the set lie in the set as well. Formally, this is subsumed by

Definition 1. A set $\mathcal{C} \subset \mathcal{V}$ is convex if

$$\forall x, y \in \mathcal{C}, \lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda)y \in \mathcal{C} \ . \tag{2.1}$$

Intuitively this means that each line connecting two points in the set will be completely contained by the set. Convex sets are closed with respect to intersection, but not with respect to the union operation.

Definition 2. A functional $f : \mathcal{V} \to \mathbb{R}$ is convex if its epigraph

$$\mathrm{epi}(f) := \{(x, \lambda) \ | \ x \in \mathcal{V} \text{ and } \lambda \geq f(x)\} \tag{2.2}$$

is a convex set.

The epigraph can be seen as the set of points "above" the graph of a real valued functional (see Figure 2.1 for an example). One can easily see that for a continuous

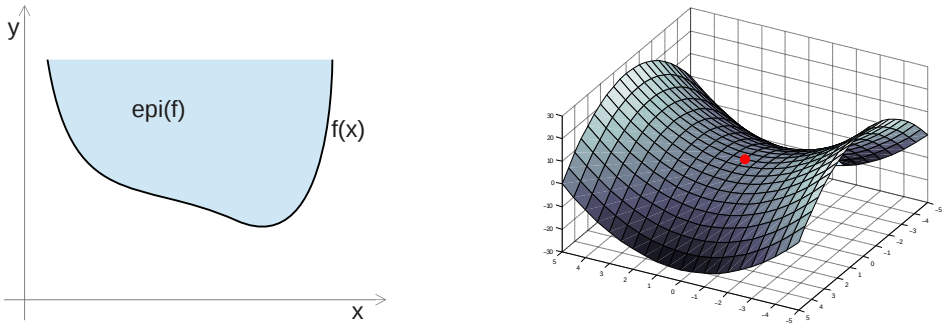


Figure 2.1: *Left*: example of a function f with its corresponding epigraph in blue $epi(f)$. *Right*: a plot of the function $f(x, y) = x^2 - y^2$ with a saddle-point at $(0, 0)$ (red).

function having two local minima, the epigraph cannot be convex, as the line connecting both minima will necessarily leave the epigraph.

By applying Definition (2.1) to points lying on the epigraph boundary, one arrives at another definition of convex functionals:

Definition 3. A functional $f : \mathcal{V} \to \mathbb{R}$ is convex if for all $x, y \in \mathcal{V}$ and $\lambda \in [0, 1]$ it holds

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)\,. \tag{2.3}$$

This definition is well suited for testing convexity of a given function. Intuitively, the definition means that the line between two function values must lie above the function graph.

2.2.2 Saddle-Point Problems

Closely related to convex functionals are functionals that are *convex-concave*.

Definition 4. A functional $E : \mathcal{U} \times \mathcal{V} \to \mathbb{R} \cup \{-\infty, \infty\}$ defined on vector spaces U and V is called *convex-concave* if

- $\forall v_0 \in \mathcal{V},\ u \mapsto E(u, v_0)$ is convex and
- $\forall u_0 \in \mathcal{U},\ v \mapsto E(u_0, v)$ is concave.

A functional E is *concave* if $-E$ is convex. Convex-concave functionals do not have minima as in the convex case, instead, one searches for a *saddle-point*. A saddle-point is a point $(\hat{u}, \hat{v}) \in \mathcal{U} \times \mathcal{V}$ for which holds that if one component is fixed, the other component yields an extremum of the resulting functional (see Figure 2.1).

Definition 5. For a concave-convex functional $E : \mathcal{U} \times \mathcal{V} \to \mathbb{R} \cup \{-\infty, \infty\}$ a point $(\hat{u}, \hat{v}) \in \mathcal{U} \times \mathcal{V}$ is called a saddle-point if

- $\hat{u} \in \operatorname{argmin}_u E(u, \hat{v})$ and
- $\hat{v} \in \operatorname{argmax}_v E(\hat{u}, v)$

Similar as for global minima of convex functionals, saddle-points of convex-concave functionals are not necessarily unique, but different saddle-points have the same energy. Another characterization of a saddle-point $(\hat{u}, \hat{v})$ is that

$$\min_u \max_v E(u, v) = \max_v \min_u E(u, v) = E(\hat{u}, \hat{v}) \ . \tag{2.4}$$

This is convenient as it allows us to commutate min and max.

An important example for a saddle-point problem is given by the *Lagrangian dual* which is used to formulate convex optimization problems under constraints. Given a convex functional $f : \mathcal{U} \to \mathbb{R}$ which is constrained by $g(x) = c$ for a convex functional $g : \mathcal{U} \to \mathbb{R}$, one can write the constrained convex optimization problem as a saddle-point problem by adding a Lagrange multiplier $\lambda \in \mathcal{V} = \mathbb{R}$ and maximizing with respect to this multiplier

$$\inf_{\substack{u \in \mathcal{U} \\ g(u)=c}} f(u) \ = \ \sup_{\lambda \in \mathcal{V}} \inf_{u \in \mathcal{U}} f(u) + \lambda(g(u) - c) \ . \tag{2.5}$$

The following proposition justifies calling (2.5) a saddle-point problem.

Proposition 1. The functional $E(u, \lambda) = f(u) + \lambda(g(u) - c)$ is a convex-concave function.

Proof. By definition $u \mapsto E(u, \lambda)$ is convex for all λ, since f and g are convex, so their sum is convex as well. Finally, for fixed u the function $\lambda \mapsto E(u, \lambda)$ is linear (and thus concave) in λ. □

2.2.3 Duality

The concept of duality plays an important role in convex and constrained optimization. It allows to formulate approximations to hard problems and to convert constrained to equivalent unconstrained problems as was shown in the last section. We will use it frequently in this thesis. A fundamental definition in this context (see e.g. [87]) is

Definition 6. *Let* $\mathcal{U}$ be a real normed vector space, $\mathcal{U}^*$ its dual and $< ., . > : \mathcal{U} \times \mathcal{U}^* \to \mathbb{R}$ the bilinear function defined as $< u, \phi > \mapsto \phi(u)$. The *Legendre-Fenchel* dual $f^* : \mathcal{U}^* \to \mathbb{R}$ of a function $f : \mathcal{U} \to \mathbb{R}$ is given by

$$f^*(y) = \sup_x < x, y > - f(x) \tag{2.6}$$

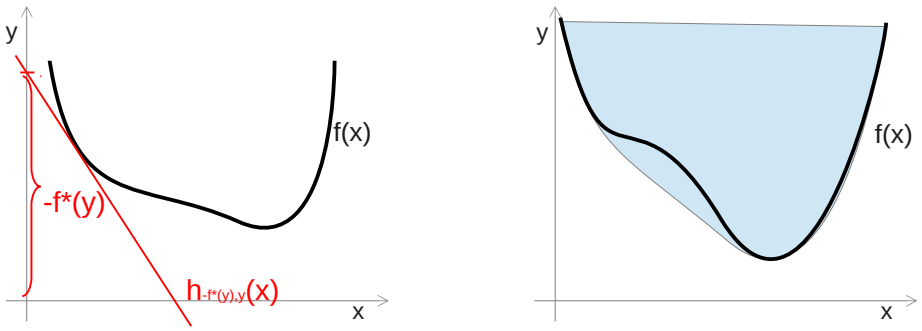


Figure 2.2: *Left*: let $h_{t,y}$ be an affine function with intercept t and slope y. The Legendre-Fenchel dual $f^*(y)$ of a function f at a point y describes the maximal negative intercept of all affine functions with slope y that lie below the graph of f. *Right*: an example of the convex envelope of a non-convex function f.

where $y \in \mathcal{U}^*$ and $\mathcal{U}^*$ is the dual space of the real normed vector space $\mathcal{U}$, i.e. the space of real-valued linear and continuous functions $f : \mathcal{U} \to \mathbb{R}$ with domain $\mathcal{U}$.

The Legendre-Fenchel dual value $f^*(y)$ can be interpreted as the largest intercept of an affine function with slope y that is below the function f (see Figure 2.2). Due to this, the dual f^* of a convex function f is convex as well. More importantly, a convex function f is the pointwise supremum of all affine functions that are below f. Therefore, it follows

Proposition 2. For a convex and lower semi-continuous function $f : \mathcal{U} \to \mathbb{R}$ it holds $f(x) = f^{**}(x)$.

Proof. Let $h_{y,t}$ be the affine function with slope y and y-intercept t. Then, since f is convex it holds by definition of f^*

$$f(x) = \sup_{h_{y,t} \leq f} h_{y,t}(x) = \sup_{y \in \mathcal{V}^*} < y, x > -f^*(y) = f^{**}(x)$$

where in the second equality we used the definition of the Legendre-Fenchel dual. □

In particular, for a *non-convex* function, f^{**} describes its convex envelope (see Figure 2.2). Finally, the dual of an indicator function $\mathbf{1}_{\mathcal{C}}(x)$ of a convex set $\mathcal{C}$ is equal to its support function, i.e. the convex function that describes the signed distances of supporting hyperplanes.

2.3 Convex Relaxations

Sometimes we are confronted with functionals that do not adhere to the definition of a convex functional simply because their domain is not a convex set - it is easy to see that in this case Definition 3 is not applicable. Such functions can be convexified by expanding their function domain.

An important example in the context of this thesis is given by functionals defined on the space of indicator functions (functions assigning a binary labeling to each point in a target domain) with bounded variation, denoted as

$$BV(\Omega; \{0, 1\}) \quad . \tag{2.7}$$

This set is obviously non-convex according to definition (2.3), as e.g. the linear convex combination of the two constant indicator functions $f_0 : x \in \Omega \mapsto 0$ and $f_1 : x \in \Omega \mapsto 1$ with $\lambda = 0.5$ is the function $f_{0.5} : x \mapsto 0.5$, which is not in the set (2.7). We can revert to the convex hull, i.e. the smallest convex set that contains all the points of the original set, of all indicator functions with bounded variation $BV(\Omega; [0, 1])$ in the functional definition and search for a minimizer of the convex, *relaxed* problem

$$\min_{u \in BV(\Omega;[0,1])} E(u). \tag{2.8}$$

Minimizing this convex problem is often much easier than optimizing the combinatorial counterpart. A relaxation is said to be *tight* if it is close to the original problem in the sense that its solution is a feasible solution to the original problem or that there is a tight bound to the original problem. An *a priori* bound for such a relaxation is given by the following relation. Let u_{opt} be the optimal solution of the original problem and u^r_{opt} the one of the relaxation. Then, since each solution of the binary problem is also feasible for the relaxed problem, it holds $E(u^r_{\text{opt}}) \leq E(u_{\text{opt}})$ and, thus, for any binary solution u_{bin} (which may be acquired by thresholding the relaxed optimum) it holds

$$E(u_{\text{bin}}) - E(u^r_{\text{opt}}) \leq E(u_{\text{bin}}) - E(u_{\text{opt}}) \tag{2.9}$$

which means that any optimal solution of the relaxed problem provides us with a bound on the optimal binary solution, which is more or less tight depending on the relaxation. Ideally, one can come up with a thresholding scheme that converts the relaxed optimal solution into a solution to the original problem which is provably optimal. Key ingredient to such a theorem is the co-area formula (2.32).

2.4 Variational Calculus

Variational Calculus deals with finding minimizers of functions with infinite dimensional domain. In this section a short overview of the basic theory is given as it will be used frequently in the chapters ahead. For a detailed introduction see e.g. [37].

2.4.1 Existence of Minimizers

Having explained the notion and benefits of convexity, the question naturally arises whether a well-defined minimum exists for a given convex functional. The conditions for a functional $f : \mathcal{U} \to \mathbb{R}$ defined on a Banach space $\mathcal{U}$ - i.e. a normed and complete vector space (see e.g. [73]) - to attain a minimum are the following

- f is lower semi-continuous, i.e. $epi(f)$ is a closed set.
- f is coercive, i.e. for any sequence $x_n \subset \mathcal{U}$ with $||x_n|| \to \infty$ it follows $\lim f(x_n) = \infty$

For saddle-point problems an analogue definition exists. A convex-concave functional E attains a saddle-point if

- $u \mapsto E(u, v)$ and $v \mapsto -E(u, v)$ are lower semi-continuous for all $u \in \mathcal{U}$ and $v \in \mathcal{V}$.
- $u \mapsto E(u, v)$ and $v \mapsto -E(u, v)$ are coercive for all $u \in \mathcal{U}$ and $v \in \mathcal{V}$.

2.4.2 Extremality Conditions

Finding the minimizer of a convex functional is the main problem we are concerned with. In the following necessary conditions for a local extremum of a functional are discussed. We are given a compact set $\Omega \subset \mathbb{R}^n$ with piecewise smooth boundary $\partial\Omega$ and a functional $E : \mathcal{C}^1(\Omega) \to \mathbb{R}$ of the form

$$E(u) = \int_\Omega \mathcal{L}(u, \nabla u, x)dx$$

where $\mathcal{L} : \mathbb{R} \times \mathbb{R}^n \times \Omega \to \mathbb{R}, (a, b, x) \mapsto \mathcal{L}(a, b, x)$ is a continuously differentiable functional called the Lagrangian. Similar to the finite dimensional case, a necessary condition for a local extremum for functions on infinite dimensional domains is that the directional derivative - given by the Gâteaux differential $\lim_{\epsilon\to 0} \frac{1}{\epsilon}(E(u + \epsilon v) - E(u))$ - has to vanish in each direction v. For a given functional $E(u)$, one can derive the following form of the Gâteaux differential

$$\int_\Omega dE(u)vdx = 0 \ , \tag{2.10}$$

where $dE(u)$ is the Fréchet derivative. According to the *fundamental lemma of variational calculus* this equation can only hold if $dE(u)$ equals zero for all $x \in \Omega$. This implies the *Euler-Lagrange equations*

$$\frac{\partial\mathcal{L}(u^*, \nabla u^*, x)}{\partial u} - \sum_i \frac{d}{dx_i}\frac{\partial\mathcal{L}(u^*, \nabla u^*, x)}{\nabla_i u} = 0 \quad \forall x \in \Omega \tag{2.11}$$

to hold for a local minimum u^*.

Optimality conditions for Saddle-points For completeness we will present the extremality condition for a saddle-point: $(\hat{u}, \hat{v})$ is a saddle-point of a convex-concave functional $E(u, v)$ if and only if

- $0 \in \partial_u E(\hat{u}, \hat{v})$ and
- $0 \in \partial_v E(\hat{u}, \hat{v})$

where $\partial_x E$ is the subgradient of E with respect to x.

2.5 Algorithms for Convex Optimization

Depending on its form different strategies for solving the partial differential equations (2.11) can be considered. In the linear case, iterative methods such as the *Gauss-Seidel* or *Jacobi* method can be applied. Closed form solutions, however, are very rare. Common iterative optimization approaches for non-linear cases will be reviewed in the following. The quality of the solution of each optimization method depends on the problem formulation.

2.5.1 Gradient Descent

Gradient descent is the most basic numerical scheme for solving equations (2.11) for convex functionals. Starting from an arbitrary initial solution u, in each iteration of the algorithm the current solution is modified in the direction of the negative gradient (or Fréchet derivative), i.e. the steepest descent of the energy functional:

$$u_{t+1} = u_t - \tau \cdot \frac{dE(u)}{du} \tag{2.12}$$

where τ is the step size. Naturally, τ has to be chosen carefully depending on the respective functional E as the numerical scheme will easily get unstable for too large step sizes.

2.5.2 Forward-Backward Splitting and FISTA

Gradient descent requires the convex functional E to be differentiable with respect to u. Forward-backward splitting [19] can be considered as a generalization of gradient descent and assumes that the objective function is a sum of a (proper) convex function $F : \mathcal{U} \to \mathbb{R}$ with a Lipschitz continuous derivative and a (proper) convex and lower semi-continuous but generally non-differentiable part $G : \mathcal{U} \to \mathbb{R}$ defined on a Banach space $\mathcal{U}$

$$\min_{u \in \mathcal{U}} F(u) + G(u)\,. \tag{2.13}$$

In each iteration the objective functional is approximated at the current point v by a quadratic functional that lies above it

$$F(u) + G(u) \leq Q_v(u) := F(v) + < u - v, \nabla F(v) > + \frac{L}{2} \|u - v\|^2 + G(u) \tag{2.14}$$

where the first three terms of Q_v yield a quadratic approximation of F and L is a Lipschitz constant. Letting go of $F(v)$ (which is constant since v is fixed) and completing the square by adding $\frac{1}{L} \|\nabla F(u)\|^2$ we arrive at the following *forward-backward splitting scheme* (for time step $t+1$)

$$u_{t+1} = \underset{u}{\operatorname{argmin}} \left\{ G(u) + \frac{L}{2} \left\| u - \left(u_t - \frac{1}{L} \nabla F(u_t) \right) \right\|^2 \right\} \quad . \tag{2.15}$$

Here, optimization over the sum of two functionals is *split* into two steps: the term in brackets is a gradient descent of function F, whereas the outer problem

$$(I + \frac{1}{L} \partial G)^{-1}(w) := \underset{u}{\operatorname{argmin}} \left\{ G(u) + \frac{L}{2} \|u - w\|^2 \right\} \tag{2.16}$$

can be interpreted as a subgradient descent step of function G. It is also called the *prox-operator* $\operatorname{prox}_{\frac{1}{L}G}(w)$. Notably, the prox-operator $\operatorname{prox}_{\chi_A}(x)$ of the characteristic function χ_A of a set A defined as

$$\chi_A(x) := \begin{cases} 0 & \text{if } x \in A \\ \infty & \text{otherwise} \end{cases} \tag{2.17}$$

is equal to the projection of x onto the set A.

Beck and Teboulle [14] extended the forward-backward splitting scheme to the *Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)*, which can be shown to converge quadratically. The iteration at time step t is given by

$$\begin{cases} u_t & = \operatorname{prox}_{\frac{1}{L}G} \left(\bar{u}_t - \frac{1}{L} \nabla F(\bar{u}_t) \right) \\ \theta_{t+1} & = \frac{1}{2} \left(1 + \sqrt{1 + 4\theta_t^2} \right) \\ \bar{u}_{t+1} & = u_t + \frac{\theta_t - 1}{\theta_{t+1}} (u_t - u_{t-1}) \quad . \end{cases}$$

The first operation is the forward-backward splitting step (2.15). Step two and three comprise an over-shooting scheme extrapolating the current solution by a factor θ (step three) which is adapted in each step (step two).

2.5.3 First Order Primal-Dual Optimization

In [21] Chambolle and Pock proposed a numerical scheme for optimizing convex problems of the form

$$\inf_{u \in \mathcal{U}} F(Ku) + G(u) \tag{2.18}$$

where $\mathcal{U}, \mathcal{V}$ are reflexive Banach spaces (those which can be identified with its bidual) and $G : \mathcal{U} \to \mathbb{R}$ and $F : \mathcal{V} \to \mathbb{R}$ are convex, proper (i.e. $< \infty$) and semi-continuous (but not necessarily differentiable) functions and $K : \mathcal{U} \to \mathcal{V}$ is a linear operator. The primal-dual algorithm is a numerical scheme for solving the primal-dual version of problem (2.18)

$$\inf_{u \in \mathcal{U}} \sup_{v \in \mathcal{V}} < Ku, v > + G(u) - F^*(v) \ . \tag{2.19}$$

Here, F^* denotes the Legendre-Fenchel dual of F. To see that (2.18) and (2.19) describe the same problem, we make use of the Legendre-Fenchel dual and the fact that since F is lower semi-continuous and convex it holds that $F^{**}(w) = \sup_{v \in \mathcal{V}} < v, w > - F^*(v) = F(w)$:

$$\inf_{u \in \mathcal{U}} F(Ku) + G(u) = \sup_{v \in \mathcal{V}} \inf_{u \in \mathcal{U}} < Ku, v > - F^*(v) + G(u) \ . \tag{2.20}$$

Note that the primal-dual is computationally easier to solve than the primal (2.18) in the sense that G and F now depend on two different variables. This splitting idea will reappear in the ADMM algorithm below.

The numerical scheme of the primal-dual algorithm for solving the saddle-point problem (2.19) consists of a gradient descent in the primal variable, a gradient ascent in the dual variable and an over-relaxation step which extrapolates the current solution in the direction of the solution change by a factor $\theta \in [0, 1]$:

$$\begin{cases} v^{t+1} & = (I + \Sigma \partial F^*)^{-1} \Big(v^t + \Sigma K \bar{u}^t \Big) \\ u^{t+1} & = (I + T \partial G)^{-1} \Big(u^t - T K^* v^{t+1} \Big) \\ \bar{u}^{t+1} & = u^{t+1} + \theta (u^{t+1} - u^t) \end{cases} \tag{2.21}$$

where Σ and T are preconditioning matrices. The scheme is iterated until convergence. There are different convergence criteria. One can measure the difference in energy of the primal and the dual problem formulation (*primal-dual gap*) and stop when it falls below a threshold. However, depending on the given problem the computation of the dual energy can be as difficult as solving the primal problem. Alternatively, one can iterate until the relative change in the solution falls below a threshold. This will be detailed in the chapters to follow. The numerical scheme (2.21) is provably convergent and can be applied to several interesting problems in computer vision that can be cast in the form (2.18).

2.5.4 Alternate Direction Method of Multipliers

Another way of optimizing problems of the form (2.18) is the Alternate Direction Method of Multipliers (ADMM) [36]. For simplicity, we will stick to the finite dimensional case, i.e. $\mathcal{U} \subset \mathbb{R}^n$ and $\mathcal{V} \subset \mathbb{R}^m$ and $K \in \mathbb{R}^{n \times m}$. The algorithm is derived by performing a variable splitting

$$\min_{u \in \mathcal{U}, v \in \mathcal{V}} F(v) + G(u) \quad \text{s.t. } Ku = v \tag{2.22}$$

and formulating the augmented Lagrangian of this problem:

$$L_\tau(u, v, \lambda) = F(v) + G(u) + \lambda(Ku - v) + \frac{\tau}{2} \|Ku - v\|^2 \ . \tag{2.23}$$

The last summand can be understood as a quadratic penalty term for the constraint $Ku = v$. The ADMM algorithm alternately minimizes the augmented Lagrangian and then does a gradient ascent step on the Lagrange Multiplier λ:

$$\begin{cases} u^{k+1} & = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} G(u) + \lambda(Ku - v^k) + \frac{\tau}{2} \left\|Ku - v^k\right\|^2 \\ v^{k+1} & = \underset{v \in \mathbb{R}^m}{\operatorname{argmin}} F(v) + \lambda(Ku^{k+1} - v) + \frac{\tau}{2} \left\|Ku^{k+1} - v\right\|^2 \\ \lambda^{k+1} & = \lambda^k + \tau(Ku^{k+1} - v^{k+1}) \end{cases} \tag{2.24}$$

where $\tau > 0$ is the step size parameter. One can pull linear term into the squared norm by substituting $\lambda = \frac{1}{\tau}\lambda$. Furthermore, an over-relaxation step can be introduced. Together this results in the following scheme, which will be used in later chapters

$$\begin{cases} u^{t+1} = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} G(u) + \frac{\tau}{2} \left\|Ku - v^t + \lambda^t\right\|^2 \\ r^{t+1} = \alpha K u^{t+1} + (1 - \alpha) v^t \\ v^{t+1} = \underset{v \in \mathbb{R}^m}{\operatorname{argmin}} F(v) + \frac{\tau}{2} \left\|r^{t+1} - v + \lambda^t\right\|^2 \\ \lambda^{t+1} = \lambda^t + r^{t+1} - v^{t+1} \end{cases} \tag{2.25}$$

where $\alpha \in (0, 2)$ is the over-relaxation parameter.

Similar to the primal-dual algorithm, the ADMM method allows for a problem decomposition owing to the variable splitting in (2.22). This can be used to separate projections on intersections of convex sets that cannot be solved in closed form. This is a major advantage of the ADMM algorithm. Note that ADMM and the primal-dual algorithm are closely related as shown in [21]. However, practically both algorithms differ in memory consumption and runtime.

A convergence criterion for scheme (2.25) is to measure the relative change in the solution and stop if it falls below a given threshold. This strategy will be detailed in the following chapters.

2.6 Total Variation

Many convex approaches to computer vision problems incorporate the *total variation norm* as a smoothness prior term. The total variation plays an important role throughout this thesis for its geometric and convex properties and will be introduced in the following.

2.6.1 Definition

Consider the following

Definition 7. The weighted total variation norm for a functional $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is defined as

$$TV_g(u) = \int_\Omega g\,|Du| = \sup\left\{ -\int_\Omega u \operatorname{div}\xi\, dx \;\middle|\; \xi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^n)\,,\;\; \|\xi\|_\infty \leq g \right\}, \tag{2.26}$$

where Du is the derivative of u in the distributional sense, $g : \Omega \to \mathbb{R}_0^+$ is a weighting function, $\|\xi\|_\infty$ is the uniform norm and C_c^1 denotes the space of smooth functions with compact support.

This definition is valid for any locally L^1-integrable function (including indicator functions of sets with measurable boundary). Several favorable properties make the TV norm well suited as a regularizer in convex optimization problems: it is a convex, lower semi-continuous function and - in contrast to a quadratic regularizer - well preserves discontinuities in the solution. Intuitively, the reason for this lies in the fact that the total variation of a function that quickly jumps from zero to one is the same as that of a function which goes from zero to one in small steps. In case the function u is differentiable, $TV(u)$ is equal to $\int_\Omega |\nabla u| dx$. The right side of Equation 2.26 can be considered the bidual of that expression.

Minimizers of TV-based functionals live in the space of bounded variation, which is defined as all locally integrable functions u that have a finite total variation [11]:

$$BV(\Omega;\mathbb{R}) := \left\{ u \in \mathcal{L}^1_{\text{loc}}(\Omega) \;\middle|\; TV(u) < \infty \right\}\,. \tag{2.27}$$

This defines a Banach space with respect to the norm

$$\|u\|_{L^1(\Omega)} + TV(u) \tag{2.28}$$

(where $TV(u)$ denotes the unweighted version of the total variation). The TV-norm is not coercive. However, typical TV-based functionals in computer vision usually incorporate a coercive data affiliation term making the functional coercive and guaranteeing the existence of minimizers.

2.6.2 Geometric Properties

The TV norm can be used to measure the perimeter of point sets. This property is the basis for convex formulations of minimal partition problems. It will also be the basis for most of the minimal surface approaches proposed in this thesis. The details are given in

Proposition 3. Let $\mathbf{1}_{\mathcal{S}}$ be the indicator function of a set $\mathcal{S} \subset \Omega$ with smooth boundary. Then the TV-norm of $\mathbf{1}_{\mathcal{S}}$ describes the perimeter of $\mathcal{S}$ measured in the metric defined through the non-negative function $g : \Omega \to \mathbb{R}_0^+$:

$$\mathrm{Per}_g(\mathcal{S}, \Omega) = TV_g(\mathbf{1}_{\mathcal{S}}) \tag{2.29}$$

Proof. By the definition in (2.26), $TV_g(\mathbf{1}_{\mathcal{S}})$ is equal to

$$\sup_{\xi:|\xi|_\infty \leq 1} \left\{ -\int_\Omega \mathbf{1}_{\mathcal{S}} \operatorname{div}(\xi)\, dx \right\} = \sup_{\xi:|\xi|_\infty \leq 1} \left\{ -\int_{\mathcal{S}} \operatorname{div}(\xi)\, dx \right\} \tag{2.30}$$

$$= \sup_{\xi:|\xi|_\infty \leq 1} \left\{ \int_{\partial\mathcal{S}} n \cdot \xi \, ds \right\} = \int_{\partial\mathcal{S}} ds \tag{2.31}$$

where in the second equality the theorem of Gauss was applied. □

Given this relation one can express the TV norm of a functional u by the perimeter of its upper level sets. This is formalized by

Proposition 4. The *co-area formula* states that

$$TV(u) = \int_{-\infty}^{\infty} TV(\mathbf{1}_{u \geq t})\, dt \tag{2.32}$$

where

$$\mathbf{1}_{u \geq t}(x) := \begin{cases} 1 & \text{if } u(x) \geq t \\ 0 & \text{else} \end{cases}. \tag{2.33}$$

A proof can be found in [38].

3 Literature Overview

Estimating 3D geometry from images has been a core research topic in computer vision for several decades. For the case of multiple input images a large variety of methods has been developed which are able to deliver exact high quality reconstruction results. For the special case that only a single still image is available the problem gets considerably more difficult. In the literature there exists a variety of methods to estimate 3D geometry that are tailored for specific classes of input images. However, a thorough comparison between the approaches has not been carried out so far.

The reason for this lies mainly in the inherent ill-posedness of the problem: during image formation, depth is irrecoverably lost. In their effort to make the problem tractable, single view methods have come up with an abundance of very different assumptions, methods and priors to infer the geometry of a depicted scene or object. This geometric information can be of different nature reaching from purely relational information, sparse measurements or dense depth information to a complete 3D model of a single object or even a scene. The reconstruction precision of such approaches exceeds that of plausible estimates only in very few cases. Consequently, the reconstruction objectives are of very different nature, which makes a comparison difficult.

In this chapter we give a brief survey on the subject of single view 3D reconstruction, which was published in [6]. We provide an introduction to the field and examine basic image information and assumptions that are used in order to compensate for ill-posedness. We then review, categorize and compare existing state-of-the-art approaches.

To this end, in the following we will first look into typical image cues and priors that are used in the literature for inferring geometric information from a single image. We will then present and categorize relevant works in the field. Finally, in Section 3.3 a systematic comparison of the presented methods is undertaken.

3.1 Image Cues and Priors

In this section we firstly provide an overview on the different types of image information ('*image cues*') used in the literature. Secondly, we will review the *priors* that are assumed in order to overcome the ill-posedness. This serves as a survey on related work.

3.1.1 Types of Image Information

Approaches to single view reconstruction extract higher or lower level information from the input image in order to infer geometric information. This is done automatically or with the help of user input. In the following we list the most important categories of image information and give prominent references.

Shading. The problem of *shape-from-shading* (SfS) is to infer a surface (height field) from a single gray level image by observing the gradual variation of shading that is induced by the interaction of surface and light. Some approaches also co-estimate lighting conditions and reflection properties. In general, the goal is to find a solution to the following image formation model

$$R(n(x)) = I(x) \ , \tag{3.1}$$

where I is the image, n is the normal field of the surface and R is the reflectance function which is dependent on the object. In most SfS approaches a Lambertian reflectance model of constant albedo is assumed. But there are other models which also consider specular materials (e.g. Wang et al. [105]). SfS is an ill-posed problem, although there has been progress on deriving conditions for unique solutions by Prados and Faugeras [83].

As shown by Durou et al. [35] and Zhang et al. [115] reconstruction from real world images is limited in the sense that each approach exhibits special and sometimes unrealistic requirements on lighting conditions or reflection properties. Especially the presence of texture is an issue. Work has been done, however, to incorporate interreflection [77], shadowing and perspective projection [29] just to name a few. One of the first minimization approaches to SfS is by Ikeuchi and Horn [54]. For a current survey see Durou et al. [35].

Shadow. The shadow that is thrown by objects conveys geometric information relative to the viewpoint of the light source. Often point light sources are assumed as soft shadows do not provide high frequency information. Furthermore, shadow is not always thrown on known geometry, which makes the problem harder. References include works by Daum and Dudek [32], Kender and Smith [60], Yu and Chang [109] and Hatzitheodorou [45].

Contour Edges. Contour edges are salient structures in the image that are induced by surface discontinuities, occlusion, object boundaries or reflectance changes. They give evidence for geometry and relative pose/position of objects. *Junction points* or *corners*, where multiple contour edges meet or end, are also important cues for single view reconstruction.

Subclasses of contour edge-based methods are contour-based and silhouette-based reconstruction methods. *Shape-from-contour* approaches try to infer geometry given the object contours alone. Approaches based on closed contour drawings include Horaud et al. [48], Ulupinar et al. [101] and Li et al. [68]. Karpenko et al. [58, 59] interpret user line drawings. Other single view reconstruction approaches that use contour edges for reconstruction include [33, 43, 64, 46, 91, 92].

Silhouette. Closely related to shape-from-contour are approaches that infer geometry given the object silhouette. The silhouette is the image of the contour generator plus its interior and the contour generator is the set of visible points on a surface, whose image rays are tangent to the surface. Silhouette based approaches find reconstructions, whose projection into the image plane agrees with the silhouette and whose normals agree with those of the contour generator. As there are always infinitely many objects that are silhouette consistent this cue suffers from inherent ambiguity if used alone.

References for silhouette based reconstruction algorithms include Prasad et al. [85, 86] and the approaches proposed in this thesis. Closely related are sketch based modeling tools such as Igarashi et al. [53], Karpenko et al. [59] and Nealen et al. [78].

Texture. The local albedo, also known as diffuse reflectivity or simply *color*, is a local property of the object material. The variation of the albedo along surface location is called *texture*. In contrast to shading information, texture is considered an inherent property of the object surface rather than a result of an interaction of light and geometry. A complete separation of shading and texture information is a hard problem (a problem also known as *intrinsic images*).

If one assumes objects to have a known or regular texture it is possible to infer their geometry from the way the texture is deformed after image projection. These *shape-from-texture* approaches, obviously, impose strong constraints on the reconstructable objects. An example constitutes the work of Malik and Rosenholtz [74].

Further single view 3D reconstruction algorithms that use texture cues include Super et al. [99], Hassner and Basri [44] and Vetter et al. [103]. Approaches that combine texture and contour edges for reconstruction by considering so-called 'superpixels' are Hoiem et al. [46] and Saxena et al. [92].

Defocus. Due to physical aspects of image formation, the sharpness of the image of an object correlates with its distance from the camera. This fact can be used to infer a dense depth map from an image. However, the accuracy of such methods is limited and camera calibration is required. References include works from Levin [67], Zhou et al. [116] and Bae and Durand [12].

Location. Relative spacial location of objects in the image can be used to infer semantic knowledge. For example, ground, floor or sky can be identified more easily from their location in the image. This information can be helpful for 3D reconstructions. Hoiem et al. [46] reconstruct vertical objects by distinguishing them from the ground and the sky. Delage et al. [33] use a Bayesian network to identify floor pixels.

3.1.2 Priors

Priors are of utter importance in single view 3D reconstruction to constraint the solution space of reconstructions. They are either applied a priori, or they are learned from sample data. Furthermore, there are low-level and high-level priors. In the following we will list priors that are most frequently assumed in single view 3D reconstruction.

Smoothness. Smoothness can be defined as the small spatial change of some property. In single view reconstruction we are often not able to infer a dense reconstruction. It is therefore good practice to choose among the possible reconstruction surfaces those which tend to be smooth. Smoothness naturally plays a significant role in the reconstruction of curved surfaces as in [114, 86].

Smoothness in a wider sense can also be learned as the consistency of object surfaces. Hoiem et al. [46] use a machine learning approach to find image features indicating the assignment of neighboring superpixels to the same object. Saxena et al. [92] use image cues and geometric relations to learn the relative depth of neighboring superpixels.

Geometric Relations. Basic geometric relations are often encountered specifically in man-made environments. One can consider e.g. *coplanarity*, *collinearity*, *perpendicularity* and *symmetry*. An early work which makes use of such simple rules is the one of Lowe [72]. By assuming planes to be parallel or perpendicular one can also derive camera parameters (see Criminisi et al. [31]). Objects are often assumed to stand vertically on the ground plane [46, 33, 43]. Symmetric objects consist of self-similar parts, which are projected to different locations in the input image. This constellation is, thus, similar to having multiple views of the object [47, 43].

Volume / Area. Single view approaches that search for smooth reconstructions often suffer from flat solutions. Requiring the surface to enclose a specific volume is a means of inflating the reconstructions. The area of the object silhouette can be used to estimate the object volume. Volume priors can be found in Li et al. [68] and in the proposed methods in this thesis.

Semantic Relations. Semantic relation priors assume high-level knowledge on the relative position and inner structure of different objects and their depth values. Han and

Zhu [43], e.g., infer occluded points based on semantic human knowledge, for instance that leaves are connected to the plant. Koutsourakis et al. [64] introduce semantic knowledge to ensure the consistency of different floors. Finally, knowledge on the location of the ground and the sky represents an important cue for 3D reconstruction. The ground is often used as starting point for the reconstruction as objects, especially walls, are usually perpendicular to this plane [43, 33, 46].

Shape Priors. Shape priors impose low-level or high-level knowledge on the reconstruction shape. Shape priors can be defined or learned. In [64], Koutsourakis et al. define a full shape grammar for the reconstruction of facades. This limits the approach to the reconstruction of buildings in urban environments. In contrast, Rother and Sapiro [89] and Chen and Cipolla [25] shape priors are learned from a database of sample objects. Hence, they are not a priori limited to a specific object class. However, their choice of samples intrinsically limits their approach to the object classes represented in the database. Silhouette priors can be regarded as a form of shape prior. Cremers et al. [30] introduced shape priors in the context of continuous level set segmentation.

The representation of shape priors ranges from specified sets of grammar rules over parametric models to probabilistic priors. In [25], Chen and Cipolla learn depth maps of human bodies by means of principal component analysis. This model imposes strong assumptions on the 3D object, but the dimension of the state space is reduced and only valid 3D reconstructions are obtained. In contrast, Rother and Sapiro [89] impose less strong assumptions on the learned model. For each object class a shape prior is learned as the relative occupancy frequency of each voxel in the object.

3.2 A Classification of Approaches to Single View 3D Reconstruction

In this section a taxonomy of single view 3D reconstruction approaches is developed. From the diversity of single view related reconstruction methods, we consider only those algorithms that focus on real world images that are not taken under controlled environments or with strong constraints on object material or lighting conditions. The selection, furthermore, focuses on works that are representative and state-of-the-art.

The following classification stems from the observation that the type of a single view approach strongly depends on its application domain, i.e. the set of objects and scenes, which it is designed for. Four categories are proposed, each named after the type of the specific reconstruction domain:

- Curved Objects
- Piecewise Planar Objects

- Learning Specific Objects
- 3D Impression from Scenes

Approaches that reconstruct *curved objects* principally aim at producing arbitrary, mostly smooth objects. Often minimal surface priors are used, which try to minimize the surface of the object given a silhouette or similar. The second class consists of methods that concentrate on *piecewise planar objects* such as buildings and man-made environments. Furthermore, we distinguish arbitrary curved and planar objects from *learning specific objects*. Approaches in this class cannot reconstruct arbitrary objects, but are inherently limited to specific object classes by shape information learned from sample databases. Finally, we discuss methods that do not aim to reconstruct exact or plausible 3D geometry but rather provide a pleasing *3D Impression from Scenes*.

In the following we will review relevant single view 3D reconstruction methods that represent each of the class well. Particular emphasis will be put on methods that are closely related or important to the approaches presented in this thesis.

3.2.1 Curved Objects

Zhang et al.

Zhang et al. [114] proposed a method for interactive depth map editing based on an input image. The depth map reconstruction is the result of minimizing a thin plate energy [34], which favors smooth surfaces and penalizes bending. User input comes as a variety of constraints on the thin plate energy which are applied interactively to the depth map. These comprise of position constraints, surface normals, surface or normal discontinuities, planar region constraints or curves along which curvature or torsion is minimized.

The mathematical formalism is based on the thin plate energy for a continuous function f on a two dimensional rectangular domain $\Omega = [0,1]^2 \subset \mathbb{R}^2$ is defined as:

$$E(f) = \int_0^1 \int_0^1 \left[\alpha(u,v) \left\| \frac{\partial^2 f}{\partial u^2} \right\|^2 + 2\beta(u,v) \left\| \frac{\partial^2 f}{\partial u \partial v} \right\|^2 + \gamma(u,v) \left\| \frac{\partial^2 f}{\partial v^2} \right\|^2 \right] du\, dv \ , \quad (3.2)$$

where functions $\alpha, \beta, \gamma : [0,1]^2 \mapsto \{0,1\}$ are weights that can be used to define local surface discontinuities. Zhang et al. [114] discretize this minimization problem by introducing a function $g_{i,j}$ that samples values of the depth map function $f : [0,1]^2 \mapsto \mathbb{R}$ on a discrete rectangular grid, that is, $g_{i,j} = f(id, jd)$, with d being the distance between neighboring grid points. For efficiency and accuracy the grid resolution can be locally refined by the user. All values $g_{i,j}$ can stacked into a vector $\mathbf{g}$ and after discretizing the partial derivatives in (3.2), the energy can be written as a quadratic form

$$\mathbf{g}^T \mathbf{C} \mathbf{g} \qquad \text{subject to} \quad \mathbf{A}\mathbf{g} = \mathbf{b} \ , \quad (3.3)$$

where $\mathbf{Ag} = \mathbf{b}$ models the linear constraints on the surface mentioned above. See [114] for a detailed description on how the constraints are incorporated into this quadratic optimization problem. A description of these constraints from the user's point of view is given later together with the experimental comparison in Section 6.5.

After introducing a Lagrange multiplier λ for the constraints the extremality condition of the optimization problem yields a sparse linear system

$$\begin{bmatrix} \mathbf{C} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \tag{3.4}$$

which can be computed efficiently with standard solvers.

Prasad et al.

The works [86] and [85] of Prasad et al. introduce a framework for single view 3D reconstruction of curved surfaces. The method is closely related to the one by Zhang et al. [114].

The main idea is to compute a parametric minimal surface by globally minimizing the thin plate energy (Equation (3.2)). However, in contrast to Zhang et al. they minimize the energy with respect to a parametrized 3D surface $f : [0,1]^2 \mapsto \mathbb{R}^3$ instead of a depth map. Consequently, function domain and image domain are no longer equivalent. The discretization of the optimization problem with constraints is done similarly to Zhang et al. (see Equation (3.4)).

The choice of constraints is mostly different from Zhang et al. [114]. The main source of reconstruction information is the silhouette: Prasad et al. [86] use the fact that surface normals can be inferred at points projecting to the contour generator $c(t)$, since at these points the normals are parallel to the image plane and have the same direction as the normal at $c(t)$. This can be formalized by the following constraints

$$\pi(f(u(t), v(t))) = c(t) \tag{3.5}$$

$$n(c(t))f(u(t), v(t)) = 0 \qquad \forall t \in [0,1] \ , \tag{3.6}$$

where $n(c(t))$ is the normal at the point $c(t)$ in $\mathbb{R}^3$ and π the orthographic projection function. The user has to determine the coordinates $(u(t), v(t))$ of the contour generator in parameter space. This is done by placing lines onto the grid of the parameter space and setting them in correspondence with the parts of the contour generator. Similar to Zhang et al. [114] the user can employ position constraints to define the object inflation locally. Also, surface discontinuities can be optionally specified to relax the surface smoothness along curves in the parameter space. Importantly, in order to define the topology of the object, the user has to define which parts of the parameter space boundary are connected. For example, the connection of the left and right boundary defines a cylindrical shape of the function domain.

The optimization of minimal surface problem (3.2) is done similarly to the method by Zhang et al. (see above). More details on the user input is given later in Section 6.5 of Chapter 6.

Other Approaches

Another subclass of 3D reconstruction algorithms for curved objects are based on surfaces of revolution (SORs) [108, 102, 27]. They are common in man-made objects and represent a subclass of Straight Homogeneous Generalized Cylinders. SOR approaches strongly rely on the assumption of rotational symmetry of the objects. In Colombo et al. [27], the task of 3D reconstruction is formulated as the problem of determining the meridian curve from the imaged object silhouette and two given imaged cross sections. Based on the computation of fixed entities such as the vanishing line or the SOR symmetry axis, camera calibration can be done and the SOR is inferred.

Francois and Medioni [39] present an interactive 3D reconstruction method based on user labeled edges and curves, which are represented by non-uniform rational basis splines (NURBS). The reconstructed objects are either modeled as generalized cylinders or as a set of 3D surfaces. Terzopoulos et al. [100] propose deformable elastic 3D shape models, which evolve around a symmetry axis and whose projection into the image is attracted by strong image gradients. Cohen and Cohen [26] propose a generalization of snakes to 3D objects based on a sequence of 2D contour models for medical images.

3.2.2 Piecewise Planar Objects and Scenes

Approaches that reconstruct piecewise planar objects and scenes use polygonal representations. Accordingly, the algorithms are based on the assumption of an idealized piece-wise planar world. In the following we give a short overview over a selection of relevant works. This will be less detailed as for the curved objects, as the piece-wise planar case is less relevant to the reconstruction approaches proposed in this thesis.

Criminisi et al. [31] describe how exact 3D affine measurements can be obtained from a single perspective view. They assume that a vanishing line of a reference plane as well as a vanishing point for a direction not parallel to the plane can be computed from the image. Given those and a known reference length, the authors derive affine scene structure from the image. Measurements between and on planes parallel to the reference plane are obtained from cross-ratios and specific image mappings. The approach is applied among others to 3D reconstruction of idealized piecewise planar scenes.

Delage et al. [33] describe an approach for the automatic reconstruction of 3D indoor scenes. Strong assumptions are made: the objects in the scene are composed of orthogonal planes and edges and the calibration of the vertical camera is available. Delage et al. show that given a segmentation of pixels into planes and edges together with respective orientations (assuming no occlusion edges are present) unique 3D geometry

can be inferred for the whole image. For the segmentation, a Markov Random Field is constructed. Image features, priors and statistics are used as data terms. Finally, to obtain a 3D reconstruction from the MRF labeling, a constrained iterative optimization problem is formulated, a generalization of Sturm and Maybanks work [98].

Another very different approach is taken by Koutsourakis et al. in [64]. They generate urban 3D reconstructions from single images by estimating the parameters of a 3D shape grammar with the help of a Markov Random Field, so that the generated building best matches the image. The main advantages of using a shape grammar are that it always produces well-defined buildings and that the complexity of the optimization as well as the dimensionality of the problem is strongly reduced. In the MRF, the unary terms ensure that object boundaries coincide with image boundaries, whereas the pairwise terms measure the appropriateness of the configuration of atomic shapes and ensure the consistency between the operator and the image.

Other approaches include the following: Kanade [57] recovers shape from geometric assumptions. The world is modeled as a collection of plane surfaces, which allows for a qualitative object recovery. Quantitative recovery is achieved by mapping image regularities into shape constraints. Piecewise planar scenes are computed in Liebowitz et al. [70] based on camera and geometric constraints such as parallelism and orthogonality, e.g. for the reconstruction of buildings.

Apart from symmetry and planarity, two additional shape constraints are introduced by Li et al. [68] for object reconstruction: maximum compactness and minimum surface. Instead of computing vanishing lines, Kushal et al. [65] perform 3D reconstruction of structured scenes by registering two user indicated world planes. Hong et al. [47] study the relation between symmetry of objects and the viewer's relative pose to the object. An important principle for the reconstruction of symmetric objects is that one image of a symmetric object is equivalent to multiple images. Li et al. [69] describe a method for reconstructing piecewise planar objects by using connectivity and perspective symmetry of objects.

3.2.3 Learning Specific Objects

The approaches in this category learn the appearance, structure or shape of certain object classes. As a consequence, reconstructing objects outside the learned classes is likely to fail with these methods. A direct comparison of results with the proposed methods in this thesis is fruitless, as approaches that rely on learning are mostly not general purpose methods in a stronger sense.

Han and Zhu [43] propose a 3D reconstruction approach based on manually defined shape priors, which can on the one hand be applied to polyhedral objects and on the other hand to grass and tree-like objects. They argue that learning priors is hard in practice, because there is not enough real world training data available. They manually define priors on planarity of faces as well as similarity in angles or edge lengths or

smooth and evenly spread curves for tree-like structures. The input image as well as the 3D scene are both represented as graphs and in a Bayesian approach the 3D scene graph is determined given the image graph.

Rother and Sapiro [89] present a general framework for pose estimation, 2D segmentation, object recognition and 3D reconstruction from a single image based on learning specific object classes. They assume to have a segmentation of the object and a corresponding color model of foreground and background. An implicit volumetric representation is chosen and a shape prior is defined by learning voxel occupancies for specific objects. A reconstruction is inferred from a probabilistic model that is based on the shape prior and a likelihood function taking into account a color model of the object and the probability of obtaining a particular pixel state (foreground or background) based on the number of full voxels projecting onto a pixel.

Chen and Cipolla [25] propose to infer 3D information directly from learned shape priors. They assume a number of given training shapes each consisting of the silhouette and the corresponding depth map. Principal component analysis is applied to the registered training data to obtain feature pairs consisting of position and depth information and learned via a Gaussian Process. Given an unknown registered silhouette, 3D information can then be inferred from the learned shape model by projecting it into the PCA subspace and asking for the most likely depth estimate at each point which is determined via an iterative optimization scheme.

Hassner and Basri [44] aim at depth reconstruction from a single image based on examples. The samples are given in a database and for each image patch centered on a pixel its depth is inferred from known depth values of the most similar patches in the database by maximizing its plausibility. The image patches overlap leading to several depth estimates for each pixel. These are combined by averaging. To ensure consistency of neighboring patches a global optimization procedure is proposed which iteratively refines depth estimates.

Other approaches that are worth mentioning include the following. Vetter [103] learned a parametric model for the reconstruction of faces by applying PCA to a database of registered 3D faces. Then the model parameters can be found, which best explain the given image of a face. In Nagai et al. [76], objects are learned from a sample database. A Hidden Markov Model is used to model the correspondence between intensity and depth.

3.2.4 3D Impression from Scenes

Approaches in this category differ from the other reviewed reconstruction methods mainly in two aspects: they focus on whole scenes instead of single objects and they put emphasis more on pleasant looking rather than exact or plausible reconstructions. Also, none of the approaches reconstruct closed geometry.

In [46], Hoiem et al. propose a fully automatic approach for creating 3D models

similar to pop-up illustrations in children's books. They divide the world into ground, sky and vertical objects. The appearance of these classes is described by image cues, which are learned from sample images. The image is segmented into superpixels, which are grouped to constellations by a probability model learned from training data. From the computed constellations a polygonal 3D modal is constructed.

In [92], Saxena et al. propose another approach for obtaining 3D structure from a single image of an unstructured environment. The only assumption the authors make is that the world consists of small planes, whose 3D position and orientation is to be estimated. Similar to Hoiem et al. [46], the authors start out from a superpixel segmentation of the image. But instead of grouping the superpixels into constellations, depth and orientation for each of them is inferred. This is done by a Markov Random Field (MRF) model whose parameters are learned from training data. The data term of the model comprises of image features such as color and texture and the pairwise term encompasses coplanarity, connectedness and collinearity.

Other works in this category include the following: in Horry et al. [50], simple 3D scenes are reconstructed based on user input such as vanishing points and foreground objects. The background of the scene is then modeled by rectangles, the foreground by hierarchical polygons. Barinova et al. [13] propose a reconstruction approach for urban scenes yielding visually pleasant results. The method is based on fitting 3D models containing vertical walls and ground plane to the scene.

3.3 Properties and Comparison of Related Works

After categorizing and reviewing the single view 3D reconstruction literature, the surveyed approaches will be summarized and compared with respect to several categories. This is done in Table 3.1, which also indicates image cues and shape priors leveraged in each approach. Note that in general the given categorizations and assigned properties are sometimes ambiguous and overlap. E.g., the approach by Rother and Sapiro [89] can reconstruct curved objects, but is listed under the *learning* category. The works listed as Toeppe et al. and Oswald et al. represent the approaches proposed in this thesis. In the following we will explain the table.

Category, Assumptions and Precision. The different reconstruction **methods** are grouped into the four **categories** of Section 3.2 (see first column of Table 3.1), that represent basic reconstruction domains.

Characteristic to each approach are the **assumptions** made. If specific assumptions are not met, the reconstruction process easily fails. Assumptions for each method are summarized in the second row of Table 3.1. Typical assumptions are a calibrated camera [33], a simplified scene composition [33, 46], an object database containing samples for learning shape priors [25, 89], a specific viewpoint [86], [5, 9] or given geometric properties

	Category	Curved Objects										Piecewise Planar										Learning Specific Objects										Scenes									
	Method	Prasad et al. [86]			Zhang et al. [114]		Oswald et al.[5]		Töppe et al. [9]		Colombo et al.[27]		Criminisi et al. [31]			Delage et al. [33]			Koutsourakis et al.[64]			Han & Zhu [43]			Rother & Sapiro [89]			Chen & Cipolla [25]			Hassner & Basri [44]			Hoiem et al. [46]			Saxena et al. [92]		
	Assumptions	characteristic sideview, max. genus 2			none		sideview, symmetry		sideview, symmetry		rotational symmetry		vanishing line, refer. height, ground plane			calibration, Manhattan			rectified image, buildings			polyhedra, plants			calibration, color models, database			database			fixed view, spec. object database			simple scene: sky, vertical walls&ground			world consists of planes		
	Precision	≥			≥		≥		≥		≥		=			=			≥			≥			≥			≥			≥			≈			≈		
	Surface Representation	[closed] parametric			depth map		closed implicit		closed implicit		closed parametric		non-closed polygonal			non-closed polygonal			closed polygonal			closed polygonal			closed implicit			depth map			depth map			pw. planar depth map			pw. planar		
	User Input	contours, [creases]			constraints		silhouette, [creases], [data term]		silhouette, [creases], [volume]		silhouette, cross sec.		all edges to be measured			none			none			none			none			silhouette			none			none			none			
Image Cues	Silhouette	x					x		x		x																	x												
	Edges												x			x			x			x													x			x		
	Location															x			x						x										x			x		
	Texture	x														x			x											x						x			x	
Priors	Smoothness	x			x		x		x							x			x			x							x						L			L	
	Volume/Area								x																															
	Semantic																		x			x																	x	
	Geometric										x		x									L														x				
	Shape												learned grammar									probabilistic			learned voxel model			learned PCA model			learned example based								

Table 3.1: Comparison of single view methods: for each approach the most important characteristics are indicated (see text). The 'L' indicates learned priors, terms in brackets are optional.

such as vanishing lines of reference planes [31].

Closely related to the assumptions of the approach is its reconstruction **precision**. The precision of a method describes the consistency of the reconstructed 3D model with the actual real-world scene. There is a trade-off between precision and reconstruction feasibility. One can witness a correlation between reconstruction precision and requirements: the higher the envisaged reconstruction precision, the more assumptions and priors have to be made on the reconstruction domain.

Exact reconstructions from a single image are only possible if strong assumptions are made, e.g. piecewise planarity with only three orientations (Manhattan assumption) [33] or known reference heights and a calibrated camera [31]. Since such strict assumptions strongly limit the applicability of the approach, most approaches revert to computing the most likely solution to the ill-posed reconstruction problem without guaranteeing accuracy. The probability of a solution is usually measured by means of manually defined priors [43], learned shape priors [25, 89], smoothness priors [86, 114], [9] or other priors. We call this a *plausible* precision. Finally, there are approaches, which do not aim for a reconstruction at all. Instead, they find solutions which look good to the viewer when animated [46, 92, 50] or can be used to synthesize approximate new views of a scene. We call these reconstructions *pleasing*. The reconstruction precision is indicated in the third column of Table 3.1. '$=$' indicates exact precision, '$\simeq$' plausible precision and '$\approx$' a pleasing approach. Surely there are smooth transitions between these classes.

Representation. Different **surface representations** have a strong impact on reconstruction domain, complexity, runtime, memory usage and versatility of the algorithm. We distinguish between parametric and implicit representations and those which are neither parametric nor implicit. Each point on a *parametric surface* can be uniquely described by a coordinate. Finding a good parametrization for an object is not straightforward and generally does not allow for arbitrary topology. *Implicit surfaces* are a remedy to this problem. In this case, the surface is a level set of a function defined on $\mathbb{R}^3$. In contrast to parametric surfaces, single points on the surface are not easily addressed. *Polygonal* surface representations are neither parametric nor implicit and can be described as a planar graph with nodes, edges and faces. Note that polygonal surfaces often describe piecewise planar objects but are also used for approximating curved parametric surfaces. Finally, representations can describe *closed* and *non-closed* 3D surfaces. A *depth map* is an example for a non-closed representation and assigns a depth value to each pixel. An implicit representation describes closed surfaces where the backside of the surface is invisible from any chosen view point.

User Input and Runtime. Most single view approaches rely on **user input** to dissolve ambiguities. The complexity of the input varies strongly among different approaches.

User input can convey low-level and high-level information. High-level input is of semantic quality which helps to dissolve ambiguities, e.g. the object silhouette.

This stands in contrast to tools, where the user *models* the reconstruction with the help of low-level operations, e.g. by specifying surface normals or cutting away object parts. Many of these *modeling tools* [56, 107, 15] are not image-based and therefore only remotely related to single view 3D reconstruction. In *sketch based modeling tools* [53, 78, 59, 112] such modeling operations are driven by user indicated strokes. The Teddy tool will be examined in more detail in Section 6.5. A pioneering work on free-form modeling was done by Welch and Witkin [106].

There is 2D and 3D user input. Most approaches use 2D input which in most cases is directly applied to the input image [9]. This involves tracing contour edges such as creases or vanishing lines. 3D input is directly applied to the reconstruction surface and is often more involved for the user as he needs to navigate in 3D space (e.g. specifying normals).

For some approaches the user input stage is separate from the reconstruction stage [25, 31]. Other methods compute a first reconstruction, then the user can add further input and the process is continued [114, 86], [5, 9]. For approaches of the latter kind, runtime is obviously an important factor.

Image Cues and Priors. The **image cues** and **priors** each method builds on are listed in the lower rows of Table 3.1. See Section 3.1 for an explanation of the respective cues and priors. Cues and priors are marked by a cross "x" if they are being used by the method and by an "L" if they are learned beforehand.

3.4 Conclusion

Single view 3D reconstruction approaches infer 3D objects from just one 2D image. Due to the ill-posed problem formulation there exist a variety of strongly differing approaches that tackle the problem by imposing strong prior information or utilizing human input.

In this chapter, we discussed a representative set of existing algorithms and grouped them into four classes: curved objects, piecewise planar objects, learning specific objects and 3D impression from scenes. These groupings are based on the different reconstruction objectives of the algorithms. We identified several important categories for comparing the approaches: the reconstruction precision, the 3D object representation, the assumptions made by the algorithm, the required user interaction, as well as image cues and priors used by the algorithms.

4 Interactive Workflow and the Minimal Surface Assumption

Several methods for single view 3D reconstruction have been reviewed in Chapter 3. It was shown that those approaches are based on largely differing assumptions, accuracy and priors and focus on diverse reconstruction scenarios. At the same time they have to deal with common problems and often exhibit similar shortcomings.

In this thesis we will concentrate on the 3D reconstruction of one or multiple closed, curved objects seen in a single image. Our approaches will neither concentrate on exact reconstructions, nor on polyhedral representations. Instead, it is designed to compute reasonable 3D geometry estimates of man-made and natural entities that often will be fairly close to the real object, which means that a comparison to a ground truth surface could be meaningfully evaluated.

More precisely, the contribution of this thesis is to explore the computation of minimal surfaces with variational methods in the context of single view 3D reconstruction. We will show that these methods lead to elegant mathematical formulations of the reconstruction problem, which can be efficiently optimized exactly or within provable bounds from the optimum, lead to compelling reconstructions by employing simple and sparse user input and address some of the common problems of state-of-the-art methods.

The following section will introduce the basic interactive framework for all the proposed reconstruction approaches in this thesis. In the second part we will detail the fundamental theoretical concepts of computing reconstructions by variational minimal surfaces. Both sections will form the basis for the proposed single view approaches in this thesis.

4.1 Interactive Work Flow

All of the proposed single view 3D reconstruction approaches in this thesis are based on an interactive framework. Design premises are the following: the tool should be easy to use, little or no expert knowledge should be required by the user, reconstruction runtimes should be small enough for enabling user interaction and results should be finished within few mouse clicks. The workflow of the reconstruction tool is depicted in Figure 4.1.

In the variational reconstruction approaches that are proposed in this thesis the object silhouette plays a central role. Firstly, it encodes a main part of the shape information

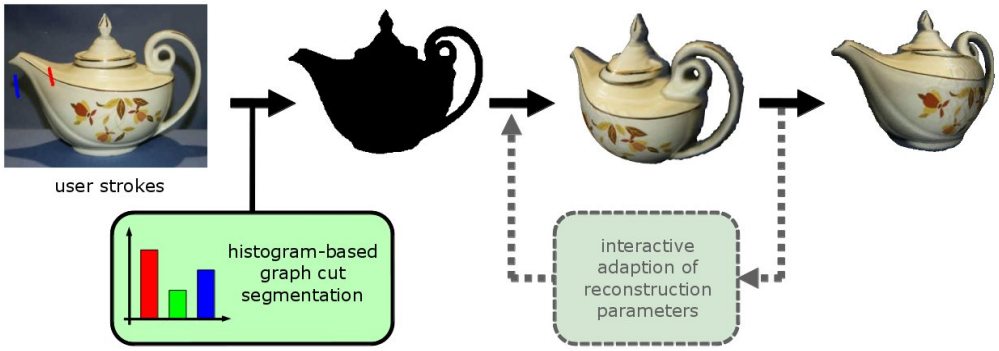


Figure 4.1: The basic workflow of the single view 3D reconstruction process: the user marks the input image with scribbles indicating foreground (red) and background (blue) exemplarily (*left*). From this a silhouette is generated by a graph cut segmentation algorithm (*second from left*). A first reconstruction estimate is then generated from the silhouette. The user can iteratively adapt the parameters (*third from left*) until a satisfactory reconstruction is acquired (*right*).

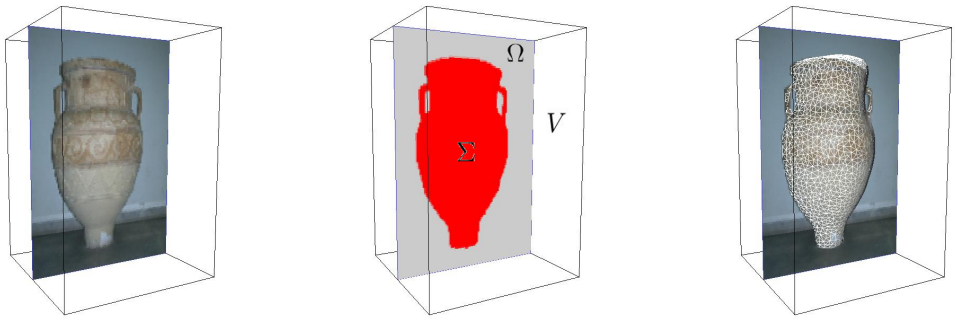
for the reconstruction. Secondly, the topology of the reconstruction surface is determined by the number of holes that the silhouette exhibits. Therefore, a high quality silhouette is the main prerequisite for a good reconstruction result. At the same time, the user effort to extract the silhouette should be kept small.

In the proposed reconstruction framework, the segmentation is obtained by utilizing an interactive graph cut scheme similar to the ones described by [17, 88] [7]. The algorithm calculates a partition of the image domain based on color histograms for foreground and background which are computed from representational user pen strokes (see Fig. 4.1).

From the information in the image and silhouette a first reconstruction is computed with corresponding parameters that are automatically estimated or hard coded. Number and types of parameters depend on the particular reconstruction method used and will be described in the following chapters.

Depending on the complexity of the object the preliminary reconstruction can already be satisfactory. If not, the user can subsequently adapt the reconstruction by specifying intuitive and simple global and local constraints, which depend on the specific reconstruction method. The editing stage can be reiterated by the user until the desired result is obtained. The texture is added to the reconstructions by orthogonally projecting the input image onto the reconstruction mesh.

Notably, for all proposed reconstruction methods the silhouette does not necessarily have to be fully connected. Disconnected parts in the silhouette will yield separate reconstruction objects automatically.



a) b) c)

Figure 4.2: Basic assumptions of all reconstruction approaches. a) The reconstruction volume $V \subset R^3$ contains the image plane Ω. b) The object silhouette Σ is part of the image plane Ω. c) To make the reconstruction feasible the object geometry is assumed to be symmetric with respect to the image plane Ω.

4.2 Basic Assumptions on the Reconstructions

To obtain feasible reconstructions we make the following assumptions, which are illustrated in Figure 4.2. Ω denotes the image plane which contains the input image and lies inside the reconstruction volume $V \subset \mathbb{R}^3$ (see Figure 4.2 a). As part of the image we also have an object silhouette $\Sigma \subset \Omega$ (Figure 4.2 b), which is obtained by means of an interactive segmentation tool. Since we reconstruct objects based on a single view only we require the object to be symmetric with respect to the image plane Ω. Based on these assumptions we can now introduce the minimal surface concept for single view reconstruction.

4.3 Minimal Surfaces

Assume we are given the silhouette of an object in an image obtained by means of an interactive segmentation tool. The goal is then to compute 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. All of the proposed approaches in this thesis are based on the prior assumption that a majority of objects in the natural and man-made environments follow the principles of minimal surfaces. In the following we will introduce the mathematical notion of this concept and introduce a basic variational framework for computing minimal surfaces that comply with a given silhouette of an

object.

We seek to compute reconstructions as minimal weighted surfaces $S \subset \mathbb{R}^3$ that are compliant with the object silhouette Σ:

$$\min \int_S g(s)ds \tag{4.1}$$

$$\text{subject to} \qquad \pi(S) = \Sigma \tag{4.2}$$

where $\pi : \mathbb{R}^3 \to \Omega$ is the orthographic projection onto the image plane Ω, $g : \mathbb{R}^3 \to \mathbb{R}^+$ is a smoothness weighting function and $s \in S$ is a surface element. In the following we derive an implicit representation for the above problem. The main advantage of an implicit surface representation is that the implementation does not need to deal explicitly with the surface topology.

To begin with, the surface S is replaced with its implicit binary indicator function $u \in BV(\mathbb{R}^3; \{0, 1\})$, where BV denotes the set of functions of bounded variation. The function u indicates the exterior ($u(x) = 0$) or interior ($u(x) = 1$) of the surface. As was shown in Chapter 2, the weighted surface area of an implicit surface $u \in BV(\mathbb{R}^3; \{0, 1\})$ is given by the total variation of u. Thus, finding a reconstruction surface with minimal area over a suitable set U of feasible surface functions yields the following problem:

$$\min_{u \in U} \int g(x)|\nabla u(x)|d^3x \tag{4.3}$$

where ∇u denotes the derivative in the distributional sense. Equation (4.3) favors smooth solutions. However, smoothness is locally affected by the function $g(x) : \mathbb{R}^3 \to \mathbb{R}^+$ which will be used in later chapters for modeling sharp edges and protrusions (see details there).

What does the set U of feasible functions look like? For simplicity, we assume the silhouette to be enclosed by the surface. Then the following set describes surface functions that are consistent with the silhouette Σ

$$U_\Sigma = \left\{ u \in BV(\mathbb{R}^3; \{0, 1\}) \;\middle|\; u(x) = \begin{cases} 0, & \pi(x) \notin \Sigma \\ 1, & x \in \Sigma \\ 0 \text{ or } 1, & \text{otherwise.} \end{cases} \right\} \tag{4.4}$$

Solving for a solution to (4.3) with respect to the set U_Σ of silhouette consistent functions will result in the silhouette itself.

As described in Chapter 3 this *surface collapsing* is a common problem in single view 3D reconstruction. Many works, therefore, revert to inflation heuristics. These techniques boil down to fixing absolute depth values. This is undesirable as true depth values cannot be recovered from a single image and fixing them to arbitrary values is tedious for the user and inhibits the flexibility of the reconstruction approach. In the following chapters different remedies for this problem are proposed that refrain from such heuristics.

5 A First Variational Method Using a Shape Prior

In this chapter we devise a first approach to 3D reconstruction from a single image. The proposed method extends the variational minimal surface formulation introduced in the last chapter by a parametric shape prior. This prior fulfills two main tasks: it determines the basic shape of the reconstruction, and it is a means of inflating the geometry in order to avoid surface collapsing.

Contributions. In this chapter a shape prior based variational single view 3D reconstruction approach is proposed, which comes with the following favorable properties:

- The variational framework is elegant and provably computes globally optimal minimal surfaces that comply with the object silhouette. It is highly parallelizable and solutions can be computed within interactive runtimes.
- The implicit surface representation allows for reconstructions with arbitrary topology and genus.
- User input is kept simple compared to state-of-the-art methods and little expert knowledge is required.
- Inflation of the reconstruction surface is done implicitly by the shape prior, which avoids surface collapsing.

This approach was published in [5]. In the following we will present the mathematical formulation of the minimal surface reconstruction problem and define the parametric shape prior. After that, the numerical optimization of the resulting problem will be discussed. In the experimental section that follows, the embedding into an interactive framework is described and qualitative results and runtimes are presented.

5.1 Variational Formulation with Shape Prior

There are different ways to model shape priors. One option is to introduce hard constraints similar to the required silhouette consistency, see Chapter 8. In this chapter a different path is taken by introducing a data affiliation term into the 3D reconstruction

energy functional, which measures how much the reconstruction surface agrees with the shape that is considered most probable.

5.1.1 Shape Prior Definition

In general, the data term takes the following form

$$E_{\text{shape}} = \int u(x)\ \phi_{\text{shape}}(x) d^3x \tag{5.1}$$

where $\phi_{\text{shape}} : \mathbb{R}^3 \mapsto \mathbb{R}$ can be chosen arbitrarily. In this chapter, we set for each voxel $x \in \mathbb{R}^3$

$$\phi_{\text{shape}}(x) \begin{cases} < 0, & \text{if } x \text{ is favored to lie inside the object} \\ > 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

In the following we will derive a parametric shape prior which depends on the silhouette Σ. For any point $p \in R^n$ let

$$dist(p, A) = \min_{s \in A} \|p - s\| \tag{5.3}$$

denote its distance to the set $A \subset \mathbb{R}^n$. Setting $A = \partial\Sigma$ and $p \in \Omega$ we, for example, obtain the distance transform of the contour $\partial\Sigma$. We now define the shape prior by means of a height map $h : \Omega \mapsto \mathbb{R}$, which maps each projected point $\pi(x)$ to the maximum height value the object shall have in this location

$$\phi_{\text{shape}}(x) = \begin{cases} -1 & \text{if } dist(x, \Omega) \leq h(\pi(x)) \\ +1 & \text{otherwise .} \end{cases} \tag{5.4}$$

For defining the height map h we make the simple assumption that the thickness of the observed object increases as we move inward from its silhouette. To this end, h is defined by means of a parametric pyramidal shape

$$h(p) = \min \left\{ \mu_{\text{cutoff}}, \mu_{\text{offset}} + \mu_{\text{factor}} * dist(p, \partial\Sigma)^k \right\} \tag{5.5}$$

based on the four parameters $k, \mu_{\text{offset}}, \mu_{\text{factor}}, \mu_{\text{cutoff}} \in \mathbb{R}^+$. μ_{offset} and μ_{cutoff} define the minimum and maximum height value, whereas μ_{factor} and the exponent k determine the slope of the height map, see Figure 5.1. The function can evolve either linearly ($k = 1$), in a concave ($k > 1$) or in a convex fashion ($k < 1$). For $k \to \infty$ the shape becomes a box with an infinite slope at the silhouette borders. Note that the impact of all shape parameters is attenuated with increasing weight λ on the smoothness term.

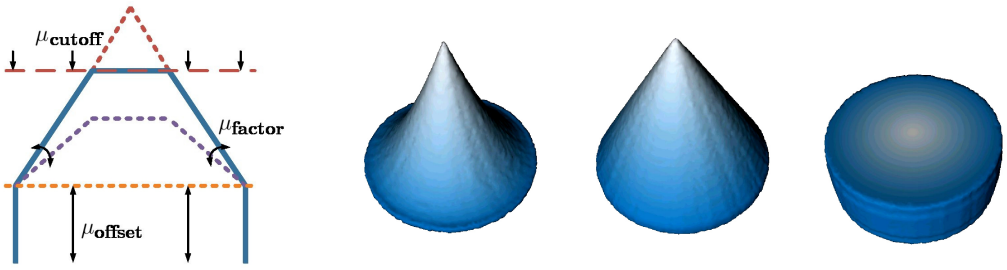


Figure 5.1: *Left*: definition of the shape prior parameters $\mu_{\text{offset}}, \mu_{\text{factor}}, \mu_{\text{cutoff}}$ in (5.5), *Right*: resulting (scaled) height map plots for a circular silhouette for various values of exponent k.

5.1.2 Complete Energy with Shape Prior

Adding shape prior (5.1) to the minimal surface energy (4.3) amounts to the following optimization problem:

$$\min_{u \in U_\Sigma} \int u(x)\ \phi_{\text{shape}}(x) d^3x + \lambda \int g(x) |\nabla u(x)| d^3x \tag{5.6}$$

where λ is a weighting parameter that determines the relative smoothness of the solution. The silhouette consistency constraint contained in U_Σ can be integrated into the shape prior by defining the following function

$$\phi_{\text{sil}}(x) = \begin{cases} \infty, & \pi(x) \notin \Sigma \\ -\infty, & x \in \Sigma \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

This function assures that all points projecting outside of the silhouette will be assigned to the background (i.e. $u(x)=0$) and that all image plane points within the contour will be assigned to the object ($u(x)=1$).

Merging the function ϕ_{sil} and the shape prior ϕ_{shape} yields the complete 3D reconstruction problem

$$\min_{u \in BV(\mathbb{R}^3;\{0,1\})} \int u(x)\ (\phi_{\text{sil}}(x) + \phi_{shape}(x)) d^3x + \lambda \int g(x) |\nabla u(x)| d^3x. \tag{5.8}$$

5.2 Efficient Optimization via Convex Relaxation

Energy (5.8) is highly non-convex since the optimization domain consists of binary functions. We will, thus, revert to a relaxation technique to arrive at a convex minimization

problem which can be optimized globally. Subsequent thresholding will yield a binary solution to the reconstruction problem. The relaxation plus corresponding thresholding scheme was first shown to yield globally optimal results for this kind of energy functionals by Chan et al. in [22].

A relaxation of the problem in (5.8) is obtained by replacing the optimization domain by its convex hull, i.e. $u \in BV(\mathbb{R}^3; [0,1])$. This allows the surface indicator function to take on values in the unit interval $[0,1]$ and the problem reads as

$$\min_{u \in BV(\mathbb{R}^3;[0,1])} E(u) \text{ where } E(u) := \int u(x)\phi(x)d^3x + \lambda \int g(x)|\nabla u(x)|d^3x \tag{5.9}$$

and $\phi := \phi_{\text{sil}}(x) + \phi_{shape}$.

Proposition 5. The optimization problem in (5.9) is convex.

Proof. For all $u_1, u_2 \in BV(\mathbb{R}^3; [0,1])$ all linear combinations $\alpha u_1 + (1-\alpha)u_2,\ \alpha \in [0,1]$ are obviously element of $BV(\mathbb{R}^3; [0,1])$, so the optimization domain is convex. The functional E is convex since for any functions u_1 and u_2 and any $\alpha \in (0,1)$ we have:

$$\begin{aligned}
&E(\alpha u_1 + (1-\alpha)u_2) \\
&\quad = \int (\alpha u_1(x) + (1-\alpha)u_2(x))\phi(x) + \lambda g(x)|\nabla(\alpha u_1 + (1-\alpha)u_2)|d^3x \\
&\quad = \int \alpha u_1(x)\phi(x) + (1-\alpha)u_2(x)\phi(x) + \lambda g(x)|\alpha\nabla u_1 + (1-\alpha)\nabla u_2)|d^3x \\
&\quad \leq \alpha \int u_1(x)\phi(x) + \lambda g(x)|\nabla u_1|d^3x + (1-\alpha)\int u_2(x)\phi(x) + \lambda g(x)|\nabla u_2|d^3x \\
&\quad = \alpha E(u_1) + (1-\alpha)E(u_2).
\end{aligned}$$

□

The Euler Lagrange Equations to (5.9) describe a non-linear diffusion process

$$0 = \phi - \lambda \text{ div}\left(g(x)\frac{\nabla u}{|\nabla u|}\right), \qquad \forall x \in \mathbb{R}^3. \tag{5.10}$$

In order to solve this system we follow the fixed-point iteration scheme first developed by Vogel and Oman [104] and applied to 3D reconstruction in [63].

Lagged Diffusivity Fixed-point Iteration. This scheme is also known as *lagged diffusivity*: the system (5.10) is linearized by treating the diffusivity $d(x) := \frac{g(x)}{|\nabla u|}$ as a constant over a number of fixed-point iterations:

$$0 = \phi - d(x)\Delta u, \qquad \forall x \in \mathbb{R}^3. \tag{5.11}$$

For solving this linear system efficiently Successive Over-relaxation (SOR) is used, a Gauss-Seidel fixed-point iteration scheme combined with an over-relaxation step for faster convergence. The diffusivity is updated every ten to twenty steps.

Discretization. The Laplace operator in (5.11) is discretized by a standard 3D Laplace kernel. An update step of the *Red-black* version of the SOR scheme then reads as

$$u_i^{(t+1)} = (1-\omega)u_i^{(t)} + \omega\left(\frac{\phi - \sum_{i\sim j} d_j u_j^{(t)}}{\sum_{i\sim j} d_j}\right) \tag{5.12}$$

where $i \sim j$ denotes neighboring voxels using 6-connectivity and $\omega \in (0,2)$ is the over-relaxation parameter. The voxels of the optimization domain can be imagined colored as *red* and *black* fields of a checkerboard. Update step (5.12) is alternatively applied to the *red* and to the *black* set of voxels. This is a parallelizable alternative to the Gauss-Seidel scheme, which has to be implemented sequentially.

For numerical stability we compute the partial derivatives in the diffusivities d_i by a non-negativity scheme. This scheme will be detailed in Chapter 7 for the 2D case in Equations (7.25)-(7.28). The 3D case is a straightforward generalization.

The diffusivities will be undefined if the norm of the gradient becomes zero. Therefore a small ϵ has to be added to the gradient before applying the norm. This is numerically problematic as values can get very large at points having a small or zero gradient. Due to this, the proposed algorithm will not compute the exact solution to (5.8). A remedy is to employ a primal-dual scheme. This strategy will be detailed in Chapter 6 and can be directly transfered to the functional at hand in (5.9).

Thresholding. To obtain a binary solution from the globally optimal computed relaxed solution a thresholding scheme is applied. In [22] it was shown that picking an arbitrary threshold $\theta \in (0,1)$ will in fact result in a globally optimal binary solution. This is known as the *thresholding theorem*.

Proposition 6. Let $u^* \in BV(\mathbb{R}^3;[0,1])$ be a globally optimal solution to the relaxed problem (5.9). Then for any $\theta \in (0,1)$, the thresholded solution $\mathbf{1}_{u^*\geq\theta}$ is a global optimum of the original binary problem (5.8).

Proof. Using the layer cake representation of $u(x) = \int_0^1 \mathbf{1}_{u\geq\mu} d\mu$ and the co-area formula $\int g(x)|\nabla u|dx = \int_0^1 \int g(x)|\nabla \mathbf{1}_{u\geq\mu}|dxd\mu$ the energy in (5.9) can be rewritten in terms of $\mathbf{1}_{u\geq\mu}$

$$E(u) = \int_0^1 \int \mathbf{1}_{u\geq\mu}\phi(x) + g(x)|\nabla \mathbf{1}_{u\geq\mu}|dxd\mu = \int_0^1 E(\mathbf{1}_{u\geq\mu})d\mu. \tag{5.13}$$

Let u_{opt} be an optimal binary solution. Assume that for some θ the thresholded solution $\mathbf{1}_{u^*\geq\theta}$ is not optimal, then for these values of θ it holds $E(u_{\text{opt}}) < E(\mathbf{1}_{u^*\geq\theta})$. For all other thresholds θ we have $E(u_{\text{opt}}) = E(\mathbf{1}_{u^*\geq\theta})$. It follows

$$E(u_{\text{opt}}) = \int_0^1 E(u_{\text{opt}})d\mu < \int_0^1 E(\mathbf{1}_{u^*\geq\mu})d\mu = E(u^*). \tag{5.14}$$

Since u^* is a minimizer of the relaxed problem it holds that $E(u_{\text{opt}}) \geq E(u^*)$, a contradiction. □

5.3 Experimental Results

In the following, we will carry out experiments based on the proposed variational reconstruction method. First we will describe the interactive modeling of the shape prior before experimental results are given.

5.3.1 Interactive Shape Prior Modeling

A good shape prior should be flexible so that it is applicable to a wide range of objects. To this end, this framework allows for interactive adaptation of the shape prior parameters in (5.4) by means of three different kinds of editing tools: global shape parameters and local shape parameters in combination with a local relaxation of the minimal surface assumption. The first two influence the shape prior directly while the third one defines the weighting function g.

Global Shape Prior Parameters. The parameters $\lambda_{\text{offset}}, \lambda_{\text{factor}}, \lambda_{\text{cutoff}}$ and the exponent k of the function (5.4) define the basic shape of the prior, refer to the explanation in Section 5.1.2 and Figure 5.1.

Local Shape Prior Parameters. The shape priors in (5.4) tend to be flat at the silhouette borders due to the distance transform in the shape prior height map definition (5.5). This behavior is not always desirable as the vase example in Figure 5.2 demonstrates. The shown vase reconstruction should be bulgy exhibiting a sharp edge at the top and the bottom. Instead, the shape prior assumes a smooth transition of height values to zero close to the silhouette boundaries.

A remedy is to let the user mark these contour parts in the image and exclude them from the calculation of the distance transform, i.e. in the *dist* function in (5.4).

Local Discontinuities. Reconstructions resulting from a minimal surface approach such as (5.8) tend to be roundish regardless of the silhouette (see Figure 5.2). The weight $g(x)$ in the surface area term of (5.8), however, can be leveraged to induce creases, discontinuities or protrusions on the reconstruction surface. By setting $g(x) < 1$ the smoothness of the surface at location x is relaxed.

The user alters local surface smoothness by drawing curves into the input image and associating them with a reduced value for $g(x)$. For all the voxels in the reconstruction volume that project to the drawn curve in the image the smoothness weight $g(x)$ is then set to the according value. The weight of the remaining voxels is set to $g(x) = 1$ by

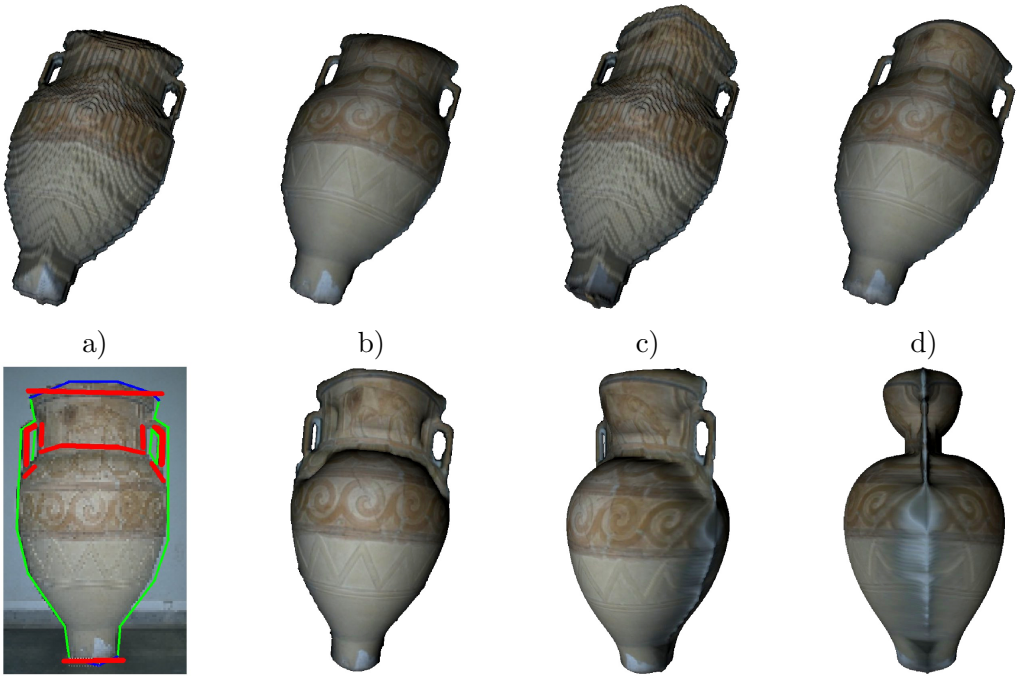


Figure 5.2: *Top:* improving the shape prior for the vase image in e) by excluding silhouette parts in the distance transform, a) original shape prior with corresponding reconstruction in b) exhibiting flat top and bottom, c) improved shape prior after excluding the top and bottom of the vase (marked in blue in e)) from the reconstruction yielding the result in d). *Bottom:* locally relaxing the smoothness where marked by the red user strokes in e) yields the reconstruction shown from three different view points in f),g) and h).

default. The effects of this input can be observed in the bottom row of Figure 5.2 and in some examples in the experimental section below.

5.3.2 Experiments

The proposed method is applied to several input images in Figure 5.3. For each object the silhouette is extracted and functional (5.8) is minimized.

The examples demonstrate different qualities of the proposed approach. The fence (top row) is a an object with complex topology of high genus. A drawback of parametric reconstruction approaches in general is that the genus of the reconstruction has to be modeled explicitly by the user [86]. In contrast, the proposed formulation in (5.6) is ignorant of the topology of the reconstruction surface, and its genus is defined implicitly

by the number of holes in the silhouette. For the fence example no user input was necessary except for the silhouette extraction from the image.

The remaining reconstruction examples were obtained by editing the parameters of the shape prior as described in Section 5.3.1. In particular, the λ_{cutoff} parameter was altered to a low value in the biker example. The lower part of the opera building and the socket of the Cristo statue were excluded from the distance transform computation as described in Section 5.3.1. Scribble-based local relaxation of the surface smoothness was applied to the beak in the cockatoo example and again to the socket of the Cristo statue in order to obtain sharp edges in the reconstruction. All editing operations are simple in the sense that firstly, they do not require expert knowledge, secondly they are intuitive and thirdly they are specified by means of user strokes in the image instead of the 3D model.

To limit the runtime and memory demand the input images were scaled down to size $256x256$. The meshes that are shown in Figure 5.3 were recovered by applying the *marching cubes* method [71] to the implicit binarized surface function after reconstruction. On an Nvidia GeForce GTX 480 graphics card the computation of the reconstruction geometry takes between 0.1 and 15 seconds, depending on the input silhouette, volume resolution and applied editing operations.

5.4 Conclusion

In this chapter a first variational approach to single view 3D reconstruction for closed, curved objects was introduced by combining the minimal surface idea of Chapter 4 with a parametric shape prior. 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 but simple user input. Globally optimal solutions are obtained via convex relaxation and subsequent thresholding of the variational energy minimization problem. The algorithm can be used interactively, since the parallel implementation of the underlying nonlinear diffusion process on standard graphics cards only requires short runtimes.

The shape prior solves the surface collapsing problem and imposes shape on the reconstruction. One problem of this approach, however, is that a parametric prior in many cases is too rigid and that the distance transform is a heuristic assumption that inhibits the flexibility of the approach during the editing phase. Furthermore, the user still has to specify several parameters, which can be tedious. The following chapter will address these problems.

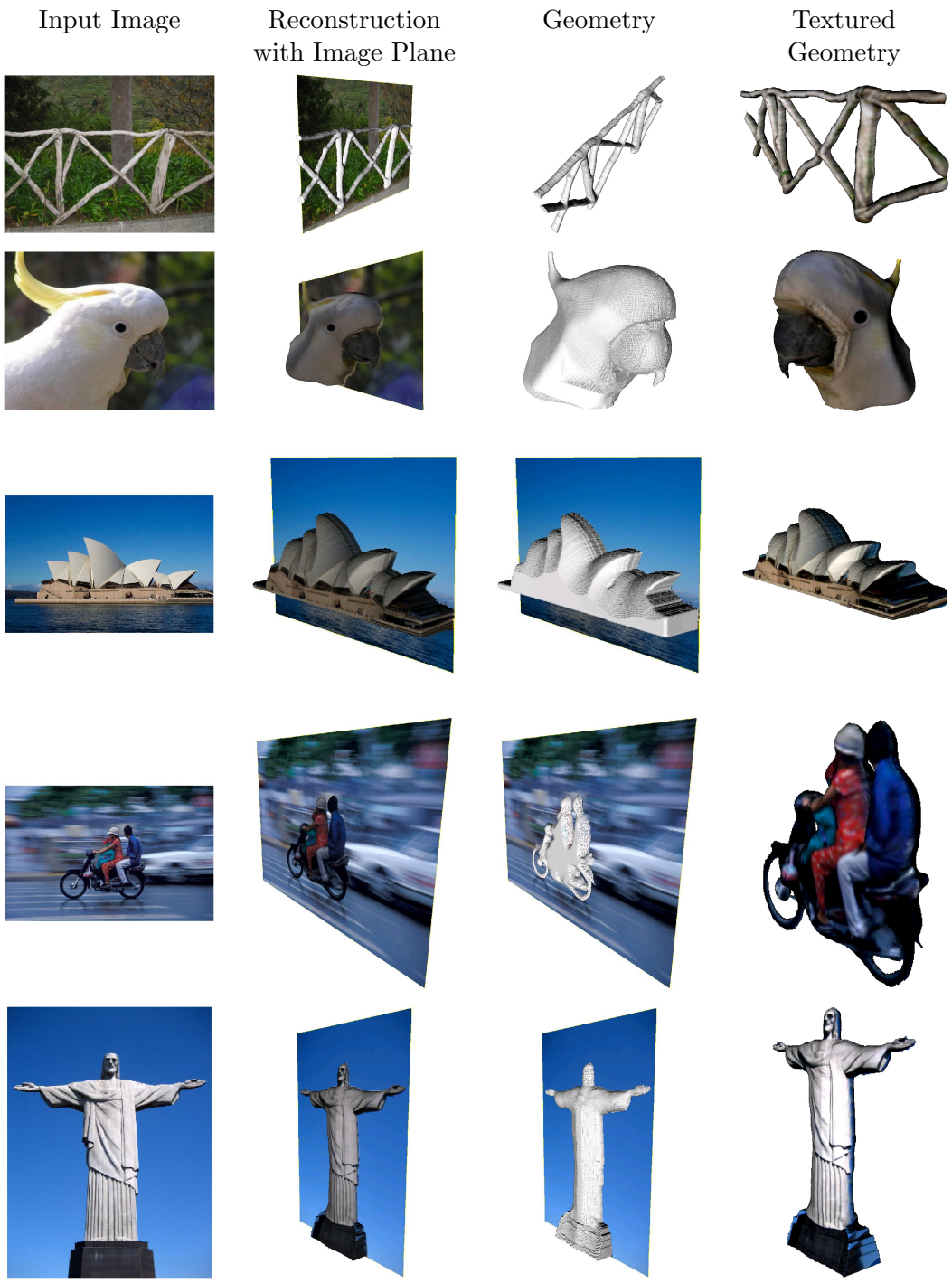


Figure 5.3: *Left:* input images, *Right:* corresponding reconstruction results of the proposed variational shape prior based approach.

6 Cheeger Set Formulation

In the last chapter a shape prior was added to the variational minimal surface problem (4.3) and it was shown that within an interactive environment this concept yields reasonable 3D reconstruction results for many images. However, one could see that a data term imposes a strong bias on the shape. Furthermore, the user has to tune several parameters in order to obtain the appropriate shape for the respective image. Ideally, we strive for a non-heuristic inflation incentive that does not inhibit the shape variety, has less parameters and results in compact surfaces. To this end, in this chapter a constraint on the size of the volume enclosed by the minimal surface is proposed.

Contributions. In this chapter another minimal surface approach is proposed that overcomes the limitations mentioned above. The new approach will share most of the advantages of the method from Chapter 5 such as speed, freedom of topology, easy surface inflation and simplicity of user input. In particular:

- It will be shown that one can efficiently compute silhouette-consistent, weighted minimal surfaces for a user-prescribed volume within provable bounds from the global optimum.
- Surface collapsing is avoided.
- User input is strongly reduced.
- No heuristic or shape prior is required, since volume constraints in combination with minimal surface assumptions lead to an automatic inflation of the object.

This approach was published in [9] together with a comparison of different optimization methods in [1]. A comparison to the shape prior approach in the previous Chapter was published in [10].

Related Work. A precursor to volume constraints are the volume inflation terms pioneered for deformable models by Cohen and Cohen [26]. However, no constant volume constraints were considered and no implicit representations were used.

In fact, the presented solution can be considered as a means of inflating the object

silhouette like a balloon. In the following we will first introduce a variational formulation of this intuition and discuss how the resulting energy can be optimized efficiently. Some theoretical considerations on solutions of the minimization problem will be given subsequently, followed by a discussion on efficient optimization techniques. After showing how the proposed approach is embedded into the interactive framework from Chapter 4, the last section will discuss qualitative experiments and runtimes of the method as well as a comparison to related state-of-the-art approaches.

6.1 Variational Formulation with Volume Constraints

The constraint on the size of the volume enclosed by the minimal surface is formulated as a soft- and as a hard constraint both of which are discussed in the following. We start from the minimal surface problem (4.3) based on the feasible set of silhouette consistent reconstructions U_Σ defined in (4.4).

Hard Constraint.

By further constraining the feasible set U_Σ of the minimal surface problem one can force the reconstructed surface to have a specific target volume V. This leads to the following formulation

$$\min_{u \in U_\Sigma \cap U_V} \int g(x)|\nabla u(x)| d^3x \tag{6.1}$$

$$\text{where} \quad U_V = \left\{ u \in BV(\mathbb{R}^3; \{0,1\}) \;\middle|\; \int u(x) d^3x = V \right\} \quad . \tag{6.2}$$

Here, U_V denotes all reconstructions with bounded variation that inscribe the specific volume V. Solutions to (6.1) is a Cheeger set problem. Cheeger sets are minimal surfaces for a fixed volume. As an important special case, for a circle shaped silhouette and the right volume the corresponding Cheeger set is a ball.

Soft Constraint.

Alternatively one can formulate a soft constraint by adding a ballooning term to the minimal surface energy (4.3):

$$E_V(u) = \lambda \left(\int u(x) d^3x - V \right)^2 . \tag{6.3}$$

The term quadratically punishes the deviation of the surface volume from a certain target value V. In contrast to the hard constraint above, this formulation comes with an extra parameter λ. To avoid additional parameters in the following we will focus on

(6.1) instead.

Notably, in contrast to the version with shape prior presented in Chapter 5 these formulations do not contain explicit information as to where the surface is inflated and to what extent. Rather, local volume inflation is a result of finding the minimal surface which only depends on the silhouette shape and the target volume V. The narrower the silhouette is the more expensive is its inflation due to the increased surface energy, so wider silhouette regions will be inflated more.

6.2 Efficient Optimization via Convex Relaxation

Optimizing (6.1) is hard, since the set of feasible reconstruction surfaces consists of binary valued functions making it highly non-convex (see Chapter 2).

In order to be able to apply continuous numerical optimization schemes such as the ones reviewed in Chapter 2, we will have to convexify the problem by relaxing the energy formulation. After having found a solution to the relaxed problem, we apply a thresholding scheme to obtain a binary solution again. We will later explore optimality bounds for this binary solution with respect to the original problem (6.1).

To relax the optimization problem we replace the sets U_Σ and U_V in the definition (6.1) by their respective convex hulls U_Σ^r and U_V^r. Formally, this is done by substituting the function domains $BV(\mathbb{R}^3;\{0,1\})$ by $BV(\mathbb{R}^3;[0,1])$ in definitions (4.4) and (6.2) yielding

$$U^r := U_\Sigma^r \cap U_V^r = \left\{ u \in BV(\mathbb{R}^3;[0,1]) \;\middle|\; \int u(x)d^3x = V \quad \text{and} \quad u(x) = \begin{cases} 0, & \pi(x) \notin \Sigma \\ 1, & x \in \Sigma \\ 0 \text{ or } 1, & \text{otherwise.} \end{cases} \right\} \tag{6.4}$$

In Chapter 5 it was shown that the energy (6.1) is convex if its feasible set is convex. Thus, it suffices to show that

Proposition 7. 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 and U_Σ^r. As an intersection of two convex sets U^r is convex as well. □

A common way of finding the globally optimal solution to this energy minimization problem is gradient descent. However, this method converges rather slowly. Moreover, as seen in Chapter 5, no exact solution to (6.1) can be found with this method, as the corresponding Euler-Lagrange equations are not defined for vanishing ∇u. Because of this and since optimization speed is an integral part of an interactive reconstruction

framework, two different provably convergent strategies for exact convex optimization are proposed in the following: the *primal-dual* algorithm published in [81] and the *alternate direction method of multipliers* (ADMM). Both algorithms were reviewed in Chapter 2. A comparison was published in [1].

6.2.1 Optimization Using the Primal-dual Scheme

In order to apply the *primal-dual* scheme, we replace the minimal surface term in (6.1) by its weak formulation (see Chapter 2)

$$\min_{u \in U^r} \int g(x) |\nabla u| d^3x = \min_{u \in U^r} \max_{|\xi(x)|_2 \leq g(x)} \left\{ \int -u \ \mathrm{div}\xi \ d^3x \right\} \ , \tag{6.5}$$

which is the primal-dual version of the reconstruction problem.

Optimization is done by alternating a gradient descent with respect to the function u and a gradient ascent for the dual variable $\xi \in \mathcal{C}_c^1(\mathbb{R}^3; \mathbb{R}^3)$. In each iteration step an over-relaxation is done on the primal variable for speed-up.

The main problem in applying the primal-dual scheme is that the solution u has to fulfill three constraints: silhouette consistency, constant volume and $u \in [0, 1]$. As proposed in Chapter 5, in order to maintain silhouette consistency (4.4) we add the data affiliation term (5.7) to energy (6.1). Alternatively, we can confine updates during optimization to those voxels which project onto the silhouette interior excluding the silhouette itself. For ease of notation we will stick to the latter strategy in the following.

The remaining two constraints (volume compliance and $u \in [0, 1]$) can either be enforced by projecting the solution to the feasible set during optimization or by introducing *Lagrange multipliers* into the energy. Both schemes have their strengths and weaknesses and are sketched in the following.

Primal-dual Optimization via Iterative Projection

Applying the primal-dual scheme described in [21] to the saddle point problem (6.5) directly yields the following steps which are iterated until convergence

$$\begin{cases} \xi^{k+1} = \Pi_{|\xi(x)|_2 \leq g(x)}(\xi^k + \tau \cdot \nabla \bar{u}^k) \\ u^{k+1} = \Pi_{U^r}(u^k + \sigma \cdot \mathrm{div}\xi^{k+1}) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \end{cases} \tag{6.6}$$

where Π_A denotes the projection onto the set A. Projection of ξ is done by simple clipping while that of the primal variable u will be detailed in the following.

A closed form solution for the projection onto the remaining constraints in U^r cannot be devised. However, an iterative algorithm which computes the Euclidean projection of a point onto the intersection of arbitrary convex sets was given by Boyle and Dykstra

[18]. It converges provably to the projection solution. Formally, for our case step i of this algorithm reduces to two separate projections for volume and domain

$$\begin{cases} u_V^i = u_D^{i-1} - v_V^{i-1} + \frac{V_d}{N} \\ v_V^i = u_V^i - (u_D^{i-1} - v_V^{i-1}) \end{cases} \tag{6.7}$$

$$\begin{cases} u_D^i = \Pi_{[0,1]}(u_V^i - v_D^{i-1}) \\ v_D^i = u_D^i - (u_V^i - v_D^{i-1}). \end{cases} \tag{6.8}$$

We initialize u_D with the current u^k in (6.6) and v_D, v_V with zero. $\Pi_{[0,1]}(u)$ simply clips the value of u to the unit interval and V_d is the difference between the target volume V and the current volume of the values $u_D^{i-1} - v_V^{i-1}$. N is the number of voxels in the discrete implementation.

The algorithm does not yield the exact projection unless it is iterated infinitely often. This states a theoretical problem, as the primal-dual algorithm is only provably convergent when the precise projection is computed in each step. However, practically this is not an issue.

Interestingly, it can be verified that the ADMM algorithm (see Chapter 2) with $K = I$, $F = \chi_C$ and $G = \chi_D$ is equivalent to Dykstra's algorithm for finding the projection onto the intersection of convex sets $C \cap D$.

Primal-dual Optimization via Lagrange Multipliers

In this formulation the volume constraint is introduced into the primal-dual energy of problem (6.1) with the help of a Lagrange multiplier λ, leading to the following Lagrangian dual problem [16]:

$$\max_{\substack{|\xi(x)|_2 \leq g(x) \\ \lambda \in \mathbb{R}}} \min_{u \in U_\Sigma^\tau} \left\{ \int -u \ \mathrm{div}\xi \ d^3x + \lambda \left(\int u \ d^3x - V \right) \right\} \tag{6.9}$$

$$= \max_{\substack{|\xi(x)|_2 \leq g(x) \\ \lambda \in \mathbb{R}}} \min_{u \in U_\Sigma^\tau} \left\{ \int -u \ \mathrm{div}\xi \ d^3x + \lambda \int u \ d^3x - \lambda V \right\} \ . \tag{6.10}$$

As shown in Chapter 2, this Lagrangian dual is a saddle point problem to which the primal-dual scheme can be applied. The Lagrange multiplier adds one step to the dual gradient ascent in the iteration scheme

$$\begin{cases} \xi^{k+1} = \Pi_{|\xi(x)|_2 \leq g(x)}(\xi^k + \tau_\xi \cdot \nabla \bar{u}^k) \\ \lambda^{k+1} = \lambda^k + \tau_\lambda \cdot (\int \bar{u} \ dx - V) \\ u^{k+1} = \Pi_{u \in [0,1]}(u^k - \sigma \cdot (-\mathrm{div} \ \xi^{k+1} + \lambda^{k+1})) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k. \end{cases} \tag{6.11}$$

The projection for the primal variable u now reduces to a clipping operation which can be done point-wisely.

Of the presented optimization schemes the Lagrange-multiplier version of the single view 3D reconstruction energy with volume constraint is more efficient. However, the projection scheme of Dykstra can be faster when more constraints are introduced. Runtimes of the algorithms will be given in the experimental section below.

6.2.2 Optimization Using Alternate Direction Method of Multipliers

In the following the ADMM algorithm is derived for problem (6.1). For simplicity, we present a spatially discretized version of the problem statement. To this end, the domain $X := \mathbb{R}^{W\times H\times D}$ is defined as a 3-dimensional regular grid with width W, height H and depth D. Then $u \in X$ is the implicit surface and $u_i, 1 \leq i \leq WHD$ its i-th element. We set $Z := (X^2, X)$ and the linear operator $K : X \to Z$ such that $Ku = \begin{pmatrix}\nabla u & u\end{pmatrix}^T$. Then for a suitable discretization $\|u\|_g$ of the TV-norm and a discretized volume constraint set U_V (see both below) and by defining $\xi = \begin{pmatrix}v & w\end{pmatrix}^T \in Z$ we can cast problem (6.1) in the following discrete form

$$\min_{u\in[0,1]^{W\times H\times D}} \|v\|_g + < f, u > +\chi_{U_V}(w) \ \text{ , s.t. } Ku = \xi \tag{6.12}$$

where the data affiliation term $< f, u >$ ensures silhouette consistency (see (5.7)) and χ_A is the characteristic function of set A, see (2.17). The variable splitting with auxiliary variable w will separate the projection on the volume constraint and on the set $[0,1]^{W\times H\times D}$ in the corresponding ADMM scheme

$$\begin{cases} u^{t+1} = \underset{u\in[0,1]^{W\times H\times D}}{\operatorname{argmin}} < f, u > +\frac{\tau}{2}\left\|Ku - \xi^t + \lambda^t\right\|_2^2 \\ r^{t+1} = \alpha K u^{t+1} + (1-\alpha)\xi^t \\ \xi^{t+1} = \underset{\xi\in Z}{\operatorname{argmin}} \|v\|_g + \chi_{U_V}(w) + \frac{\tau}{2}\left\|r^{t+1} - \xi + \lambda^t\right\|_2^2 \\ \lambda^{t+1} = \lambda^t + \tau(r^{t+1} - \xi^{t+1}) \end{cases} \tag{6.13}$$

where $\lambda = \begin{pmatrix}\lambda_1 & \lambda_2\end{pmatrix}^T \in Z$ are the Lagrange multipliers, $r = \begin{pmatrix}r_1 & r_2\end{pmatrix}^T \in Z$ is the over-relaxed solution and $\alpha \in (0, 2)$. Solutions for the subproblems are detailed in the following.

Solving the subproblem in u. Deriving the problem with respect to u results in the following Euler-Lagrange equations

$$(\mathcal{I} - \Delta)u = -\frac{f}{\tau} + w^t - \lambda_2^t - \operatorname{div} v^t + \operatorname{div} \lambda_1^t \ . \tag{6.14}$$

The constrained problem can be approximately solved by the *Projected Gradient Descent* method.

However, we rather solve approximately for u by alternating Jacobi iterations with pointwise projection to the constraint set U_V in the same fashion as [42]. In practice this strategy turned out to be slightly faster than projected gradient descent, while producing the same results. Five iterations for this subproblem suffice to make the ADMM algorithm converge.

Solving the subproblem in v. The closed-form solution for the minimizer of the optimization problem

$$v^{t+1} = \operatorname*{argmin}_v \|v\|_g + \frac{\tau}{2} \left\|v - (r_1^{t+1} + \lambda_1^t)\right\|_2^2 \tag{6.15}$$

is given by the *coupled shrinkage* formula

$$v^{t+1} = \operatorname{shrink}_g(r_1^t + \lambda_1^t, \tau)$$

which we define similarly to [42, 93] as

$$(\operatorname{shrink} g(x,\tau))_i = \begin{cases} 0, & \text{if } \|x_i\|_2 \leq \frac{g_i}{\tau} \\ x_i - \frac{g_i}{\tau} \frac{x_i}{\|x_i\|_2}, & \text{otherwise.} \end{cases} \tag{6.16}$$

See the Appendix for the derivation of the shrinkage formula.

Solving the subproblem in w. According to the definition of the *prox-operator* reviewed in Chapter 2, the closed form solution to the optimization problem in w is given by the orthogonal projection of $r_2^{t+1} + \lambda_2^t$ onto the convex set U_V, since

$$\operatorname*{argmin}_w \chi_{U_V}(w) + \frac{\tau}{2} \left\|w - (r_2^{t+1} + \lambda_2^t)\right\|_2^2 = \operatorname{prox}_{\frac{1}{\tau}\chi_{U_V}}(r_2^{t+1} + \lambda_2^t) = \Pi_{U_V}\left(r_2^{t+1} + \lambda_2^t\right) \quad . \tag{6.17}$$

It is shown in the next section, that this projection can be closely computed by

$$\Pi_{U_V}(u) = u - \frac{1}{|X|}\left(\sum_{i \in X} u_i - V\right) \quad .$$

6.2.3 Implementation Details

Discretization. To implement both optimization schemes we have to discretize the spatial optimization domain. As described for the ADMM scheme, we define a regular grid $X := \mathbb{R}^{W \times H \times D}$ and a surface function $u \in X$. The gradient field of u is given as $\nabla u \in Y := X^2$. Accordingly, the TV-norm is discretized as

$$\|\nabla u\|_g = \sum_{i \in \Omega} g_i \|\nabla u_i\|_2 \quad . \tag{6.18}$$

Similarly, the constraint set of volume compliant surfaces u is equal to

$$U_V := \left\{ u \in X \Big| \sum_i u_i = V \right\} \quad . \tag{6.19}$$

Forward differences are used to compute the gradient field ∇u. At the boundary of X von Neumann boundary conditions are assumed

$$(\nabla u(x,y,z))_x := \begin{cases} u(x+1,y,z) - u(x,y,z) & \text{if} \quad x < W \\ 0 & \text{otherwise} \end{cases} \tag{6.20}$$

$$(\nabla u(x,y,z))_y := \begin{cases} u(x,y+1,z) - u(x,y,z) & \text{if} \quad y < H \\ 0 & \text{otherwise} \end{cases} \tag{6.21}$$

$$(\nabla u(x,y,z))_z := \begin{cases} u(x,y,z+1) - u(x,y,z) & \text{if} \quad z < Z \\ 0 & \text{otherwise.} \end{cases} \tag{6.22}$$

The divergence operator has to be chosen adjoint to the gradient operator, i.e. such that it holds $< \nabla u, \xi >= - < u, \operatorname{div} \xi >$. It is straightforward to show that this is given for the following discretization of $\operatorname{div} \xi$ with backward differences and Dirichlet boundary conditions

$$\partial_x \xi(x,y,z) := \begin{cases} \xi(x,y,z)_1 - \xi(x-1,y,z)_1 & \text{if} \quad 1<x<W \\ \xi(x,y,z)_1 & \text{if} \quad x = 1 \\ -\xi(x-1,y,z)_1 & \text{if} \quad x = W \end{cases} \tag{6.23}$$

$$\partial_y \xi(x,y,z) := \begin{cases} \xi(x,y,z)_2 - \xi(x,y-1,z)_2 & \text{if} \quad 1<y<H \\ \xi(x,y,z)_2 & \text{if} \quad y = 1 \\ -\xi(x,y-1,z)_2 & \text{if} \quad y = H \end{cases} \tag{6.24}$$

$$\partial_z \xi(x,y,z) := \begin{cases} \xi(x,y,z)_3 - \xi(x,y,z-1)_3 & \text{if} \quad 1<z<D \\ \xi(x,y,z)_3 & \text{if} \quad z = 1 \\ -\xi(x,y,z-1)_3 & \text{if} \quad z = D \end{cases} \tag{6.25}$$

and $\operatorname{div} \xi := \partial_x \xi(x,y,z) + \partial_y \xi(x,y,z) + \partial_z \xi(x,y,z)$ where $\xi \in Y$. For the ADMM, the Laplace operator in equation (6.14) is discretized as $\Delta := -\nabla^T \nabla$. Both algorithms were parallelized on recent graphics hardware by using the CUDA framework.

Projection Scheme. An orthogonal projection on the set of volume compliant surfaces U_V has to be done in both the primal-dual scheme with iterative projection as well as in the ADMM algorithm. In the following we will devise this step analytically. The Euclidean projection of u' onto the relaxed set U_V (see (6.1)) can be described as the following optimization problem:

$$\min_u \frac{1}{2} \int_\Omega \left\| u - u' \right\|^2 dx \quad \text{s.t.} \quad \int_\Omega u \, dx = V \ . \tag{6.26}$$

By introducing the Lagrange multiplier $\lambda \in \mathbb{R}$ and calculating the partial derivatives of the corresponding Lagrangian function we obtain the following extremality conditions:

$$0 = u - u' + \lambda \quad \forall x \in \Omega \tag{6.27}$$

$$0 = \int_\Omega u \, dx - V \tag{6.28}$$

Inserting (6.27) into (6.28) results in a solution for λ

$$\lambda = \frac{\int_\Omega u' dx - V}{\int_\Omega dx} \ , \tag{6.29}$$

back substituting into 6.27 yields

$$u = u' + \Big(\frac{V - \int_\Omega u' \, dx}{\int_\Omega dx} \Big) \cdot \mathbf{1}_\Omega \tag{6.30}$$

as a simple update scheme for the volume projection in which $\mathbf{1}_\Omega$ is an indicator function being 1 at every point $x \in \Omega$ and 0 elsewhere. Intuitively, this means that the residual volume is evenly distributed over all function values of u in Ω.

6.2.4 Thresholding Scheme and Optimality Bound

The optimization algorithm above computes a globally optimal solution u_{opt} of the relaxed primal-dual reconstruction problem (6.5). The question remains how a solution to the binary problem can be obtained from there and how these solutions relate to each other energetically. Unfortunately, no thresholding theorem holds, which would imply the binary optimality of the thresholded relaxed optimum for arbitrary thresholds such as in [22]. We can, nevertheless, construct a binary solution u_{bin} as follows:

Proposition 8. The relaxed solution can be projected onto the set of binary functions in such a way that the resulting binary function preserves the user-specified volume V.

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

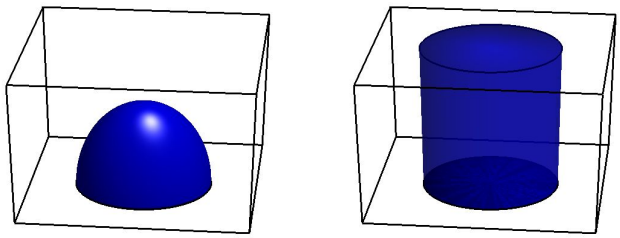


Figure 6.1: The two cases considered in the analysis of the compactness of solutions. *Left:* a hemispherical concentration of the material. *Right:* the material is distributed uniformly over the volume.

Concerning an optimality bound the following holds:

Proposition 9. Let u^r_{opt} be the globally optimal solution of the relaxed energy and u_{opt} the (unknown) globally optimal solution of the binary problem. Then

$$E(u_{\text{bin}}) - E(u_{\text{opt}}) \leq E(u_{\text{bin}}) - E(u^r_{\text{opt}}) \ . \tag{6.31}$$

A proof for this bound was shown in Chapter 2.

6.3 A Theoretical Analysis of the Compactness of Solutions

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 a posteriori optimality bounds, one may take a closer look at the given problem and ask why the relaxed solution u should favor the emergence of solid objects rather than distribute the assigned volume equally over all voxels within the reconstruction domain.

In the following, an analytical proof is derived that the proposed functional energetically prefers compact solutions. For simplicity, we will consider the case that the object silhouette in the image is a disk. And we will compare the two extreme cases of all volume concentrated within a ball (a known solution of the Cheeger problem) compared to the case that the same volume is distributed equally over the whole feasible space (namely a cylinder) – see Figure 6.1. Note that in the following proof it suffices to consider the volume only on one side of the silhouette.

Proposition 10. Let u_{sphere} denote the binary solution which is 1 inside the sphere and 0 outside – Figure 6.1, left-hand side – and let u_{cyl} denote the non-binary solution which is uniformly distributed (i.e. constant) over the entire cylinder – Figure 6.1, right-hand

side. Then we have

$$E(u_{\text{sphere}}) < E(u_{\text{cyl}}), \tag{6.32}$$

independent of the height of the cylinder.

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

$$E(u_{\text{sphere}}) = \int |\nabla u_{\text{sphere}}| d^3x = 2\pi R^2. \tag{6.33}$$

If instead the volume of this hemisphere, i.e. $V = \frac{2\pi}{3}R^3$, is distributed uniformly over the cylinder of height $h \in (0, \infty)$ with volume $\pi R^2 h$, we have

$$u_{\text{cyl}}(x) = \begin{cases} \frac{V}{\pi R^2 h} = \frac{2\pi R^3}{3\pi R^2 h} = \frac{2}{3}\frac{R}{h} & \text{if } x \text{ is inside cylinder} \\ 0 & \text{otherwise.} \end{cases} \tag{6.34}$$

The surface energy of u_{cyl} is given by the area of the cylinder weighted by the respective jump size at the bottom and top disks of the cylinder and for all other boundary voxels:

$$E(u_{\text{cyl}}) = \int |\nabla u_{\text{cyl}}| d^3x = \left(1 - \frac{2R}{3h}\right)\pi R^2 \; + \; \frac{2R}{3h}(\pi R^2 + 2\pi R h) = \frac{7}{3}\pi R^2 > E(u_{\text{sphere}}). \tag{6.35}$$

□

6.4 Experimental Results

In this section we study the properties and applicability of the proposed interactive single view 3D reconstruction method with volume constraint. Note that both the primal dual and the ADMM optimization method compute the global optimum of the energy in (6.1) and thus yield exactly the same results. Qualitative experiments are presented, which are used to highlight practical and theoretical aspects of the approach. Furthermore, limitations and runtimes of the framework are explored and an experimental comparison to the minimal surface approach with shape prior presented in Chapter 5 is given. Failure cases of the approach will be discussed in the subsequent Chapter 8, in Section 8.5.

6.4.1 Interactively Controlling Volume and Smoothness

Embedding the above presented optimization schemes into the interactive framework described in Chapter 4 is straightforward. User specified parameters include the smoothness weight $g(x)$ in (6.5) and the object volume.

Volume control is implemented as a simple slider indicating the percentage of the volume the reconstruction surface should inscribe. Each time the user alters the volume, a recomputation of the reconstruction surface is triggered. This is illustrated in Figure 6.2.

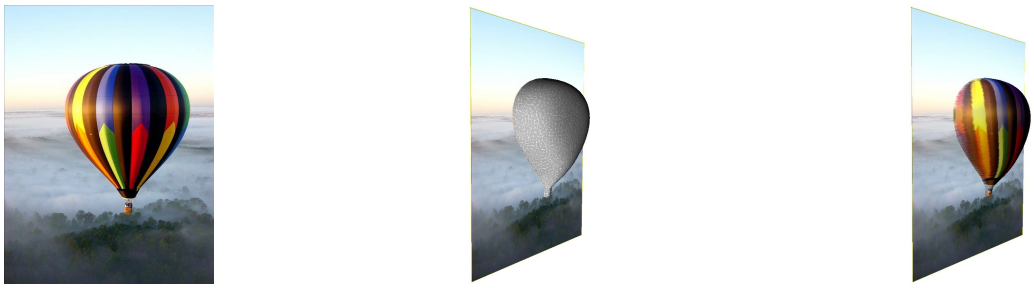
Input Image | Reconstruction | +25% Volume | +35% Volume

Figure 6.2: By interactively increasing the target volume, the reconstruction is intuitively inflated. Broader parts of the silhouette are inflated more than thinner ones.

One disadvantage of minimal surface approaches is that they tend to result in roundish reconstructions. For several objects this is not desirable. We, therefore, employ the user scribble interface as described in Section 5.3.1 for controlling the local smoothness parameter $g(x)$ of the surface are term in (5.6).

6.4.2 Inflation Behavior

Figures 6.3 and 6.4 illustrate how solutions to (6.1) are typically inflated. The former shows that roundish silhouettes lead to balloonish reconstructions. The latter figure shows that thin structures in the silhouette - such as the twig - are inflated less whereas points that have a longer distance to the silhouette boundary are inflated more - see e.g. the bird. Thus, in a way the heuristic shape prior approach of Chapter 5 mimics the Cheeger set method by introducing a silhouette distance transform. Both examples are unweighted, i.e. $g(x) = 1$ for all x.



Input Image | Reconstructed Geometry | Textured Geometry

Figure 6.3: The proposed Cheeger set approach favors minimal surfaces for a user-specified volume. Therefore the reconstruction algorithm is ideally suited to compute smooth, round reconstructions.

Input Reconstruction Different View Geometry

Figure 6.4: Volume inflation dominates where the silhouette area is large (the bird) whereas thin structures (the twigs) are inflated less.

6.4.3 Weighted Minimal Surfaces

Figure 6.5 shows how the weight g can be adapted in order to model high frequency features such as the edges of the box-like shape of the watering can. The user input for this reconstruction is kept simple, although choosing the right scribble parameter requires some experience.

In the air plane example in Figure 6.6 the surface weight was used to create protrusions in the reconstruction. This was achieved by setting the parameter of the wing scribble to a very low value (0.03 in this case). Notably, inducing protrusions in this way is not possible in the approach of Chapter 5, since the shape prior of energy (5.6) inhibits the flexibility of the surface.

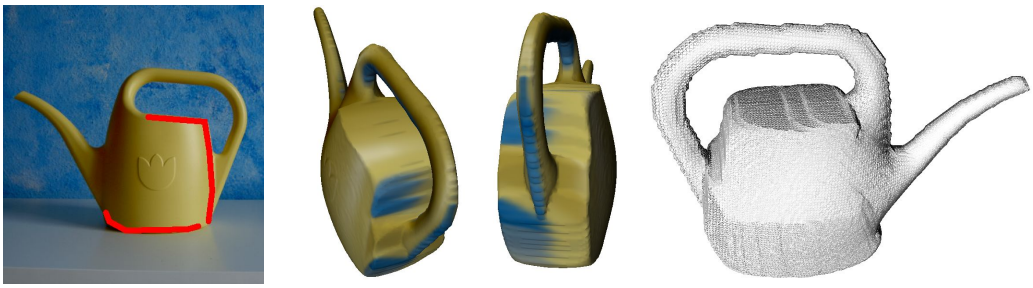


Image with User Input Reconstructions Geometry

Figure 6.5: The proposed approach allows to generate $3D$ models with sharp edges marked by the user as locations of low smoothness in the image. Along the red user strokes the local smoothness weighting function g is decreased.

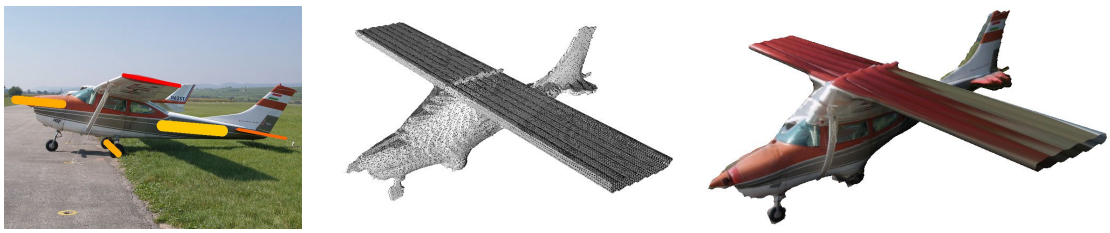


Image with User Strokes Reconstructed Geometry Textured Geometry

Figure 6.6: An example for a minimal surface with user defined volume and local smoothness adaptation for creating protrusions. Colored lines in the input image mark user input, which locally alters the surface smoothness. Red marks decrease, yellow marks increase the smoothness.

6.4.4 Compactness of the Solution

It was theoretically shown in Section 6.3 that the relaxed problem formulation of 6.1 energetically favors compact solutions. In Figure 6.7 an experimental evaluation of the compactness of the reconstructions is shown for several examples. Voxel occupancies of the optimal relaxed surface function u are visualized. Values range from 0 (completely transparent) to 1 (white). We can see that most voxels are either 0 or 1 which demonstrates the compactness of the solutions.

6.4.5 Shape Prior versus Volume Prior

Figure 6.8 compares multiple reconstruction examples of the proposed Cheeger set approach with those of the shape prior method from Chapter 5. The experiments show that the reduced number of parameters does not have a negative impact on the reconstruction quality for most of the shown examples. For some examples the shape priors inflicts a strong bias on the reconstructions (see e.g. the pyramidal shapes of the *egret* and the *giraffe*). Increasing the smoothness parameter λ in the energy (5.6) will mitigate the influence of the shape prior. However, with higher smoothness the result tends to be less voluminous making it hard to achieve ball-like shapes. This relationship is shown in Figure 6.9. The shape prior also does not allow for protrusions as induced by the weight $g(x)$ described above, since the data term is too strong.

6.4.6 Runtime Comparison for PD and ADMM

Table 6.1 compares the ADMM and the Lagrange Multiplier version of the PD algorithm with respect to the runtime. Both algorithms were run until the RMSE error $\|u - u^*\| 2/\sqrt{|\Omega|}$ dropped below a threshold of $\varepsilon = 5.0 * 10^{-3}$. The optimal solution u^*

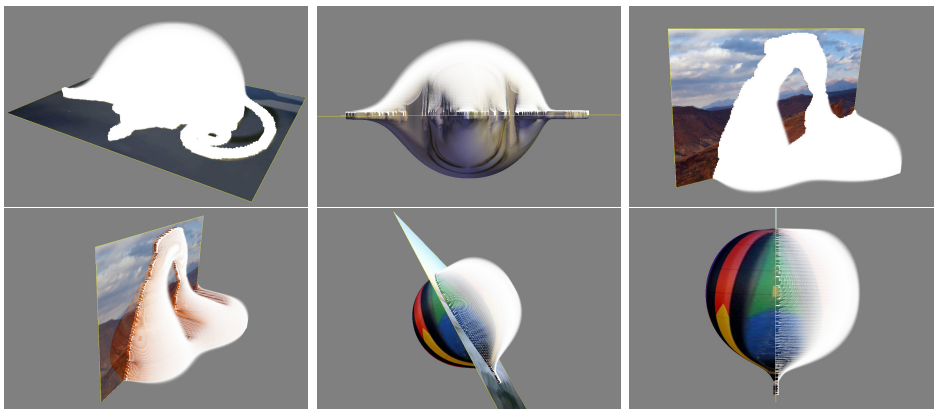


Figure 6.7: Voxel occupancies visualized for the optimal relaxed solution computed on several examples. Occupancies range from 0 (transparent) to 1 (white). Almost all voxels are either 1 or 0 demonstrating the compactness of the reconstructions. Additionally, some images contain the iso-surface, i.e. the final reconstruction, that was generated by the described thresholding scheme. One can see that the relaxed solution is reasonably close to the final iso-surface for most cases, which further supports the compactness thesis.

	Runtime in Seconds		Number of Iterations	
$\|\Omega\|$	PD	ADMM	PD	ADMM
sm	0.15 ($\pm$ 0.10)	0.10 ($\pm$ 0.05)	1015 ($\pm$ 618)	300 ($\pm$ 152)
med	13.47 ($\pm$ 11.64)	8.35 ($\pm$ 7.32)	5716 ($\pm$ 5664)	1471 ($\pm$ 1399)
big	115.30 ($\pm$ 113.60)	79.16 ($\pm$ 91.80)	11724 ($\pm$ 12136)	3171 ($\pm$ 3776)

Table 6.1: Average and standard deviation of the runtime and number of iterations for the PD and ADMM algorithm over 18 different single view 3D reconstruction examples. The results show that ADMM performs better with respect to the runtime.

of the energy minimization problem was computed by letting the algorithms run for a sufficiently long time. For the ADMM parameters $\tau = 1$, $\alpha = 1.5$ were chosen and for the PD algorithm it was set $\theta = 1$ with preconditioned step sizes. All numerical experiments were carried out on a PC with a 3.4GHz Intel i7-3770 CPU with 32GB RAM and a NVIDIA GeForce GTX680 graphics card on a Linux distribution.

For experiments 18 test images were used and the reconstruction was computed for each of them using PD and ADMM. To examine how the algorithms scale with the resolution of the reconstruction domain Ω three different resolutions were used. A rather

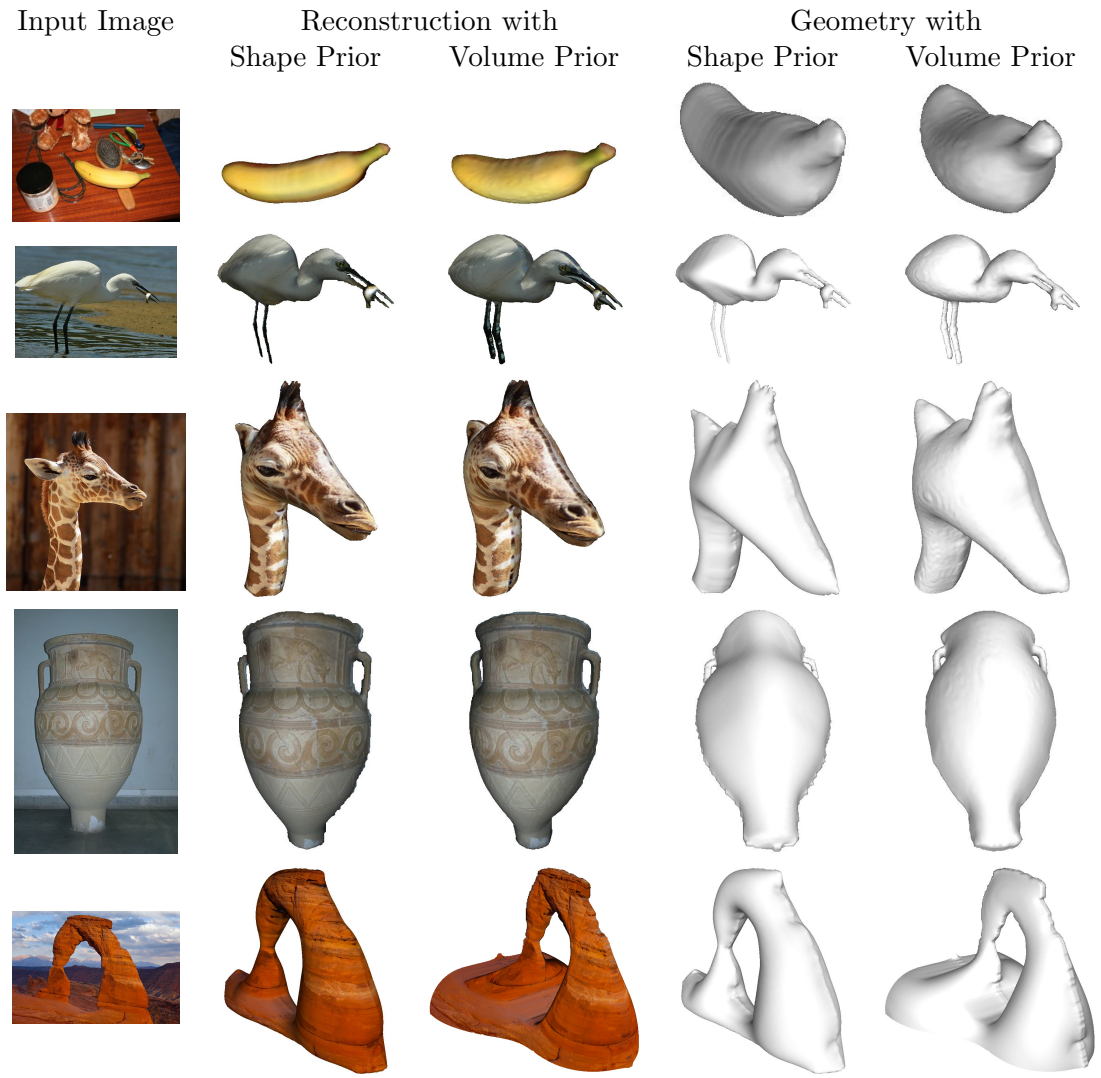
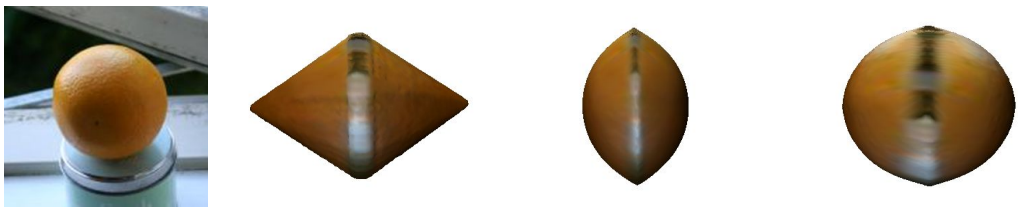


Figure 6.8: Direct comparison of the reconstruction method with shape prior given in Chapter 5 to the Cheeger set formulation in this chapter. The results demonstrate the superior reconstruction performance of the Cheeger set formulation in contrast to the shape prior approach which often yields biased solutions (see egret and giraffe).



Input Image | Data Term as Shape Prior | Reconstruction with Shape Prior | Reconstruction with Volume Prior

Figure 6.9: Using a silhouette distance transform as shape prior the relation between the data term and the final reconstruction is not obvious for a user. In contrast, the Cheeger set approach results in more intuitive solutions controlled only by the volume parameter.

coarse resolution with few ($\approx 3.4{\cdot}10^4$) voxels (sm), an intermediate ($\approx 8.0{\cdot}10^5$) resolution (med) and a high resolution with a large number ($\approx 3.6 \cdot 10^6$) of voxels (big).

For each resolution the average runtime and standard deviation over all test images is given in Table 6.1. Figure 6.10 shows an example for the different performances of PD and ADMM on the balloon reconstruction image in Figure 6.3. The horizontal line indicates the termination criterion where the desired accuracy of the solution is reached. Results on the other test images look similar. From the results we can conclude that the ADMM algorithm converges faster than the PD algorithm for the 3D reconstruction problem instance of the minimal partition problems with volume constraints.

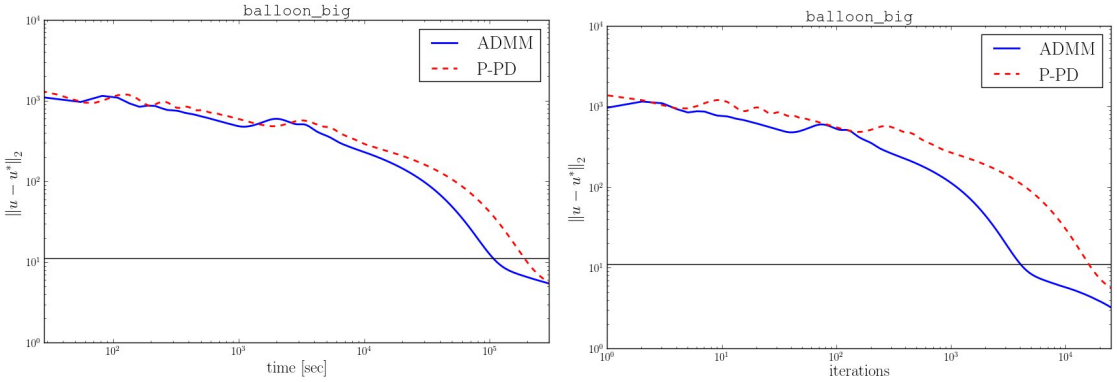


Figure 6.10: Exemplary convergence of the ADMM and the Lagrange Multiplier version of the PD algorithm. The horizontal line indicates the termination criterion based on the accuracy of the algorithm (vertical axis). The ADMM algorithm converges significantly faster than the PD algorithm.

Example	Zhang et al. [114]	Igarashi et al. [53]	Prasad et al. [86]	Shape Prior	Volume Prior
Banana	20 min	< 1 min	10 min	5 min	< 1 min
Fish	15 min	< 1 min	2 min	8 min	1 min
Dory	40 min	< 1 min	5 min	7 min	1 min
Vase	20 min	< 1 min	2 min	13 min	4 min
Squash	12 min	< 1 min	2 min	2 min	1 min
Orange	14 min	< 1 min	<1 min	3 min	< 1 min
Ostrich	30 min	< 1 min	15 min	7 min	2 min
Donut	55 min	< 1 min	10 min	3 min	1 min
Jelly Bean	15 min	< 1 min	2 min	4 min	1 min
Teapot	35 min	< 1 min	20 min	15 min	4 min

Table 6.2: Approximate modeling times for a user with intermediate experience for all five methods based on the examples in Figures 6.11, 6.12 and 6.13. In combination with the reconstruction results this table reveals significant differences in the efficiency of the methods.

6.5 Comparison to State-of-the-art Single View Approaches

In this section the minimal surface reconstruction algorithms presented so far are compared to the most relevant related single view 3D reconstruction approaches that have been published over the last years. Our focus lies on methods that aim for the reconstruction of curved surfaces. In particular, we discuss the approaches by Zhang et al. [114], Prasad et al. [86] and Igarashi et al. [53]. Although Igarashi et al. is rather a modeling tool it inflates object contours and is thus strongly related to silhouette based 3D reconstruction.

For all experiments, the single view modeling tool by Zhang et al. [114] and the SmoothTeddy software - which is based on several works by Igarashi et al. [53, 51, 52] - was used. Both of them are publicly available for download. The reconstruction results by Prasad et al. are taken from [85, 86, 84].

Figures 6.11, 6.12 and 6.13 show the reconstruction results of all five methods on ten different examples, covering various issues such as object shape, topology, viewing angle and image type.

User Input and Modeling Times. The measured modeling times are listed in Table 6.2. Times include all the preprocessing (such as silhouette extraction) and respective user input and only excludes the coloring of the model in the approach by Igarashi et al. [53].

Among the reconstruction approaches, Zhang et al. and Prasad et al. exhibit the

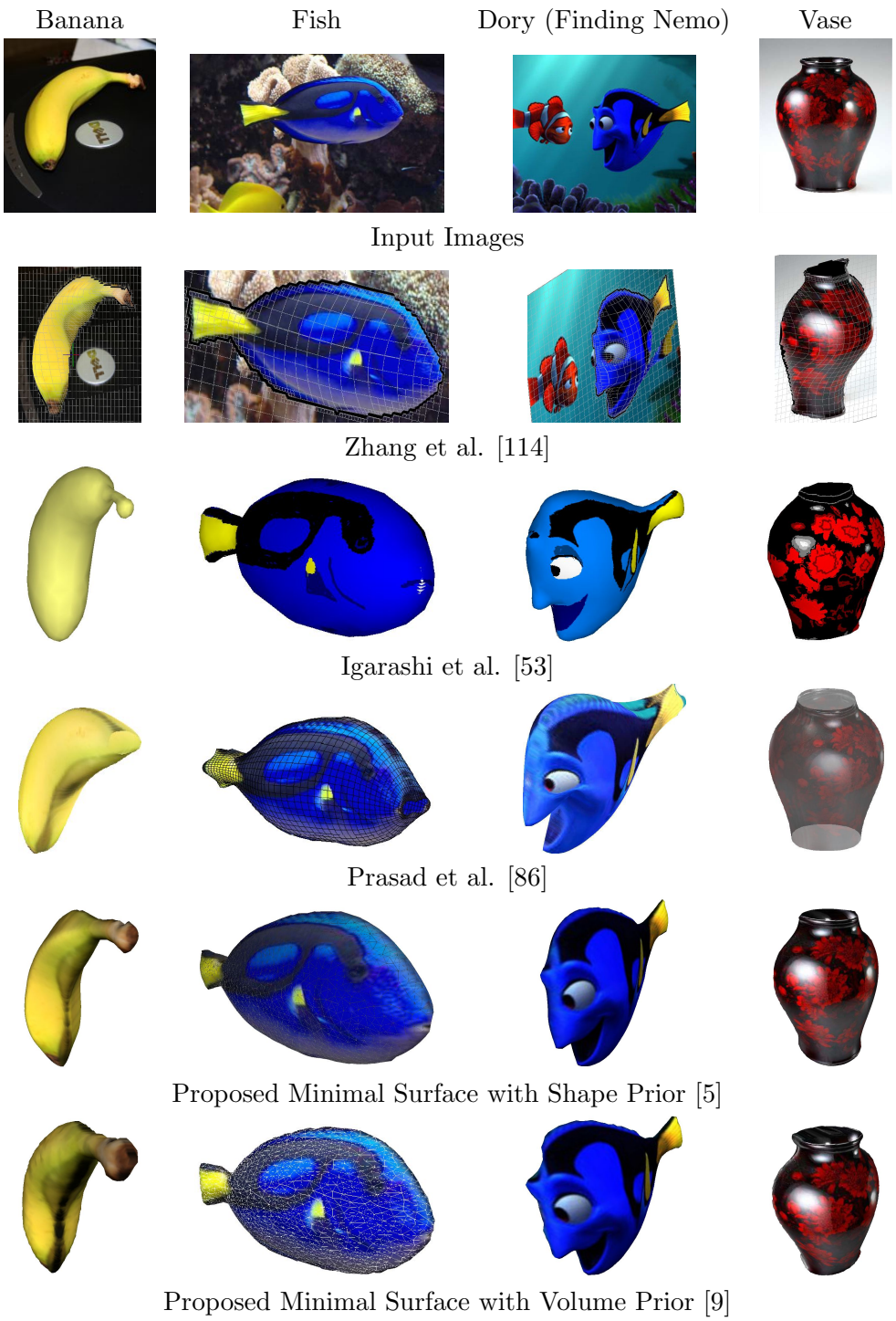


Figure 6.11: Experimental comparison of several methods for curved object reconstruction. The results obtained by the method by Prasad et al. are taken from [84].

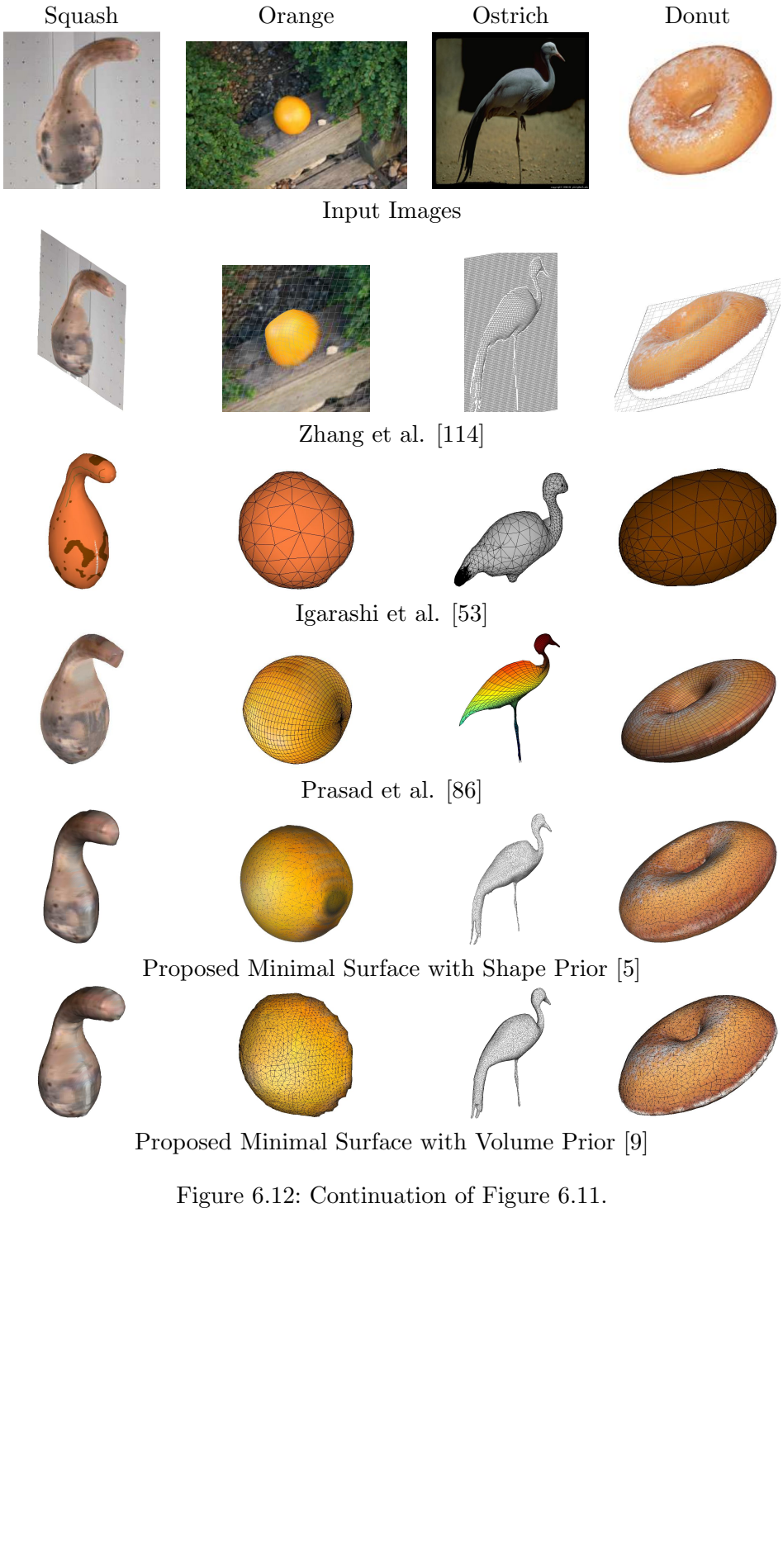


Figure 6.12: Continuation of Figure 6.11.

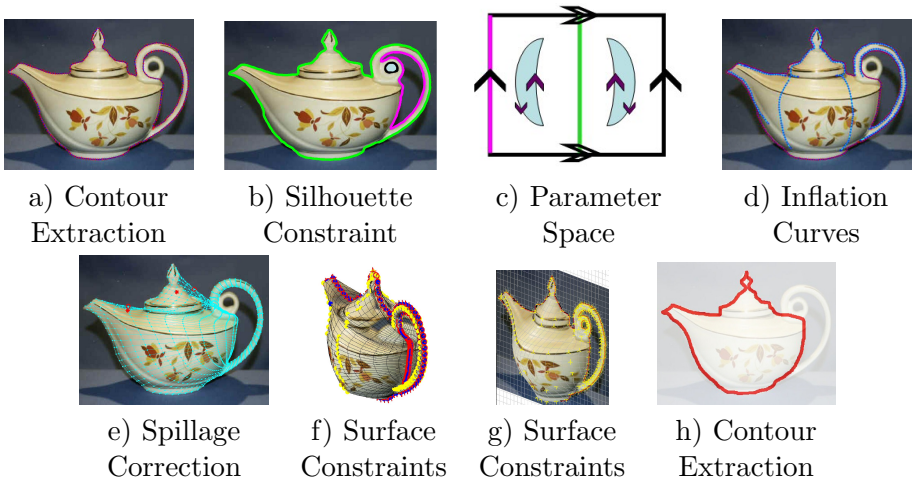
Figure 6.13: Continuation of Figure 6.12.

highest demand on user experience and modeling times, see Table 6.2. Although Zhang et al. and Igarashi et al. rather belong to the category of modeling tools, they differ largely in the required user effort. This is because with the method of Zhang et al. the user has a variety of choices for surface manipulations. These include *position and normal constraints* for fixing depth values and normal directions at selected points, *discontinuity constraints* for modeling sharp edges, *planar region constraints* for modeling planar surfaces and *manual mesh-subdivision* for increasing the resolution locally. Usually many of these constraints are necessary for reasonable reconstructions leading to modeling times of several minutes to hours even for experienced users. As moderately experienced users, 20-40 minutes were spent for each of the examples shown in Figure 6.11 - 6.13.

Similarly, the method by Prasad et al. requires concise input and expert knowledge (see Figure 6.14): the user has to assign parts of contour lines to lines in the parameter space, which becomes harder for objects of higher genus. Moreover, for volume inflation one needs to define a set of *interpolation constraints*. In subsequent steps the user can add further constraints for allowing *surface creases*, which are also represented as lines in parameter space. During optimization the constraints for the silhouette consistency may be violated, which makes it necessary to manually add a so called *spillage correction*.

Surface Representation and Topology. A comparison to Zhang et al. is problematic in that only height maps can be modeled as opposed to closed surfaces. Also, no segmentation is done of the object silhouette, which is why the background is part of the height field. Igarashi et al. allow only for genus zero surfaces as they model meshes explicitly. This can lead to mesh inconsistencies as can be seen for the neck of the ostrich in Figure 6.12. To allow for holes during modeling extensions such as FiberMesh [78] were proposed. Prasad et al. can theoretically represent surfaces of higher genus, but are practically limited to genus two (e.g. the *teapot*) due to the parametric surface representation. The proposed minimal surface approaches are the only ones that can reconstruct surfaces of arbitrary topology. However, even though they are capable of representing arbitrary surfaces, in practice reconstructions will be height fields mirrored along the symmetry plane, e.g. there will be no self-occlusion in viewing direction due to the minimal surface assumption. This problem will be addressed in Chapter 8.

Prasad et al. compute a non-trivial surface parametrization from user input (see Section 3.2.1). This is beneficial as the parametrization can encode unique characteristics of the object in combination with non-planar contour generators. Drawbacks are the required user effort. In addition, uniformly distributed points in parameter space can end up being stretched or compressed on the object surface leading to oscillations such as those in the teapot in Figure 6.13 or difficulties in modeling thin structures such as the legs of the ostrich in Figure 6.11.



a) Contour Extraction | b) Silhouette Constraint | c) Parameter Space | d) Inflation Curves

e) Spillage Correction | f) Surface Constraints | g) Surface Constraints | h) Contour Extraction

Figure 6.14: a)-f): User input required by Prasad et al. [86] for the genus 1 version of the teapot example (figures are taken from [84]), a) Contour extraction, b) Lines of the contour have to be related to lines in the parameter space in c), d) Input for the inflation heuristic, e) Spillage correction, f) Final reconstruction with inflation constraints in yellow, g) User input for the method by Zhang et al. Yellow crosses are position constraints, red ones denote discontinuity constraints, h) User input for the method by Igarashi et al. which consists only of the contour extraction.

Assumptions on the Input Image. The proposed minimal surface approaches and the method by Igarashi et al. both make the assumption that the object in the input image is symmetric with respect to the image plane. This implies that the contour generator is planar. If this assumption is violated, reconstructions can still look pleasing, but are erroneous (see e.g. the *donut*, or the *banana*). The approach by Prasad et al. [86] allows for non-planar contour generators and, thus, in some cases for slightly more general view points than just a side-view. The approach by Zhang et al. works for the broadest range of images but at the same time results in the most unrealistic reconstructions.

Quality. It is difficult to assess the quality of the reconstructions. One criterion is how geometrically close the reconstructions come to the real objects. In that respect Prasad et al. and the minimal surface approaches are clearly dominating. Still there are also subjective differences in quality that are hard to measure. The *teapot* example by Prasad et al., e.g., suffers from surface over-oscillation at the handle, whereas the minimal surface approaches have problems with the reconstruction of the *donut* since the image is not taken from a side perspective. An advantage of the method by Zhang et al. is the large freedom in modeling details such as, e.g., the round shaped eye of the

fish Dory in Figure 6.11 or its side fin which bends away from its body.

The method by Igarashi et al. is the only one not capable of modeling creases and sharp edges. This lowers the quality of some reconstructions such as the tail fin of Dory in Figure 6.11 or the top and bottom of the vase (Figure 6.11) and the teapot (Figure 6.12).

Runtimes. Zhang et al. and Prasad et al. use thin plate spline energies to compute minimal surfaces, which is done in a matter of seconds on modern CPUs as it amounts to solving small systems of linear equations. Igarashi et al. construct their surface by a heuristic, for which probably the lowest computational effort is needed among the discussed approaches. In the proposed minimal surface approaches large systems of variables have to be solved and the computational complexity rises cubically with a growing reconstruction volume. By parallelizing the computation on graphics hardware runtimes between below one second and several seconds are obtained depending on the resolution of the reconstruction volume.

6.6 Conclusion

A second single view 3D reconstruction approach was proposed in this chapter that is based on the idea of Cheeger sets, i.e. minimal surfaces for a fixed user-specified volume. To this end, the minimal surface idea in Chapter 4 was extended by a constraint on the volume contained by the reconstruction surface. The resulting variational problem was solved via convex relaxation and either PD or ADMM algorithms. The proposed thresholding scheme delivers a binary solution that complies with the volume and is provably within bounds from the global optimum.

In an experimental evaluation we showed that the results are qualitatively comparable to the shape prior approach of Chapter 5 but without the drawback of imposing a bias on the reconstruction shape. In this way, the weight of the total variation will have a bigger impact on the surface locally and can be used to model sharp edges and protrusions with the help of simple user strokes. Finally, it was shown that the convex approaches compare equally well or favorably over existing state-of-the-art single view 3D reconstruction methods for curved surfaces.

There are several drawbacks of the Cheeger set approach. The computational complexity rises cubicly with the resolution of the optimization volume, which, therefore, should be kept small. Small resolutions, however, will result in discretization artifacts. Another problem is that the thresholding scheme does not guarantee a global optimum of the combinatorial problem.

Other problems concern the flexibility of the approach and the quality of the results: no additional shape information enters the reconstruction process, which is why firstly, reconstructions tend to be roundish and secondly, they do not allow for more complex

shapes such as occlusions or protuberances, and they are always representable as mirrored height fields.

The latter line of drawbacks will be addressed in Chapter 8, while the next chapter will deal with the remaining issues.

7 A Globally Optimal Minimal Surface Approach Based on Height Fields

We have seen that formulating single view 3D reconstruction as a silhouette based minimal surface problem requires to have some inflation stimulus to avoid trivial flat solutions. Up to this point two approaches were presented, one of which is based on a parametric shape prior (Chapter 5), the other one on a volume constraint (Chapter 6). We found the latter approach favorable in most cases as it is simpler, less heuristic, avoids the shape prior bias and has only one parameter with intuitive effect on the reconstruction surface.

However, the presented globally optimal height field approach comes with a couple of drawbacks, which have already been indicated in the last chapter:

- The volumetric representation is computationally more involved than other state-of-the-art minimal surface based single view 3D reconstruction approaches. For moderate resolution reconstructions the method requires more than a second of computation time despite an efficient GPU-accelerated primal-dual algorithm. As a consequence, higher-resolution 3D models cannot be generated at interactive speeds.
- Although the Cheeger set method was shown to provide smooth and volume-consistent solutions, the algorithm is still based on convex relaxation and thresholding. Consequently, in the absence of a threshold theorem, the method is not guaranteed to provide a globally optimal minimal surface of specified volume. Furthermore, despite the compactness of the solution it is not clear whether subsequent thresholding of the relaxed solution actually leads to a spatially coherent structure (rather than a scattered set of voxels).
- Low resolutions of the optimization volume lead to discretization artifacts since depth is discretized into a finite number of voxels.

Contributions. In this chapter a reformulation of the Cheeger set approach is proposed which solves the issues outlined above. The method exploits the observation that the surfaces computed by the fully volumetric Cheeger set approach are representable as height fields (mirrored along the image plane to obtain a closed reconstruction). More precisely the contributions are the following:

- The proposed single view 3D reconstruction approach is based on a height field representation. As a consequence, depth values are represented continuously.
- Requirements on memory and computation time are substantially reduced with a quadratic rather than a cubic complexity. This enables us to compute high resolution reconstructions based on high resolution input images and silhouettes within interactive runtimes.
- In contrast to the fully volumetric Cheeger set approach the proposed method does not require a convex relaxation or thresholding. As a consequence, the algorithm provably computes globally optimal silhouette-consistent minimal surfaces of a specified volume.

This approach was published in [4].
In the following we will introduce a proper mathematical formulation for a height field approach to volume-constrained minimal surfaces and will consider suitable methods for optimizing the problem efficiently. Finally, we will evaluate the method qualitatively, compare it to related approaches and examine runtimes.

7.1 A Globally Optimal Minimal Surface Approach Based on Height Fields

7.1.1 Variational Formulation

The paradigm shift in the following approach is to avoid the implicit formulation of the reconstruction surface. Instead, we assume that the surface can be fully described as a height map defined on the silhouette domain Σ (see Chapter 4)

$$u : \Sigma \to \mathbb{R}_0^+, \quad \Sigma \subset \Omega \ . \tag{7.1}$$

The function u assigns a depth value $u(x, y)$ to each point $(x, y) \in \Sigma$ of the silhouette which is a subset of the image domain Ω. As shown in the schematic plot in Figure 7.1, an infinitesimal surface area element dA of the surface represented by the function u is given by

$$dA = \left| \begin{pmatrix} dx \\ 0 \\ u_x\, dx \end{pmatrix} \times \begin{pmatrix} 0 \\ dy \\ u_y\, dy \end{pmatrix} \right| = \sqrt{1 + |\nabla u|^2}\, dx\, dy \ . \tag{7.2}$$

The overall area of the surface u is obtained by summing over all its surface elements. Since we wish to minimize the surface area we define the following energy

$$E(u) = \int dA = \int_\Sigma \sqrt{1 + |\nabla u|^2}\, d^2x \ . \tag{7.3}$$

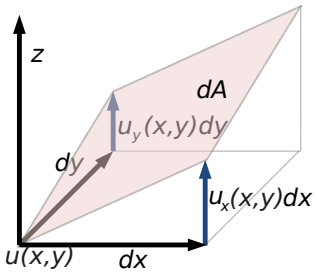


Figure 7.1: Area of an infinitesimal surface element dA computed in (7.2) based on the partial derivatives of u.

For brevity, d^2x now denotes a two dimensional variable of integration. The task of computing a minimal surface enclosing the volume V is, therefore, equivalent to the following problem

$$\min_{u\in\mathcal{C}} E(u), \quad \text{with} \quad \mathcal{C} = \left\{ u \in \mathcal{C}^1(\Omega;\mathbb{R}_0^+) \;\middle|\; \int_\Sigma u\, d^2x = V \right\} \quad . \tag{7.4}$$

Proposition 11. The two-dimensional fixed volume minimal surface problem defined in (7.4) is convex.

Proof. The volume constraint on u is obviously a linear constraint, thus the optimization domain is convex. Moreover, the functional E is convex because for any functions u_1 and u_2 and any $\alpha \in (0,1)$ it holds:

$$\begin{aligned}
&E(\alpha u_1 + (1-\alpha)u_2) \\
&\quad = \int \sqrt{1 + |\big(\alpha\nabla u_1 + (1-\alpha)\nabla u_2\big)|^2}\, d^2x \\
&\quad = \int \left\| \begin{pmatrix} \alpha\nabla u_1 + (1-\alpha)\nabla u_2 \\ \alpha + (1-\alpha) \end{pmatrix} \right\| d^2x \\
&\quad = \int \left\| \alpha \begin{pmatrix} \nabla u_1 \\ 1 \end{pmatrix} + (1-\alpha) \begin{pmatrix} \nabla u_2 \\ 1 \end{pmatrix} \right\| d^2x \\
&\quad \leq \int \alpha \left\| \begin{pmatrix} \nabla u_1 \\ 1 \end{pmatrix} \right\| + (1-\alpha) \left\| \begin{pmatrix} \nabla u_2 \\ 1 \end{pmatrix} \right\| d^2x \\
&\quad = \int \alpha\sqrt{1+|\nabla u_1|^2} + (1-\alpha)\sqrt{1+|\nabla u_2|^2}\, d^2x \\
&\quad = \alpha E(u_1) + (1-\alpha)E(u_2)
\end{aligned} \tag{7.5}$$

□

In contrast to the Cheeger set formulation (6.1) proposed in the last chapter, the two-dimensional height field formulation is convex. As a consequence, we do not need to revert to the generally suboptimal strategy of finding a convex relaxation and subsequent thresholding. Instead, we can directly compute globally optimal solutions by solving (7.4).

7.1.2 Theoretical Analysis of Solutions for the Disc Case

Before we detail how this can be done efficiently in the next section, let us undertake a theoretical analysis of basic solutions to problem (7.4). This will be examined for the case where the silhouette Σ is a disc. Consider the following

Proposition 12. Given a silhouette the shape of a disc with radius R, the solution to problem (7.4) for given volume $V < \frac{2}{3}\pi R^3$ is a sphere segment. For $V = \frac{2}{3}\pi R^3$ the minimal surface becomes a hemisphere of radius R.

Proof. For the special case of a disc, problem (7.4) becomes rotationally invariant and we can revert to a one-dimensional formulation. To this end, we substitute

$$u(x) = v(|x|) \qquad \text{and} \qquad \nabla u(x) = v'(|x|)\frac{x}{|x|} \tag{7.6}$$

with $v : \mathbb{R} \to \mathbb{R}$, so that integral (7.3) reads as

$$\int_\Sigma \sqrt{1 + |\nabla u|^2}\, d^2x = \int_\Sigma \sqrt{1 + \left| v'(|x|)\frac{x}{|x|} \right|^2}\, d^2x = \int_\Sigma \sqrt{1 + |v'(|x|)|^2}\, d^2x \tag{7.7}$$

From this, we can reduce the constant volume minimal surface problem to its one-dimensional form (without loss of generality we will assume a disc of unit radius $R = 1$)

$$\min_{v \in \mathcal{C}} c \cdot \int_0^1 r\sqrt{1 + v'(r)^2}dr \quad \text{s.t.} \quad \mathcal{C} = \left\{ v \,\middle|\, c \cdot \int_0^1 rv(r)\, dr = V \right\} \ , \tag{7.8}$$

where c is the area of a unit circle. We add the constraint as a Lagrange multiplier λ to the energy functional. Then, the optimality conditions for a solution composes of the Euler-Lagrange equations on the one hand and the derivative with respect to λ on the other:

$$-\frac{v'}{\sqrt{1 + v'^2}} - r\left(\frac{v''}{\sqrt{1 + v'^2}} - \frac{v'^2 v''}{\sqrt{1 + v'^2}^{\frac{3}{2}}} \right) + \lambda = 0 \quad \forall r \in [0, 1] \tag{7.9}$$

$$c \cdot \int rv(r) - V = 0 \tag{7.10}$$

Now, a sphere segment is given by the function

$$v(r) := \sqrt{R_V^2 - r^2} - m_V \tag{7.11}$$

where R_V is the radius and m_V the shift in y-direction that depend on the given volume V. It is easy to see that for given $V \leq \frac{2}{3}\pi$, R_V and m_V are determined uniquely (for a sagitta of length two, i.e. $v(-1) = v(1) = 0$). Since we choose $v(r)$ to describe a sphere segment of volume V, Equation (7.10) is trivially satisfied. In order to show that $v(r)$ satisfies condition (7.9) we plug in the derivatives

$$v'(r) = -r(R_V^2 - r^2)^{-\frac{1}{2}} \tag{7.12}$$

$$v''(r) = -(R_V^2 - r^2)^{-\frac{1}{2}} - r^2(R_V^2 - r^2)^{-\frac{3}{2}} \tag{7.13}$$

and end up with

$$\begin{aligned}
&- \frac{-r(R_V^2 - r^2)^{-\frac{1}{2}}}{(1 + r^2(R_V^2 - r^2)^{-1})^{\frac{1}{2}}} - r\left(\frac{-(R_V^2 - r^2)^{-\frac{1}{2}} - r^2(R_V^2 - r^2)^{-\frac{3}{2}}}{(1 + r^2(R_V^2 - r^2)^{-1})^{\frac{3}{2}}} + \right. \\
&\qquad \left. \frac{r^2(R_V^2 - r^2)^{-\frac{3}{2}} + r^4(R_V - r^2)^{-\frac{5}{2}}}{(1 + r^2(R_V^2 - r^2)^{-1})^{-\frac{3}{2}}}\right) + \lambda r \stackrel{!}{=} 0 \\
\Leftrightarrow\ & r - r\left(-\frac{1}{R_V} - \frac{r^2}{R_V(R_V^2 - r^2)} + \frac{r^2}{R_V^3} + \frac{r^4}{R_V^3(R_V^2 - r^2)}\right) + \lambda r \stackrel{!}{=} 0 \\
\Leftrightarrow\ & r - r\left(-\frac{1}{R_V} - \frac{r^2}{R_V^3}\left(\frac{(R_V^2 - r^2)}{(R_V^2 - r^2)} - 1\right)\right) + \lambda r \stackrel{!}{=} 0 \\
\Leftrightarrow & \left(1 + \frac{1}{R_V}\right) r + \lambda r \stackrel{!}{=} 0
\end{aligned} \tag{7.14}$$

which can be satisfied for all r by choosing $\lambda = -\left(1 + \frac{1}{R_V}\right)$.

□

7.2 Efficient Optimization

Globally optimal minimal solutions of the convex problem (7.4) can be obtained by solving the Euler-Lagrange extremality condition given by the partial differential equation

$$\frac{dE}{du} = -\mathrm{div}\left(\frac{1}{\sqrt{1 + |\nabla u|^2}}\,\nabla u\right) = 0 \ . \tag{7.15}$$

Solving this equation with respect to the function u will yield the global optimum due to the convexity of problem (7.15). Note, however, that theoretically there may be several

solutions of the same energy. Condition (7.15) is a nonlinear diffusion equation which is similar to the well-known model by Perona and Malik [80] for edge-preserving image smoothing, but with a different diffusivity $g(x) = 1/\sqrt{1 + |\nabla u|^2}$ which was proposed by Charbonnier et al. [24].

Deriving the diffusion (7.15) from Equation (7.4) therefore provides a geometric interpretation of the Perona-Malik diffusion with the Charbonnier-diffusivity: in image diffusion the image gray values can be interpreted as a height map whose surface area is being minimized as the diffusion proceeds (see also [95] for more details).

However, here we use Equation (7.4) in a completely different setting. Instead of using a data term we impose a global volume constraint and special boundary conditions which are derived from the input silhouette. In the following we will describe how these constraints are chosen and incorporated into the numerical optimization of Equation (7.15).

7.2.1 Numerical Schemes for Finding the Global Optimum

There are multiple optimization schemes for solving problem (7.4). In the following we will present four different numerical algorithms. We will explain their theoretical properties, point out their relations, detail their implementation and compare their performance in the context of single view 3D reconstruction. The first numerical algorithm for solving the minimal surface energy (7.3) was given as early as 1967 by Paul Concus [28], however, he did not consider a volume constraint in his formulation.

Projected Gradient Descent. A simple iterative scheme for solving constrained energy minimization problems is projected gradient descent, which is similar to gradient descent as described in Chapter 2, only that in each iteration the solution is projected onto the feasible set $\mathcal{C}$, i.e.

$$u_{t+1} = \text{prox}_{\chi_{\mathcal{C}}}\left(u_t - \tau \cdot \frac{dE}{du}\right) \tag{7.16}$$

where τ is the step size and $\text{prox}_{\chi_{\mathcal{C}}}$ is the Euclidean projection onto the convex set C. In each step the energy of the solution will decrease. The algorithm converges when the solution does not change anymore. Since the minimization problem (7.4) is convex, the gradient descent method will converge to the global optimum of the constrained energy. We will discuss later how the projection can be computed analytically. For faster convergence, the projection is usually performed only every l-th iteration with $l > 1$.

Fast Iterative Shrinkage-Thresholding Algorithm - FISTA. As reviewed in Chapter 2, the fast iterative shrinkage-thresholding algorithm [14] is a splitting method capable of solving minimization problems with non-differentiable components. Energy (7.4) has no

non-differentiable term which in our case makes FISTA equivalent to the Nesterov scheme [79] of which it is a generalization. Nesterov introduced an over-relaxation step, which leads to a faster convergence by over-shooting the current solution into the direction of its recent change. The over-relaxation step size is adapted in each iteration by a special scheme which can be shown to guarantee a global convergence rate of $O(1/t^2)$, t being the iteration.

$$\begin{cases} u_t & = \bar{u}_t - \tau \cdot \frac{1}{L} \frac{dE(\bar{u}_t)}{du} \\ \tau_{t+1} & = \frac{1}{2}\left(1 + \sqrt{1 + 4\tau_t^2}\right) \\ \bar{u}_{t+1} & = \text{prox}_{\chi_C}\left(u_t + \frac{\tau_t - 1}{\tau_{t+1}}(u_t - u_{t-1})\right) \end{cases}$$

Here L is a Lipschitz constant of the functional E. Again, the projection in the last step is not required in each step.

Primal-Dual. The first order primal-dual scheme presented in [21] and reviewed in Chapter 2 solves problems of the form

$$\inf_{u \in \mathcal{V}} F(Ku) + G(u) \quad . \tag{7.17}$$

We can apply the algorithm to the minimal surface problem (7.4) by substituting the variables as follows:

$$\begin{aligned} K : \mathcal{C}^1(\Omega; \mathbb{R}) \to \mathcal{C}^1(\Omega; \mathbb{R}^2), u \mapsto \nabla u \\ F : \mathcal{C}^1(\Omega; \mathbb{R}^2) \to \mathbb{R}, v \mapsto \int \sqrt{1 + < v, v >}\, d^2x \\ G = 0 \quad . \end{aligned} \tag{7.18}$$

The following Proposition derives the primal-dual formulation of the height field minimal surface approach with volume constraint for single view 3D reconstruction.

Proposition 13. The corresponding primal-dual problem for the energy in (7.4) reads as

$$\inf_u \sup_{\substack{\xi : |\xi|_\infty \leq 1 \\ \lambda \in \mathbb{R}}} \int_\Sigma < \nabla u, \xi > + \sqrt{1 - < \xi, \xi >}\, d^2x + \lambda \left(\int_\Sigma u \, d^2x - V \right) \quad . \tag{7.19}$$

Proof. With substitutions (7.18) the primal-dual formulation of problem (7.17) takes the following form (cf. Equation (2.19))

$$\min_{u \in \mathcal{C}} \sup_\xi < \nabla u, \xi > - F^*(\xi) \quad . \tag{7.20}$$

It remains to calculate the Legendre-Fenchel dual F^* to the function F. To this end, we seek to solve the following problem

$$F^*(\xi) = \sup_v \int < v, \xi > -\sqrt{1+ < v, v >}\ d^2x. \tag{7.21}$$

The Euler-Lagrange equations to this problem are given by $\xi(1+ < v, v >)^{\frac{1}{2}} = v,\ \forall x \in \Omega$ for which the larger solution is $v = \frac{\xi}{\sqrt{1-<\xi,\xi>}}$. Inserting this solution into Equation (7.21) yields

$$F^*(\xi) = \begin{cases} -\int \sqrt{1- < \xi, \xi >}\ d^2x, & \text{if } \ |\xi|_\infty \leq 1 \\ \infty, & \text{otherwise} \end{cases}. \tag{7.22}$$

Finally, the volume constraint in (7.4) is added to (7.19) by means of a Lagrange multiplier λ. This concludes the proof. □

A straightforward application of the primal-dual algorithm to problem (7.19) yields the following optimization scheme.

$$\begin{cases} \xi^{k+1} = \Pi_{|\xi(x)|_2 \leq 1} \left(\xi^k + \tau_\xi \cdot (\nabla \bar{u}^k - \frac{\xi^k}{\sqrt{1-<\xi^k,\xi^k>}}) \right) \\ \lambda^{k+1} = \lambda^k + \tau_\lambda \cdot (\int \bar{u}^k\ dx - V) \\ u^{k+1} = u^k - \sigma \cdot (-\mathrm{div}\xi^{k+1} + \lambda^{k+1}) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \end{cases} \tag{7.23}$$

However, care has to be taken with the gradient ascent step of the dual variable ξ in the implementation of the scheme. The term $\frac{\xi}{\sqrt{1-<\xi,\xi>}}$ is not defined for $\|\xi\|_2 = 1$. As a remedy, one can subtract a small value ϵ in the denominator $\frac{\xi}{\sqrt{1-<\xi,\xi>+\epsilon}}$, although this numerical correction prevents us from finding exact solutions to (7.4) and thus renders the primal-dual approach suboptimal for this functional. Interestingly, the primal-dual formulation for the fully volumetric approach in Chapter 6 did not have this problem, instead, the gradient descent scheme to the primal problem had.

The primal-dual scheme has the advantage that no explicit projection onto the feasible set of volume compliant surfaces has to be performed.

Successive Over-Relaxation by Means of Lagged-Diffusivity. The lagged-diffusivity approach was already discussed in Chapter 5. By keeping the diffusivity $1/\sqrt{1+|\nabla u|^2}$ in Equation (7.15) constant the problem turns into a sparse linear system $\mathrm{div}(\nabla u) = 0$, which can be solved efficiently with numerical solvers such as Jacobi, Gauss-Seidel or Successive Over-Relaxation (SOR). The diffusivity is updated every few iterations.

This numerical algorithm will not provably converge to the global optimum when applied to our constrained problem (7.4) because of the additional projection step that

is needed (a projection was not necessary in the unconstrained case of problem (5.6)). Intuitively, with projection step the convergence is only guaranteed if in each step the solution is modified in the direction of the negative gradient. Nevertheless, we will explore this scheme further in the following as it leads to low runtimes and good results.

7.2.2 Implementation Details

We solve optimization problem (7.4) with each of the methods mentioned above and compare their performance. Prior to this, we discuss several implementation details of the algorithms in this section.

Projection Scheme. Except for the primal-dual scheme, each presented algorithm involves an orthogonal projection onto the convex set $\mathcal{C}$ of volume compliant height fields. This projection was devised in Chapter 6, Section 6.2.3 for implicit surfaces. An equivalent scheme applies to the case where the surface u represents a height field.

Boundary Conditions. We apply Dirichlet boundary conditions at the silhouette boundary $\partial\Sigma$ and Neumann boundary conditions if the silhouette coincides with the image boundary $\partial\Omega$:

$$u(x) = 0 \quad \forall x \in \partial\Sigma\,; \quad \frac{du}{dx}(x) = 0 \quad \text{if } x \in \partial\Omega \ . \tag{7.24}$$

The first condition ensures silhouette consistency of the solution, together with the fact that we only optimize over the silhouette Σ. The second condition makes the object surface touch the borders of the reconstruction volume orthogonally (see experiments below).

Discretization and Parallel Implementation. For numerical stability, we employed a non-negative discretization for the projected gradient descent and the FISTA algorithm. This works as follows. The expression $v(x) := \frac{\nabla u}{\sqrt{1+|\nabla u|^2}}$ inside the divergence is computed for discrete positions $v_{i-\frac{1}{2},j}$, $v_{i+\frac{1}{2},j}$, $v_{i,j+\frac{1}{2}}$ and $v_{i,j-\frac{1}{2}}$ on the basis of the following stable derivatives of u:

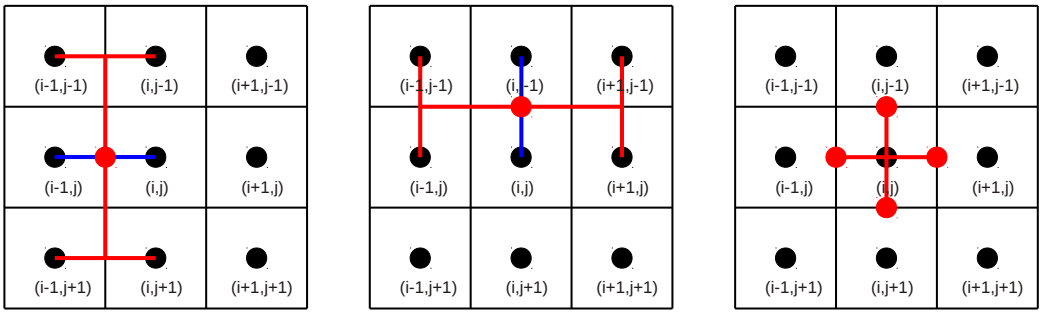


Figure 7.2: Illustration of the discretization scheme used for the FISTA implementation. *Left*: finite differences scheme used to compute the derivatives of u left (and right) of the target pixel in y-direction (red) and x-direction (blue) (cf. Equations (7.25) and (7.26)). *Center*: the analogous scheme used for the derivatives in u above (and below) the target pixel in x-direction (red) and y-direction (blue) (cf. Equations (7.27) and (7.28)). *Right*: the divergence for the central pixel is finally computed by central differences.

$$\left(\frac{\partial u}{\partial x_1}\right)_{i-\frac{1}{2},j} := u_{i,j} - u_{i-1,j} \tag{7.25}$$

$$\left(\frac{\partial u}{\partial x_2}\right)_{i-\frac{1}{2},j} := \frac{1}{2} \cdot \left(\frac{(u_{i-1,j+1} + u_{i,j+1})}{2} - \frac{(u_{i-1,j-1} + u_{i,j-1})}{2}\right) \tag{7.26}$$

$$\left(\frac{\partial u}{\partial x_1}\right)_{i,j-\frac{1}{2}} := \frac{1}{2} \cdot \left(\frac{(u_{i+1,j-1} + u_{i+1,j+1})}{2} - \frac{(u_{i-1,j-1} + u_{i-1,j})}{2}\right) \tag{7.27}$$

$$\left(\frac{\partial u}{\partial x_2}\right)_{i,j-\frac{1}{2}} := u_{i,j} - u_{i,j-1} \tag{7.28}$$

where the subscripts $(i - \frac{1}{2}, j)$, $(i, j - \frac{1}{2})$ denote the derivatives defined "on the grid". Definitions (7.28) and (7.25) can be understood as the central difference between two mean values (see Figure 7.2, *left* and *center*). The other two discretizations denote simple backward differences. The derivatives $\left(\frac{\partial u}{\partial x}\right)_{i+\frac{1}{2},j}$ and $\left(\frac{\partial u}{\partial x}\right)_{i,j+\frac{1}{2}}$ are defined equivalently for directions x_1 and x_2. Once the $v(x)$ are computed from these derivatives, the divergence is calculated simply by central differences (see Figure 7.2, *right*).

For the implementation of the primal-dual scheme, the discretization was chosen equivalently to the volumetric case - see Chapter 6 for details. For parallelization of the SOR method a *Red-black* scheme was implemented. All minimization methods described in Section 7.2.1 were parallelized on recent graphics hardware using the CUDA framework

including the projection step.

7.3 Experimental Results

In the following the proposed minimal surface approach (7.4) is applied to several real world images. First we discuss the integration of the formalism within an interactive framework for single view 3D reconstruction. Then, in the first part of the experiments the results are discussed and qualitatively compared to the implicit Cheeger set approach of the last chapter and to selected related single view 3D reconstruction approaches reviewed in previous chapters. The second part of the experimental section concentrates on the runtimes of the proposed minimal surface approaches and the performance of the discussed optimization algorithms.

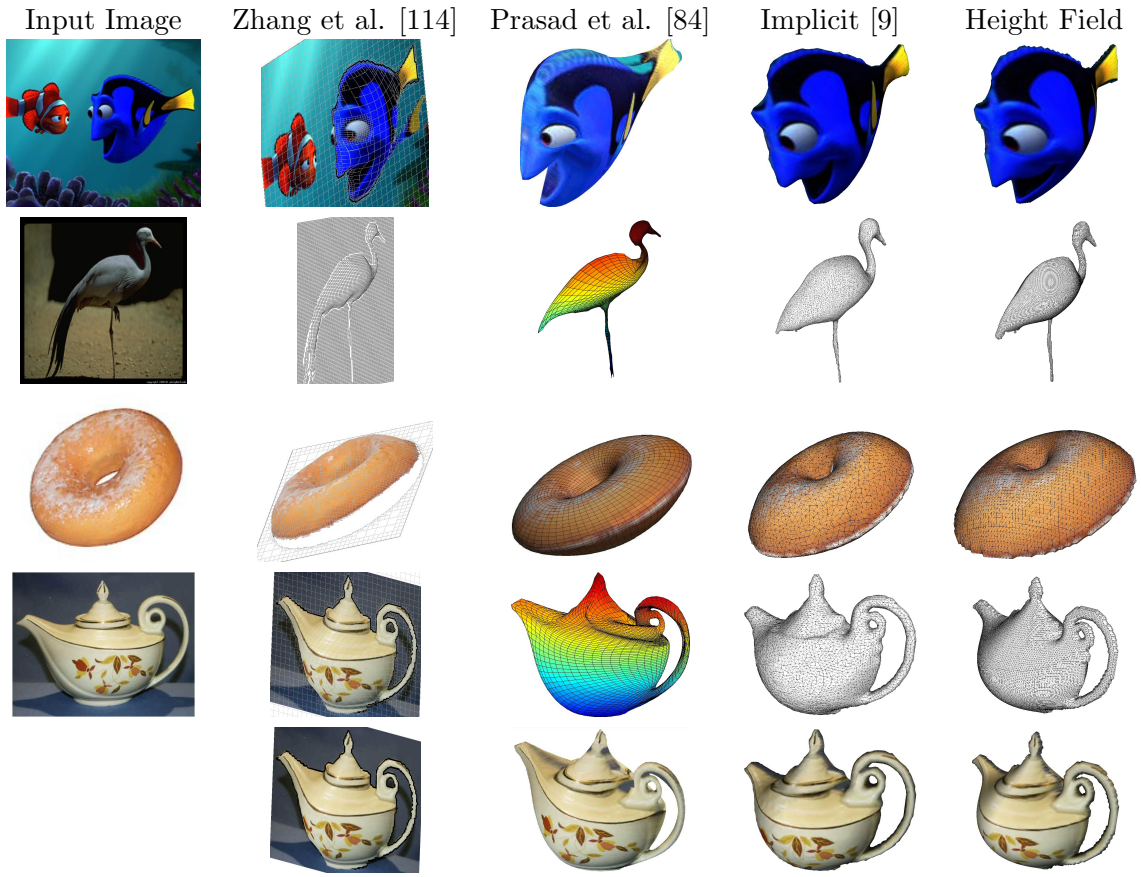


Figure 7.3: Qualitative comparison of the proposed height field approach to several single view methods.

7.3.1 Interactively Controlling Volume and Smoothness

The reconstruction volume is controlled by the user in the same way as described in Chapter 6. However, compared to the implicit formulation, the substantially reduced computation times of the height field approach make the volume adaptation much more convenient. This will be verified in the experimental section below.

We can control the amount of surface smoothness locally by introducing a weighting function similar to the weighted total variation in (6.1). Equation (7.3) then becomes

$$E(u) = \int_{\Sigma} g(x)\sqrt{1 + |\nabla u|^2}\, d^2x \ . \tag{7.29}$$

Fortunately, the introduction of the weighting function $g(x) : \Sigma \mapsto \mathbb{R}^+$ does not affect the convexity of the energy. This can be seen by a straightforward extension of the proof in Proposition 11. Modification of $g(x)$ is done by a scribble based system similar as in the implicit approaches.

7.3.2 Qualitative Comparison

Figures 7.3 and 7.4 show reconstruction results of the height field method for various input images. The examples represent objects with very different properties reaching from natural to man-made objects.

Solutions of the proposed algorithm are qualitatively similar to the ones of the implicit approach of Chapter 6. Differences can be found mainly in the resolution of the reconstructions. Since for the proposed height field method less memory and computation time is needed, input images with considerably higher resolution are feasible which result in more highly detailed silhouettes and reconstruction meshes as can be seen in most examples shown. Also, surfaces of the height field method appear smoother (see e.g. the balloon), since continuous depth values are computed. In contrast, depth values in the implicit approach have to be discretized into a finite number of voxels.

No weighting function was used for the experiments in Figures 7.4 and 7.3. This is different in Figure 7.6 where the user scribbles were used to define locations of reduced surface smoothness. Setting the surface weight $g(x)$ to less than 1 locally allows for sharp edges and surface extrusions such as the airplane wings.

7.3.3 Runtime Evaluation

We employed a gradient descent scheme, FISTA and SOR for solving problem (7.4) as described in Section 7.2.1. All experiments were done on a PC with a 2.27GHz Intel Xeon CPU, 12GB RAM equipped with a NVIDIA GeForce GTX480 graphics card running a recent Linux distribution.

For comparing runtimes of the respective optimization algorithms, each was run on a reconstruction example until convergence. For each time step t the distance $d(u_t, u^*)$ of

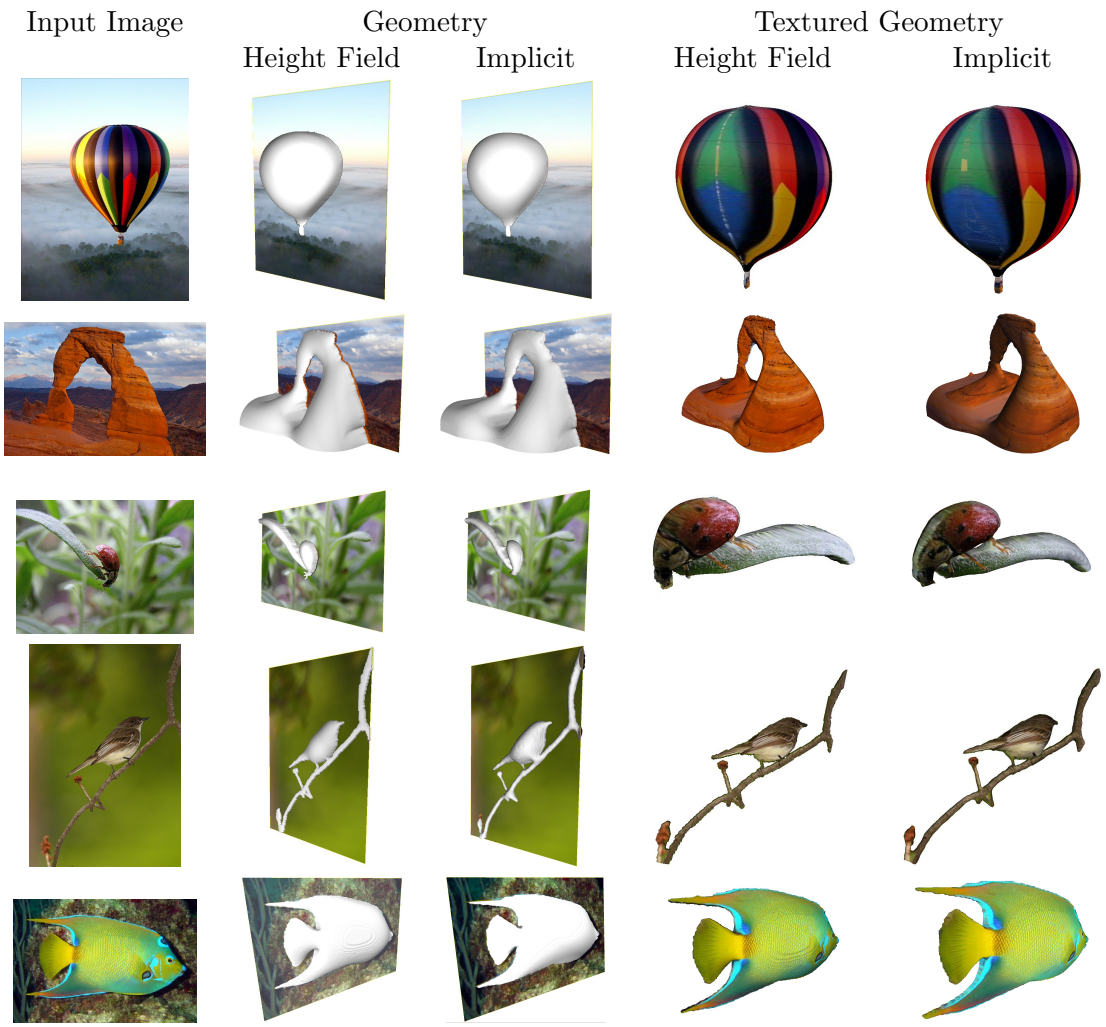


Figure 7.4: Reconstruction results of the proposed height field approach in this chapter are similar to the results of the implicit approach but are obtained with higher resolutions, less memory, lower computation times and increased depth precision.

the intermediate result u_t to the precomputed converged result u^* was plotted.

$$d(u_t, u^*) := \int_{(x,y)\in\Sigma} (u_t(x,y) - u^*(x,y))^2 dx\, dy \tag{7.30}$$

The convergence criterion for all experiments was set to

$$\left| \frac{E(u_{t-1}) - E(u_t)}{E(u_t)} \right| < \theta \tag{7.31}$$

with $\theta = 10^{-15}$. Figure 7.5 shows the results for three optimization schemes. The lagged diffusivity SOR approach of section 7.2.1 is clearly the most efficient algorithm. It also

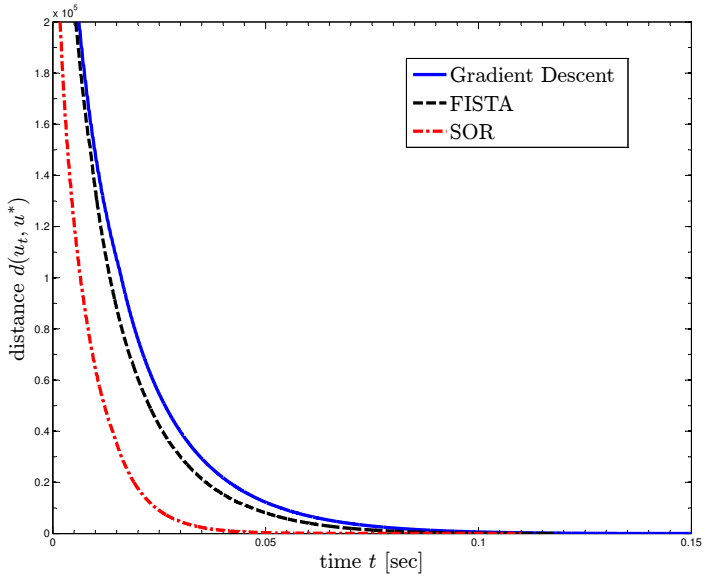


Figure 7.5: Runtime comparison of different algorithms minimizing Equation (7.4) measured on the teapot example without user-scribbles.

example		implicit	height field	speedup
teapot	size	131x101x58	131x101	
	time	1.82s	0.14s	13.0
arch	size	179x137x79	179x137	
	time	6.24s	0.99s	6.3
ladybug	size	151x122x27	151x122	
	time	1.62s	0.15s	10.8
bird	size	157x244x4	157x244	
	time	2.12s	0.2s	10.6
balloon	size	82x97x44	82x97	
	time	2.65	0.15s	17.7

Table 7.1: Runtime comparison of the height field and the implicit minimal surface approach for the experiments in Figure 7.4

Figure 7.6: Reconstruction results for the input images based on relaxed local smoothness at the locations of the user scribbles.

has the steepest convergence curve, i.e. it quickly attains a surface that is near to the converged solution. The FISTA algorithm is only slightly faster than gradient descent. The performance gain stems from the adaptive over-relaxation step. Note that due to the constraints on the feasible set, we have no proof that SOR converges to the global optimum (see the point *Projection Scheme* above). However, the results of SOR are almost identical to the those from methods that provably attain the global optimum.

The computational efficiency of the globally optimal height field approach is compared to that of the implicit minimal surface methods from Chapters 5 and 6 in Table 7.1. For the height field version the fastest optimization scheme was employed. The table lists runtimes for the experiments in Figure 7.4. It clearly shows that the proposed height field method is significantly faster than the implicit version. This difference mainly stems from the fact that the height field approach is two-dimensional, whereas the implicit versions are three-dimensional.

7.4 Conclusion

In this chapter a globally optimal height field based method for single view reconstruction was presented. In contrast to its implicit version of Chapter 6 it directly computes continuous depth values on a height field making the algorithm several orders of magnitude faster and provably globally optimal while offering similar solutions with higher resolution. The speedup is ideally suited to support the interactivity of the approach enabling instant user feedback for volume adaptation. To obtain maximum efficiency multiple numerical algorithms for solving the optimization problem (7.4) were introduced and compared against each other. SOR outperformed the other algorithms, although we were not able to show global optimality of this method.

Obviously, all these advantages come with the main drawback that the complexity of the set of reconstructable surfaces is limited, which means, e.g., that no self-occlusion or protuberances can be modeled. In the next chapter we will, therefore, revisit the implicit minimal surface approach and explore how shape constraints can be formulated in order to extend the flexibility and versatility of variational single view 3D reconstruction based on minimal surfaces.

8 Single View 3D Reconstruction with Relative Volume Constraints

Up to this point three different approaches to single view 3D reconstruction were proposed and discussed all of which were based on computing minimal surfaces complying with a given silhouette. Although the two implicit methods presented in Chapters 5 and 6 are theoretically capable of representing arbitrary surfaces in 3D, we observed that due to the minimal surface assumption and the lack of an additional shape prior reconstructions are limited to height fields. As a consequence, in Chapter 7 we looked into an equivalent volume-constrained minimal surface approach formulated on the basis of a height field representation. In spite of the advantages that came with this formulation (global optimality, strongly reduced computation times and higher precision) more complex reconstructions that cannot be represented by height fields, e.g. self-occluding objects, protrusions or dents, could not be handled, yet. In addition, an important source of shape information, i.e. the shading information in the image, has not been exploited so far. These points will be addressed in the following chapter in order to extend the class of reconstructable objects while reducing the required user guidance.

Contributions. To account for the general ill-posedness in 3D reconstruction from a single image, in this chapter we will introduce the notion of *relative volume constraints*. This concept will lead to two additional constraints on the minimal surface approach of Chapter 6, one that directly imposes shape knowledge on the reconstruction surface and one that relates the volume of different object parts. The imposed shape can either be determined by user sketches or by semi-automatically extracting it from shading information in the image. In this context the following contributions will be made:

- We impose characteristic object shape by means of *relative depth profiles* and *partial volume ratios*.
- The reconstructions are not limited to height maps and allow for self-occlusions, protuberances, dents and holes.
- A semi-automatic method is proposed to infer depth profiles from the shading information in the input image.

These contributions are embedded into the variational framework and optimized within bounds from the global optimum via convex relaxation. The approach is easily parallelized and can be run on graphics hardware. It was published in [8].

The chapter will be structured as follows: firstly, an intuition is given on how the newly introduced constraints can be used for modeling reconstruction surfaces. We then formalize the relative volume constraints mathematically, discuss their integration into the variational framework in Chapter 6 and detail the efficient optimization of the resulting problem. In Section 8.4 the new method for semi-automatic profile estimation from shading data is presented. Finally, the approach will be evaluated qualitatively and results will be compared to related methods.

8.1 Absolute and Relative Volume Constraints

In Chapter 6 a constraint on the surface volume was introduced as a means for inflating the reconstruction. This volume constraint is absolute in the sense that the specified volume does not depend on any other quantity. In the following relative volume constraints will be defined which introduce dependencies between object part volumes. In particular, two types of shape priors are derived from these relative volume constraints:

- user defined or shading based **relative depth profiles**, which define the object shape along its cross sections and
- **volume ratio constraints**, which specify the volume ratio of object parts with respect to the full object.

Intuition Behind the Constraints. To give a clear idea of the reconstruction process based on relative volume constraints, an example of the reconstruction of the watering can is shown in Figure 8.1. Figure 8.1 a) shows the original image of the watering can. If we apply the reconstruction method from Chapter 6 and look for a minimal surface that is consistent with the object silhouette while imposing a global constraint on the object volume we obtain the ball shaped reconstruction with flat handle in Figure 8.1 b). To improve the result we introduce a depth profile constraint, which defines the rough shape of the object along a cross section. In the example above, the profile in Figure 8.1 c) is imposed along the vertical cross section of the can indicated in red in Figure 8.1 a). It can either be given by the user or estimated from shading information. By imposing this profile (see Section 8.1) we obtain the result with handle in Figure 8.1 d). The object shape now resembles a realistic watering can instead of a ball. Yet, the handle is reconstructed as a solid object. To further improve the reconstruction we apply a volume ratio constraint. 'Volume ratio' means that we restrict the object volume within the indicated pink region to a specific ratio of the full object volume, e.g. to 0 for the

a) Original | b) Volume Constraint | c) Depth Profile | d) Partial Result | e) Volume Ratio | f) Result

Figure 8.1: 3D reconstruction of the watering can using absolute and relative volume constraints, see text for explanation.

region below the handle indicated in pink in Figure 8.1 e) in order to obtain a hole. We finally obtain the improved reconstruction in Figure 8.1 f).

Note that the imposed profile constraints define relative instead of absolute depth values, i.e. the depth of one pixel is proportional to the depth of a reference pixel within the profile. Since the depth values are relative the profiles and thus the object shape automatically scale with increasing volume. An example is shown in Figure 8.2.

In the following both constraints will be devised formally and integrated into the variational framework for single view 3D reconstruction.

8.1.1 Relative Depth Profiles

Relative depth profiles indicate the shape of the object along a given cross section. Such a profile consists of two ingredients:

- the *profile line* which marks the location of the profile in the image plane (see the red line in Figure 8.1 a))
- the desired qualitative (not absolute) depth values (*depth profile*) along the line (see the pink sketch in Figure 8.1 c))

The depth profile can either be sketched by the user or computed from shading information (see Section 8.4).

Let $C \subseteq \Sigma$ denote the profile line across the object within the image plane, which indicates the desired location of the shape profile. Let

$$R_y := \{x \in \mathbb{R}^3 \,|\, \pi(x) = y\} \tag{8.1}$$

denote the ray of all voxels which project onto the image pixel $y \in C$. Let the depth ratio $c_y \in \mathbb{R}_0^+$ indicate the depth of the object at pixel y with respect to that of a reference

a) Input Profile b) 20% Volume c) 40% Volume

Figure 8.2: Application of the shape profile in a) to a spherical 2D shape with b) 20% and c) 40% volume. Since the depth constraints are relative, the shape scales naturally with increasing volume.

profile pixel, which can be picked arbitrarily from those within the profile C. We set $c_{\text{ref}} = 1$ for the ray R_{ref} at the reference pixel. The constraint set of all reconstruction surfaces that obey the relative depth profile is defined as

$$U_C := \left\{ u \in BV(\mathbb{R}^3; \{0,1\}) \;\middle|\; \int_{R_y} u(x) d^3x = c_y \int_{R_{\text{ref}}} u(x) d^3x \;,\;\; \forall y \in C \right\}. \tag{8.2}$$

Equivalent constraint sets can be defined for additional relative depth profiles specified with respect to other profile lines C_i. Their concatenation defines the set of surfaces that comply with all these relative depth profiles simultaneously. However, one has to take care not to define contradicting relative depth constraints as their intersection will result in the empty set.

8.1.2 Relative Volume Ratios

The second type of constraint are relative volume ratios. A volume ratio constraint defines a fixed volume ratio for part of the surface with respect to the whole object, e.g. we can define that the wings of the plane in Figure 8.8 should contain 25% of the volume of the whole plane. Such relative volume constraints allow for protuberances, dents, self-occlusions and holes in the reconstruction.

Let $T \subset \mathbb{R}^3$ denote the set of voxels on which the volume ratio constraint is imposed. This set can be indicated interactively by the user as discussed in the following section. The constraint set can then be defined as follows

$$U_T := \left\{ u \in BV(\mathbb{R}^3; \{0,1\}) \;\middle|\; \int_T u(x) d^3x = r_p \int u(x) d^3x \right\}. \tag{8.3}$$

Again, each volume ratio constraint will yield one such set. The intersection of these sets will be the set of surfaces that comply with all volume ratio constraints.

8.2 Efficient Optimization via Convex Relaxation

In this section the final energy minimization problem with respect to the reconstruction surface u is derived. A convex relaxation approach similar to the one in Chapter 6 will be taken.

We follow the definition (6.4) of the relaxed set U^r of surfaces that are compliant with silhouette and global volume V and relax constraint sets (8.2) to U_C^r and (8.3) to U_T^r accordingly by reverting to $u : \mathbb{R}^3 \to [0, 1]$. Taking the intersection of all the constraint sets we arrive at the following optimization problem

$$\min_{u \in U^r \cap U_T^r \cap U_C^r} \int g(x) |\nabla u| d^3x. \tag{8.4}$$

The constraints for the global volume, the depth profiles and the volume ratios are all linear and thus their constraint sets as well as their intersection are convex. For efficient optimization we consider the primal-dual formulation of (8.4) by introducing Lagrange multipliers ν, μ^i, and γ_y^i for each constraint. This results in the following saddle point problem:

$$\begin{aligned} \max_{\substack{|\xi(x)| \leq g(x) \\ \nu, \gamma_y^i, \mu^i \in \mathbb{R}}} \min_{u \in U_\Sigma} & \int -u \text{ div } \xi \; d^3x + \nu \left(\int u \; d^3x - V \right) + \\ & \sum_i \sum_{y \in C_i} \gamma_y^i \left(\int_{R_y} u \; d^3x - c_y^i \int_{R_{\text{ref}_i}} u \; d^3x \right) + \\ & \sum_i \mu^i \left(\int_{T_i} u \; d^3x - r_p^i \int u \; d^3x \right). \end{aligned} \tag{8.5}$$

Again, optimization is done with the primal-dual method by Chambolle and Pock [21]. It consists of alternating a gradient descent with respect to the function u and a gradient ascent for the dual variables ξ, ν, γ_y and μ interlaced with an over-relaxation step on the primal variable. For the sake of readability we will regard the case for only one given depth profile and one given volume ratio constraint with respective Lagrange multipliers γ_y and μ.

$$\begin{cases} \xi^{k+1} = \Pi_{|\xi(x)|_2 \leq g(x)}(\xi^k + \tau_\xi \cdot \nabla \bar{u}^k) \\ \nu^{k+1} = \nu^k + \tau_\nu \cdot (\int \bar{u}^k \; dx - V) \\ \mu^{k+1} = \mu^k + \tau_\mu \cdot (\int_T \bar{u}^k \; dx - r_p \int \bar{u}^k \; dx) \\ \gamma_y^{k+1} = \gamma_y^k + \tau_{\gamma_y} \cdot (\int_{R_y} \bar{u}^k \; dx - c_y \int_{R_{\text{ref}}} \bar{u}^k \; dx) \\ u^{k+1} = \Pi_{u \in [0,1]} \left(u^k - \sigma(-\text{div}\xi^{k+1} + \nu^{k+1} + \mu^{k+1}(1 - r_p)\mathbf{1}_T + \sum_{y \in C} \gamma_y^{k+1}(1 - c_y)\mathbf{1}_C \right) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \end{cases} \tag{8.6}$$

where Π_A denotes the projection onto the set A and $\mathbf{1}_A$ is its indicator function. 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 discretization is done as described in Chapter 6.

8.3 Interactive Modeling with Relative Volume Constraints

We have already provided an intuition on how *relative depth* and *volume ratio* constraints can be used for improving the quality of single view 3D reconstructions. This section deals with the details of how these constraints are implemented in an interactive framework.

8.3.1 Relative Depth Profiles

Obtaining the Profiles. Relative depth profiles help to define the shape of the reconstruction. The question remains how the profiles themselves can be obtained. For simple object shapes, rough profile sketches can easily be outlined by the user, e.g. the profile of the watering can in Figure 8.1 c). This is done by specifying control points of splines or piecewise linear curves. A method to semi-automatically estimate the profiles from shading data in the image is given in the next section.

Application of the Profiles. Although in principle profiles can be applied along arbitrary curves $C \subseteq \Sigma$, for simplicity we let the user either define profiles along *vertical* or *horizontal* lines in the input image whose position is specified by a simple mouse click.

Since we wish to impose the profile shape upon the whole object - not only the marked cross-section - a copy of the relative depth profile specified by the user is applied to each object cross-section parallel to the reference line. In the watering can example in Figure 8.1 profile c) was specified along the vertical cross-section in a) and then the same profile was applied to every parallel column in the input image. Solving problem (8.4) with the corresponding constraints will result in a smooth reconstruction surface since the profiles only specify relative depth within each cross-section. Another example for a surface constructed with this strategy is given in Figure 8.2.

Profile Attenuation. One can also choose to soften the profile constraints with increasing distance to the reference cross-section. To this end, we suggest to put a limit on the Lagrange multipliers for each profile constraint depending on its distance to the reference profile. The Lagrange multiplier limits can be chosen according to a function with maximum value on the reference cross-section and decreasing values along neighboring cross-sections, e.g. a Gaussian. Let C denote the user indicated reference cross-section. Then for any neighboring profile constraint C_n we compute the minimum distance d_n

between C_n and C to obtain the limit $L_m(C_n)$ on the Lagrange multiplier for each cross-section C_n

$$L_m(C_n) = m\left(1 - \frac{d}{\max\{d_n\}}\right)^2. \tag{8.7}$$

Here m is the maximum Lagrange multiplier for any of the profiles. Setting m to infinity yields the same user indicated hard depth profile constraint for each of the cross-sections. The following proposition gives a theoretical justification for pruning the Lagrange multipliers:

Proposition 14. Confining the Lagrange multiplier of a volumetric constraint in the primal-dual formulation (8.5) amounts to an L_1 soft constraint.

Proof. We regard the minimal surface problem with a global volume constraint, whose Lagrange multiplier is confined by k:

$$\begin{aligned}
& \max_{\substack{|\xi(x)|\leq g(x) \\ |\nu|\leq k}} \min_u \int -u \operatorname{div} \xi \; d^3x + \nu \left(\int u \; d^3x - V \right) \\
= \quad & \max_{|\xi(x)|\leq g(x)} \min_u \int -u \operatorname{div} \xi \; d^3x + \nu \left(\int u \; d^3x - V \right) - \delta_{|.|\leq k}(\nu) \\
= \quad & \max_{|\xi(x)|\leq g(x)} \min_u \int -u \operatorname{div} \xi \; d^3x + k \left| \int u \; d^3x - V \right|
\end{aligned}$$

where for the second equality we used that the Fenchel-dual of $\max_\nu \nu \cdot x - \delta_{|.|\leq k}(\nu)$ is given by $k|x|$. □

Multiple Profiles. It can make sense to define profiles for multiple parallel (vertical or horizontal) cross-sections. In 8.3 e), e.g, relative depth profiles for multiple horizontal cross-sections are specified for the watering can. For the cross-sections that lie between two different profile definitions, the profile constraints are linearly blended. This will result in a smooth transition of the profiles. The result for the watering can after optimization can be seen in Figure 8.7.

8.3.2 Volume Ratio Constraints

Defining Volume Ratios Interactively. To indicate the part of the object, where the volume ratio constraint is imposed, the user draws a region into an arbitrary 3D view of the reconstruction as shown by the pink region in Figure 8.1 e). Then he specifies a volume ratio r_p relative to the overall object volume V. Each voxel in the reconstruction volume is then projected onto the viewing plane of the camera. All voxels in $\mathbb{R}^3$ which project into the user drawn region constitute the constraint set $T \subset \mathbb{R}^3$ on which the volume ratio constraint is imposed.

Modes of Optimization. Constraints on volume ratios can either be imposed as an additional constraint from the beginning or as a subsequent optimization problem after convergence of the original problem. For the latter case optimization is done only within the voxel set T. This can be used to limit the effects of the subvolume constraint to the selected voxels.

8.4 Estimating Relative Depth Profiles from Shading Information

Rather than drawing the depth profiles by hand, which can be tedious, in this section a method is proposed, which estimates them directly from the shading information in the input image. The following assumptions are made:

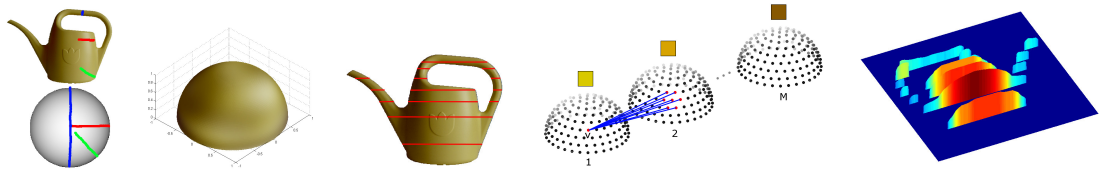
- At the locations where we estimate the profiles, the object is made of a homogeneous material with constant albedo.
- The distances of the light sources to the object are large compared to the object size.

The last assumption is given for most scenes. These two assumptions imply that points with similar normals result in similar irradiance. In general, our framework allows for arbitrary reflectance properties including shiny objects with specular surfaces.

The proposed interactive approach for estimating the profile consists of three steps, which are illustrated in Figure 8.3. In the first step the reflectance function of the target object is estimated from user given samples. In the second step the user defines the profiles to be estimated by marking their respective locations in the input image. Finally, relative depth along the profiles is computed automatically by finding the shortest path in a specific graph. In the following we will detail these steps.

Estimation of Reflectance Function. The process of estimating the reflectance function is illustrated in Figure 8.3 a) and b). For performing regression on the reflectance function $\rho : S^2 \to \mathbb{R}^3$, which maps each normal direction to its corresponding reflected color, we need samples. These are specified by pairs of curves $s_1, s_2 : [0, 1] \to \mathbb{R}^2$ given by the user. The first curve of each pair is drawn onto the input image, the second one onto the image of a sphere, whose points represent normal directions. For each pair, the sequence of colors from the input image described by s_1 is mapped to the normal directions given by s_2. This step is illustrated in Figure 8.3 a).

Given the color samples, we perform regression on the reflectance function. To this end, we represent it as a sum of spherical harmonics basis functions and obtain their coefficients through a least squares estimate (see Figure 8.3 b). Each color channel is



a) b) c) d) e)

Figure 8.3: The different steps for extracting profiles from an input image using shading information: a) The user provides color samples of the reflectance function by marking corresponding scribbles in the input image and on a sphere. b) The color samples are used to estimate the complete reflectance function of the input object by means of regression. c) The user marks horizontal lines in the input image for which the height profiles will be estimated. d) For estimating a single profile a shortest path is computed on the graph indicated. Each node in the graph combines a profile pixel with a possible normal direction at this point. e) Each shortest path then corresponds to a sequence of normals and thus a depth profile which determines the shape of the watering can .

estimated separately. After drawing a new curve pair, regression can be recomputed on the fly. For our experiments we used spherical harmonics up to degree 5.

Marking Profile Locations. In the second step the user marks the profile lines in the input image for which relative depth profiles will be estimated (Figure 8.3 c). The lines are arbitrary as long as they start and end at contour points and the corresponding profiles do not contradict. For each of the profile lines we estimate the corresponding depth profile by computing a shortest path in a graph, which is described in the following.

Depth Computation on a Graph. We start by defining the set $D = \{n_1, n_2, ..., n_N\} \in \mathbb{R}^3$ of uniformly sampled normal directions and the color sequence $c_1, c_2, ..., c_M \in \mathbb{R}^3$ along the profile line C. The graph consists of a set of M connected domes (half spheres), one dome for each pixel in the profile line C (see Figure 8.3 d). Each dome consists of N nodes, each representing one possible normal direction of the set D. Thus, the node v_{ij} in the graph represents the j-th sampled normal direction in dome i for profile pixel i. Each node of dome i is connected to the neighborhood of the same node in dome $i+1$ containing all nodes of similar normal directions (see the neighborhood connections of node v in Figure 8.3 d).

Each path in the graph consists of M nodes (one in each of the domes, i.e. one normal direction for each pixel in the profile) representing one possible sequence of surface

normals from the start to the end point of the profile line. The start and end normals are known, since the start and end points of the profile line lie on the object contour. Hence, their normals coincide with those of the silhouette at these points.

We assume that the most likely path connecting the start and end normal is the one with minimal color difference between reflectance value and image color for each node and minimal surface curvature in the sequence. The weight for each edge in the graph is, therefore, defined as

$$w(v_{ij}, v_{i+1k}) = \lambda \cdot \|c_{i+1} - \rho(n_k)\| + \cos^{-1} < n_j, n_k > . \tag{8.8}$$

The first term ensures that the color reflected in normal direction n_k is similar to the observed pixel color c_{i+1}. The second term penalizes large deviations of neighboring normals along the profile. We compute the shortest path in this graph with Dijkstra's algorithm to obtain the most likely sequence of normals $(j_1, .., j_M) \in D^M$ by minimizing the energy

$$E(v_{1j_1}, .., v_{Mj_M}) = \sum_{i=1}^{M-1} w(v_{ij_i}, v_{i+1j_{i+1}}). \tag{8.9}$$

If the desired profiles are symmetric we can increase the stability and accuracy of the algorithm by adding the constraint that each normal in the first half of the sequence must be the mirrored version of its corresponding normal in the second half. Symmetric profiles do not only occur for symmetric objects, as there are many non-symmetric objects that exhibit symmetric parts.

After the sequence of normals have been computed for the profile, we integrate it to obtain the relative depth values.

For $\lambda = 0$ in (8.8), minimizing energy (8.9) will result in the geodesic on the unit sphere from the start to the end normal. This regularizer is motivated by the reasonable assumption that the curvature of objects, i.e. the variation of their surface normals, is minimized.

Differences to Shape-from-shading Approaches. The proposed approach has some points in common with shape-from-shading algorithms, since the ultimate goal is to derive shape from reflectance. It was already mentioned in Section 3.1 that in general the shape-from-shading problem [49] has ambiguous solutions. To overcome the inherent ambiguities *interactive* methods have been proposed such as the one by Zeng et al.

[113]. Although our method is interactive as well, the approach differs from existing shape-from-shading methods mainly in the following point: by estimating shape profiles from shading information we directly infer qualitative shape knowledge instead of taking the detour over numerically accurate dense normal maps to obtain shape information. In this way several drawbacks of typical shape-from-shading methods are avoided:

- The computation of shape profiles is simpler than the computation of dense normal maps and thus less error prone.
- The proposed method of estimating reflectance functions is well suited for deriving *qualitative* shape characteristics instead of numerically accurate ones. The approach is not designed to compute accurate height-fields as even for state-of-the-art shape-from-shading methods real world images pose a serious challenge (see e.g. [35, 115]).
- In contrast to common shape-from-shading algorithms which mostly assume Lambertian reflectance properties, the proposed method assumes the object to be made of a homogeneous but otherwise arbitrary material with constant albedo only at those locations where depth profiles are to be estimated. Thus, our reconstruction approach can deal with textured objects, color and shadows. Since the user only indicates profile lines in regions of constant albedo reasonable profile estimates can be computed and then propagated to textured and shadowed regions by means of the smoothness assumption on the surface. For this task, flexible scalable shape profiles are much better suited than point-wise absolute normal information.
- User input in our approach is not used to improve the surface normal inference, but merely to estimate the reflectance function of the object.

8.5 Experimental Results

In this section 3D reconstruction results with imposed relative volume constraints, i.e. profile constraints and volume ratios, are presented. Examples come with the following constraints (in addition to global volume)

- manually sketched relative depth profile constraints (attenuated and constant)
- multiple, semi-automatically estimated relative depth profile constraints
- volume ratio constraints
- relative depth profile constraints *and* volume constraints in combination

Results are also compared to related single view 3D reconstruction approaches as well as to the approaches presented in Chapters 6 and 7.

8.5.1 3D Reconstruction Results with Relative Volume Constraints

Without shape constraints such as relative depth profiles or volume ratios the reconstruction approach from Chapter 6 fails in many situations. This is illustrated in Figure 8.4.

Figure 8.4: *Top:* Input images, *Bottom:* 3D reconstruction results obtained by computing minimal surfaces with constant volume (see Chapter 6). With no further shape constraints the reconstructions fail for various reasons.

These poor results are due to various reasons: the shoe, e.g., lacks the characteristic curved *shape in viewing direction*. Similar problems occur with the pyramid profile, the vase and the handle of the watering can. For some objects, the *minimal surface assumption* turns out to be an unsuitable prior. E.g, the wheels of the car collapse to a thin layer. The same happens with the watering can handle. Finally, *self-occlusion* as in the handle of the watering can and the plane wings cannot be reconstructed.

Figures 8.5 and 8.8 show reasonable improvements of the reconstructions from Figure 8.4 after adding relative volume constraints. Details are given in the following.

Relative Depth Profiles

User Drawn Profiles. Relative depth profiles determine the basic shape of an object along an arbitrary cross-section. Figure 8.5 shows several reconstruction results based on user drawn depth profiles. Since the profiles scale with the global volume, it suffices to indicate the profile line on the image plane (here in red) together with a rough sketch of the corresponding depth profile (here in pink). The same relative depth profile constraints are then applied to each parallel cross-section in the way explained in Section 8.3.1.

The profile of the shoe, for example, indicates that the shoe is wider at the front and back and narrow in the center. The profile imposed on the vase makes it slimmer and more bulgy at the top. The profile constraints of the vase are, in addition, attenuated by Lagrange multiplier pruning towards the left and right (see Section 8.3.1 and Equation (8.7)).

a) b) c) d)

Figure 8.5: 3D reconstruction results with relative depth profile constraints. a) The 2D input images with respective profile locations (in red). b-d) Reconstructions with corresponding depth profiles drawn in pink.

For the pyramid, the value of the weight function g was reduced along the base line (see energy (8.4)) in a first step to obtain the sharp edge at the bottom. The result is similar to the pyramid in Figure 8.4. To model the pyramid's triangular shape we additionally imposed the shape profile indicating a linear depth increase from the top to the bottom (see pink line in Figure 8.5), which results in the reconstruction shown.

For the watering can we first imposed a user drawn vertical profile as shown in Figure 8.1. The depth profile constraints were attenuated with increasing distance from the reference profile.

Shading Based Profiles. To obtain shading based depth profiles the reflectance function of the object has to be estimated in a first step. Examples of estimated reflectance functions for various input images are shown in Figure 8.6. The corresponding user provided samples are presented in the center row. For the orange, the porcelain mug and the watering can the reflectance functions could be estimated easily from a few simple user strokes. The estimated reflectance function for the watering can is shown in Figure 8.3. Although high frequent reflectance functions such as that of the metal mug can be reconstructed reasonably, many samples and a high number of spherical harmonics

Figure 8.6: Estimated reflectance functions for several examples. *Top:* smoothed input images, *Center:* shading samples extracted by the user. *Bottom:* estimated reflectance function (even for tiny and textured objects), see Section 8.3.1 for details.

basis functions are necessary. The teapot is an example for a partially textured object. Samples are only taken from untextured regions. The shown reflectance function thus represents the homogeneous porcelain material the teapot is made of. One can see that the reflectance function has a yellowish glow at the top and a darker shading at the bottom. These colors stem from the lighting in the original image and help to distinguish normal directions.

Based on these reflectance functions shading based reconstructions can be computed. Figure 8.7 shows reconstruction examples based on depth profiles which were computed from the estimated reflectance functions. To this end, the semi-automatic procedure described in Section 8.3.1 was used. No further constraints have been manually applied. Note that we can estimate the depth profile equally well on shiny (mug) and diffuse (watering can) materials since we estimate the reflectance function of the target object prior to the object shape. The estimated depth profiles for the watering can are shown in Figure 8.3 e). The examples were computed with $0.01 < \lambda < 0.1$ (see Equation (8.8)).

Relative Volume Ratio Constraints

Volume ratio constraints can be imposed to obtain protuberances, dents, self-occlusions and holes.

Figure 8.7: 3D reconstruction results with automatically estimated relative depth profile constraints based on shading information. The center row shows the reconstruction geometry without texture.

Figure 8.8 shows the reconstruction of a tuba with a zero volume ratio constraint for modeling the opening and a 30% volume constraint for inflating the thin tubes and reducing the inflation of the lower part of the opening.

For the airplane example, without relative volume constraints the only chance to model the wings is by reducing the weight g. However, they turn out unnaturally rectangular that way (see Figure 8.4). By adding volume ratio constraints for the wings requiring the side wings to contain 25% and the tail wing 5% of the object volume we obtain the results with self-occluding wings on the right. The characteristic wing bending was achieved by positioning the camera to show the plane slightly from behind. Then the wing is marked in that view so that all voxels that project into this drawn 2D region belong to the wing and are inflated.

In the car example the reconstruction without additional volume constraints after reducing the weight g along the tire borders yields two very long tires instead of four normal ones. By adding a volume ratio constraint with fraction zero we can separate the tires.

a) b) c) d)

Figure 8.8: 3D reconstruction results with application of volume ratio constraints. a) Original images with relative depth profiles marked in red (location) and pink (depth function), b-d) Reconstructions with application of volume ratio constraints.

For the watering can we increased the thickness of the spout by adding a 4% volume ratio constraint.

8.5.2 Comparison to State-of-the-art Reconstruction Approaches

Finally the approach with relative volume constraints is compared to the state-of-the-art single view 3D reconstruction approaches reviewed in Chapter 3. Figure 8.9 shows reconstructions by Prasad et al. [86], Zhang et al. [114] and the fully volumetric approach

a) b) c) d)

Figure 8.9: Different 3D reconstruction results obtained with a) the method by Prasad et al. [86], b) the approach in Chapter 6, c) the method by Zhang et al. [114] and d) the proposed method with relative volume constraints.

in Chapter 6. All reconstructions are based on the same input images. Note that by comparing to the fully volumetric minimal surface approach, we indirectly also compare to the height field method of Chapter 7 as they yield qualitatively very similar results.

The results show that the proposed method compares well to previous approaches, e.g. some reconstructions are less ball shaped and thus look more realistic than for other methods. In addition, the approaches by Zhang et al. and Prasad et al. require substantially more user input.

8.6 Conclusion

In this chapter we introduced relative volume constraints into 3D reconstruction from a single image and added them to the implicit minimal surface approach presented in Chapter 6. Two types of such constraints, relative depth profiles and volume ratios, allow to impose shape knowledge on the object. Due to their relative nature it suffices to let the user sketch the profiles by hand. In addition, we showed how to exploit shading information in the image for semi-automatic profile estimation in order to reduce the required user guidance in the reconstruction process.

The new shape constraints were designed to address shortcomings of the volumetric minimal surface approaches in earlier chapters. With the novel approach shape knowledge can now be introduced along object cross sections, balloonish reconstructions are prevented and self-occlusions as well as dents and protuberances can occur in the reconstructions.

The presented shape constraints were introduced into the optimization framework in Chapter 6 by adding Lagrange Multipliers and optimizing with the primal-dual scheme reviewed in Chapter 2. Although an increasing runtime could be observed stemming from the additional constraints, reconstructions can still be computed within a matter of seconds.

9 Conclusion

The goal of this thesis was to show that variational convex optimization methods can be used to formulate elegant and simple approaches to interactive single view 3D reconstruction. To this end, four different algorithms were proposed that are based on the idea of searching for surfaces with weighted minimal area under constraints whose projections onto the image plane are congruent with the object silhouette. To avoid trivial solutions, additional constraints were conceived and evaluated.

The proposed single view approaches are interactive by design, but a premise was to keep user input intuitive and as simple as possible. The basic common workflow starts by extracting the object silhouette with the help of scribble-based segmentation. A silhouette compliant minimal surface is then computed by minimizing a continuous energy functional under constraints with efficient optimization algorithms. Computed reconstructions can be edited by adapting the reconstruction parameters and recomputing the surface. In addition, the framework allows the user to adapt the weight of the surface area term in the energy for inducing sharp edges or protrusions in the reconstruction.

For optimization provably convergent algorithms were employed that were parallelized on graphics hardware. Convex formulations were devised for all proposed energy minimization problems that can be optimized either globally or within bounds from the global optimum. One major advantage of these convex relaxation approaches is that their results do not depend on the initialization.

Except for the approach in Chapter 7, all proposed minimal surface formulations represent the reconstruction surface implicitly in the form of a binary valued function defined on $\mathbb{R}^3$, indicating inside and outside of the surface. This representation enables reconstructions of arbitrary topology. Moreover, it can easily be transformed into relaxed formulations that lead to convex energies.

For the first silhouette compliant minimal surface approach to single view 3D reconstruction presented in Chapter 5 a parametric data affiliation term was constructed that served as a shape prior and at the same time inflated the reconstruction surface. The resulting formulation is equivalent to a binary 3D segmentation problem and can be solved globally optimally. The extremality condition of the relaxed segmentation energy is a partial differential equation that is solved by means of lagged diffusivity and a successive over-relaxation scheme. Although experiments showed that the algorithm could be used to reconstruct objects from quite different areas in man-made and natural environments, it also became evident that the data affiliation term introduced a strong

shape bias, which inhibited modeling flexibility and exhibited too many parameters.

Therefore, in Chapter 6 the shape prior term was replaced by a hard constraint on the surface volume. This approach comes with several advantages: it is based on a non-heuristic inflation technique, required user input is reduced, shape naturally evolves by computing a weighted minimal surface with fixed volume and weight adaptation can be used more effectively for influencing the surface shape allowing for protrusions and box shaped objects. Although the corresponding combinatorial energy minimization problem cannot be globally optimized by convex relaxation, we presented a thresholding scheme leading to solutions that adhere to an energy bound with respect to the global optimum and comply with the user specified target volume. For optimization two algorithms were proposed, one based on the primal-dual scheme, the other based on the ADMM algorithm. Although both algorithms are similar, they differ in runtime and memory consumption. We showed that reconstructions can be computed within a matter of seconds depending cubicly on the resolution of the reconstruction volume. Experiments showed that the proposed approach leads to reconstructions that are qualitatively comparable to the approach with shape prior.

The fact that the volume based algorithm results in reconstructions that are representable as a height-field (mirrored along the image plane) inspired two different directions. On the one hand an equivalent height-field based energy minimization formulation was proposed in Chapter 7 to reduce the complexity of the original approach, and on the other hand further constraints to the fully volumetric approach were introduced in Chapter 8 to enable reconstructions of more complex objects, which are not representable by height fields.

For the first direction an energy minimization approach was introduced to compute weighted minimal height-fields with fixed volume. For optimization different algorithms were compared, among them FISTA, primal-dual and successive over-relaxation. In particular, the reconstruction surfaces are globally optimal solutions that can be computed without convex relaxation techniques. Although the algorithms exhibited different runtimes for the problem, they all lead to speed-ups of several orders of magnitude compared to the fully volumetric approach. As a direct consequence we were able to reconstruct objects using a significantly higher resolution with quadratic complexity. Furthermore, depth values could be computed on a continuous domain without the need for an implicit voxel-based representation, which raises the quality of reconstruction surfaces.

Although the height-field based reconstruction formulation is globally optimal, elegant and fast, its extendability to more complex reconstructions is limited. Therefore, in the last chapter we returned to the fully volumetric, fixed volume, weighted minimal surface formalism and extended it by relative volume constraints. Two additional shape constraints were introduced: relative depth profiles and volume ratio constraints. The former impose manually drawn or semi-automatically inferred shape profiles on the reconstruction that scale with the global reconstruction volume. The latter relate sizes of object parts and can be drawn by the user. The additional constraints are introduced by

means of Lagrange multipliers into the primal-dual optimization scheme. Experiments show that the constraints strongly increase the versatility of the minimal surface reconstruction approach and are still easy to handle for the user. Furthermore, a graph-based shortest path algorithm was proposed in order to estimate the relative depth profiles automatically from shading data in the image and thus compute reconstructions automatically based on estimated reflectance functions. A big advantage of the approach compared to shape-from-shading algorithms is that shape information from shading can be propagated to textured or shadowed regions of the object and that qualitative shape information are sufficient in contrast to exact normal maps.

In a nutshell, this thesis intended to show mainly two things. Firstly, that the simple idea of weighted minimal surfaces together with convex relaxation methods leads to powerful and versatile interactive reconstruction frameworks even without high-level learning approaches. And secondly, that these frameworks are sustainable in the sense that they can be extended by more dynamic constraints that can incorporate further information and cues from the input image.

Own Publications

[1] T. Möllenhoff, C. Nieuwenhuis, E. Toeppe, and D. Cremers. Efficient convex optimization for minimal partition problems with volume constraints. In *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition*, 2013.

[2] C. Nieuwenhuis, E. Toeppe, and D. Cremers. Space-varying color distributions for interactive multiregion segmentation: Discrete versus continuous approaches. In *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition*, pages 177–190, 2011.

[3] C. Nieuwenhuis, E. Toeppe, and D. Cremers. A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *Int. J. of Computer Vision*, 2013.

[4] M. R. Oswald, E. Toeppe, and D. Cremers. Fast and globally optimal single view reconstruction of curved objects. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2012.

[5] M. R. Oswald, E. Toeppe, K. Kolev, and D. Cremers. Non-parametric single view reconstruction of curved objects using convex optimization. In *Pattern Recognition (Proc. DAGM)*, 2009.

[6] M. R. Oswald, E. Toeppe, C. Nieuwenhuis, and D. Cremers. A survey on geometry recovery from a single image with focus on curved object reconstruction. In *Proceedings of the 2011 Conference on Innovations for Shape Analysis: Models and Algorithms*. Springer-Verlag, 2011.

[7] F. R. Schmidt, E. Toeppe, and D. Cremers. Efficient planar graph cuts with applications in computer vision. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 351–356, Miami, Florida, June 2009.

[8] E. Toeppe, C. Nieuwenhuis, and D. Cremers. Volume constraints for single view reconstruction. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2013.

[9] E. Toeppe, M. R. Oswald, D. Cremers, and C. Rother. Image-based 3d modeling via Cheeger sets. In *Asian Conf. on Computer Vision*, 2010.

[10] E. Toeppe, M. R. Oswald, D. Cremers, and C. Rother. Silhouette-based variational methods for single view reconstruction. In *Proceedings of the 2010 International Conference on Video Processing and Computational Video*, pages 104–123. Springer-Verlag, 2011.

Bibliography

[11] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[12] S. Bae and F. Durand. Defocus magnification. In *Eurographics*, 2007.

[13] O. Barinova, V. Konushin, A. Yakubenko, K. Lee, H. Lim, and A. Konushin. Fast automatic single-view 3-d reconstruction of urban scenes. In *Europ. Conf. on Computer Vision*, pages 100–113, 2008.

[14] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2:183–202, 2009.

[15] M. Botsch and L. Kobbelt. An intuitive framework for real-time freeform modeling. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, volume 23, pages 630–634, 2004.

[16] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[17] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary region segmentation of objects in n-d images. In *IEEE Int. Conf. on Computer Vision*, pages 105–112, 2001.

[18] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Journal of Statistical Planning and Inference*, 37:28–47, 1986.

[19] R. J. Bruck. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 61:159–164, 1977.

[20] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, University of Bonn, 2008.

[21] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Im. Vis.*, 40(1):120–145, 2011.

[22] T. Chan, S. Esedoḡlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Mathematical Analysis*, 66(5):1632–1648, 2006.

[23] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62:1019–1043, 2002.

[24] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge–preserving regularization in computed imaging. *IEEE Trans. on Image Processing*, 6(2):298–311, 1997.

[25] Y. Chen and R. Cipolla. Single and sparse view 3d reconstruction by learning shape priors. *Computer Vision and Image Understanding*, 115:586–602, 2011.

[26] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[27] C. Colombo, A. Del Bimbo, A. Del, and F. Pernici. Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:99–114, 2005.

[28] P. Concus. Numerical solution of the minimal surface equation. *Mathematics of Computation*, 21:340–350, 1967.

[29] F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos. Towards shape from shading under realistic photographic conditions. In *Proc. International Conference on Pattern Recognition*, pages 277–280, 2004.

[30] D. Cremers, N. Sochen, and C. Schnörr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In L. Griffith, editor, *Int. Conf. on Scale Space Theories in Computer Vision*, volume 2695 of *LNCS*, pages 388–400, Isle of Skye, 2003. Springer.

[31] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *Int. J. of Computer Vision*, 40(2):123–148, 2000.

[32] M. Daum and G. Dudek. On 3-d surface reconstruction using shape from shadows. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 461–468, 1998.

[33] E. Delage, H. Lee, and A. Ng. Automatic single-image 3d reconstructions of indoor Manhattan world scenes. In *Proceedings of the International Symposium of Robotics Research (ISRR)*, 2005.

[34] J. Duchon. *Constructive Theory of Functions of Several Variables, Oberwolfach 1976*, volume 571, chapter 'Splines minimizing rotation-invariant semi-norms in Sobolev spaces', pages 85–100. Springer, 1977.

[35] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008.

[36] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[37] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[38] H. Federer. *Geometric measure theory*. Springer-Verlag New York Inc., New York, 1969.

[39] A.R.J. Francois and G.G. Medioni. Interactive 3d model extraction from a single image. *Image and Vision Computing*, 19(6):317–328, 2001.

[40] B. Goldluecke and D. Cremers. Superresolution texture maps for multiview reconstruction. In *IEEE Int. Conf. on Computer Vision*, 2009.

[41] B. Goldluecke and D. Cremers. Convex relaxation for multilabel problems with product label spaces. In *Europ. Conf. on Computer Vision*, 2010.

[42] T. Goldstein, X. Bresson, and S. Osher. Geometric applications of the split bregman method: Segmentation and surface reconstruction. Technical Report 09-06, UCLA, Los Angeles, 2009.

[43] F. Han and S. C. Zhu. Bayesian reconstruction of 3d shapes and scenes from a single image. In *Higher-Level Knowledge in 3D Modeling and Motion Analysis*, pages 12–20, 2003.

[44] T. Hassner and R. Basri. Example based 3d reconstruction from single 2d images. In *Beyond Patches Workshop at IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[45] M. Hatzitheodorou. The derivation of 3-d surface shape from shadows. In *Proc. DARPA Image Underst.*, pages 1012–1020, 1989.

[46] D. Hoiem, A. Efros, and M. Hebert. Automatic photo pop-up. *ACM Transactions on Graphics*, 24(3):577–584, 2005.

[47] W. Hong, A. Yang, K. Huang, and Y. Ma. On symmetry and multiple-view geometry: Structure, pose, and calibration from a single image. *Int. J. of Computer Vision*, 60:241–265, 2004.

[48] R. P. Horaud and M. Brady. On the geometric interpretation of image contours. *Artificial Intelligence*, 37(1-3):333–353, 1988.

[49] B. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA, 1970.

[50] Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pages 225–232, 1997.

[51] T. Igarashi. Adaptive unwrapping for interactive texture painting. In *ACM Symposium on Interactive 3D Graphics*, pages 209–216, 2001.

[52] T. Igarashi. Smooth meshes for sketch-based freeform modeling. In *Proceedings of the symposium on Interactive 3D graphics*, pages 139–142, 2003.

[53] T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: a sketching interface for 3d freeform design. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pages 409–416, 1999.

[54] K. Ikeuchi and B. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–185, 1981.

[55] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, 2003.

[56] P. Joshi and N. Carr. Repoussé: Automatic inflation of 2d art. In *Eurographics Workshop on Sketch-Based Modeling*, 2008.

[57] T. Kanade. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17:409 – 460, 1981.

[58] O. Karpenko, J. Hughes, and R. Raskar. Free-form sketching with variational implicit surfaces. *Computer Graphics Forum*, 21(3):585–594, 2002.

[59] O. A. Karpenko and J. F. Hughes. Smoothsketch: 3d free-form shapes from complex sketches. *ACM Transactions on Graphics*, 25/3:589–598, 2006.

[60] J. Kender and E. Smith. Shape from darkness. In *Int. J. of Computer Vision*, pages 539–546, 1987.

[61] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *Europ. Conf. on Computer Vision*, 2008.

[62] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Continuous global optimization in multview 3d reconstruction. *Int. J. of Computer Vision*, 2009.

[63] K. Kolev, M. Klodt, T. Brox, S. Esedoglu, and D. Cremers. Continuous global optimization in multiview 3d reconstruction. In *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition*, 2007.

[64] P. Koutsourakis, L. Simon, O. Teboul, G. Tziritas, and N. Paragios. Single view reconstruction using shape grammars for urban environments. In *IEEE Int. Conf. on Computer Vision*, 2009.

[65] A. Kushal, S. Sanyal, V. Bansal, and S. Banerjee. A simple method for interactive 3d reconstruction and camera calibration from a single view. In *In Proceedings Indian Conference on Computer Vision, Graphics and Image Processing*, 2002.

[66] J. Lellmann, J. H. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multiclass image labeling by simplex-constrained total variation. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, volume 5567, pages 150–162, 2009.

[67] A. Levin. Analyzing depth from coded aperture sets. In *Europ. Conf. on Computer Vision*, pages 214–227, 2010.

[68] Y. Li, Z. Pizlo, and R. Steinman. A computational model that recovers the 3d shape of an object from a single 2d retinal representation. *Vision Research*, 49:979–991, 2009.

[69] Z. Li, J. Liu, and X. Tang. A closed-form solution to 3d reconstruction of piecewise planar objects from single images. *Int. Conf. on Computer Vision and Pattern Recognition*, pages 1–6, 2007.

[70] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, 1999.

[71] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 21(4):163–169, 1987.

[72] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395, 1987.

[73] D. G. Luenberger. *Optimization by Vector Space Methods.* John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1997.

[74] J. Malik and R. Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *Int. J. of Computer Vision*, 23(2):149–168, 1997.

[75] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers. Video super resolution using duality based TV-L1 optical flow. In *Pattern Recognition (Proc. DAGM)*, 2009.

[76] T. Nagai, T. Naruse, M. Ikehara, and A. Kurematsu. HMM-based surface reconstruction from single images. In *Int. Conf. on Image Processing*, volume 2, pages 561–564, 2002.

[77] S.K. Nayar and E Computer Society. Real-time focus range sensor. In *IEEE Int. Conf. on Computer Vision*, pages 995–1001, 1995.

[78] A. Nealen, T. Igarashi, O. Sorkine, and M. Alexa. Fibermesh: designing freeform surfaces with 3d curves. *ACM Transactions on Graphics*, 26(3):41, 2007.

[79] Y. Nesterov. A method of solving a convex programming problem with convergence rate o(1/sqr(k)). *Soviet Mathematics Doklady*, 27:372–376, 1983.

[80] P. Perona and J. Malik. Scale-space and edge-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[81] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *IEEE Int. Conf. on Computer Vision*, 2009.

[82] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *Europ. Conf. on Computer Vision*, Marseille, France, October 2008.

[83] E. Prados and O. Faugeras. Shape from shading: a well-posed problem? In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 870–877, 2005.

[84] M. Prasad. *Class-based Single View Reconstruction.* PhD thesis, University of Oxford, July 2009.

[85] M. Prasad, A. Zisserman, and A. W. Fitzgibbon. Fast and controllable 3D modelling from silhouettes. In *Eurographics*, pages 9–12, 2005.

[86] M. Prasad, A. Zisserman, and A. W. Fitzgibbon. Single view reconstruction of curved surfaces. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 1345–1354, 2006.

[87] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[88] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, 2004.

[89] D. Rother and G. Sapiro. Seeing 3d objects in a single 2d image. In *IEEE Int. Conf. on Computer Vision*, pages 1819–1826, 2009.

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

[91] A. Saxena, S. Chung, and A. Ng. 3-d depth reconstruction from a single still image. *Int. J. of Computer Vision*, 76, 2007.

[92] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):824–840, 2009.

[93] S. Setzer. Split bregman algorithm, douglas-rachford splitting and frame shrinkage. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 464–476, 2009.

[94] D. Shulman and J.-Y. Herve. Regularization of discontinuous flow fields. In *Workshop on Visual Motion*, pages 81–86, 1989.

[95] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Trans. on Image Processing*, 7:310–318, 1997.

[96] E. Strekalovskiy, B. Goldluecke, and D. Cremers. Tight convex relaxations for vector-valued labeling problems. In *IEEE Int. Conf. on Computer Vision*, 2011.

[97] E. Strekalovskiy, C. Nieuwenhuis, and D. Cremers. Nonmetric priors for continuous multilabel optimization. In *Europ. Conf. on Computer Vision*, 2012.

[98] P. Sturm and S. Maybank. A method for interactive 3d reconstruction of piecewise planar objects from single images. In *Proceedings of the British Machine Vision Conference*, pages 265–274, 1999.

[99] B. Super and A. Bovik. Shape from texture using local spectral moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:333–343, 1995.

[100] D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-seeking models and 3d object reconstruction. *Int. J. of Computer Vision*, 1:211–221, 1987.

[101] F. Ulupinar and R. Nevatia. Shape from contour: Straight homogeneous generalized cylinders and constant cross section generalized cylinders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:120–135, 1995.

[102] S. Utcke and A. Zisserman. Projective reconstruction of surfaces of revolution. In *Pattern Recognition (Proc. DAGM)*, pages 93–102, 2003.

[103] T. Vetter. Synthesis of novel views from a single face image. *Int. J. of Computer Vision*, 28, 1998.

[104] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. on Image Processing*, 7:813–824, 1998.

[105] G. Wang, W. Su, and Y. Song. A new shape from shading approach for specular surfaces. In *Artificial Intelligence and Computational Intelligence*, volume 7004 of *Lecture Notes in Computer Science*, pages 71–78. Springer, 2011.

[106] W. Welch and A. Witkin. Free-form shape design using triangulated surfaces. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pages 247–256, 1994.

[107] L. Williams. 3d paint. *Proceedings of the 1990 symposium on Interactive 3D graphics*, pages 225–233, 1990.

[108] K. Wong, P. Mendonça, and R. Cipolla. Reconstruction of surfaces of revolution from single uncalibrated views. In *Proceedings of the British Machine Vision Conference*, pages 265–272, 2002.

[109] Y. Yu and J. Chang. Shadow graphs and surface reconstruction. In *Europ. Conf. on Computer Vision*, pages 31–45, 2002.

[110] C. Zach, D. Gallup, J.-M Frahm, and M. Niethammer. Fast global labeling for realtime stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*, 2008.

[111] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Pattern Recognition (Proc. DAGM)*, LNCS, pages 214–223. Springer, 2007.

[112] R. Zeleznik, K. Herndon, and J. Hughes. Sketch: An interface for sketching 3d scenes. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pages 163–170, 1996.

[113] G. Zeng, Y. Matsushita, L. Quan, and H. Shum. Interactive shape from shading. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 343–350, 1999.

[114] L. Zhang, G. Dugas-Phocion, J. Samson, and S. Seitz. Single view modeling of free-form scenes. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 990–997, 2001.

[115] R. Zhang, P. Tsai, J. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

[116] C. Zhou, S. Lin, and S. Nayar. Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. *Int. J. of Computer Vision*, 93(1):53, May 2011.

Appendix: Derivation of the Shrinkage Formula

In the following we will derive the closed form solution of the following minimization problem

$$\underset{v}{\operatorname{argmin}} \|v\|\, g + \frac{\tau}{2} \|v - x\|\, 2^2 = \operatorname{prox} \frac{1}{\tau}\Phi(x) =: \operatorname{shrink} g(x, \tau)$$

where

$$\Phi(v) = \|v\|\, g = \sum_{i \in \Omega} g_i \|v_i\|\, 2$$

First, we will calculate the convex conjugate $\Phi^*(v)$. It is given by the indicator function of the dual norm ball. The dual norm is given by

$$\|v\|\, g^* = \sup_{i \in \Omega} g_i \|v_i\|\, 2$$

so we arrive at

$$\Phi^*(v) = \iota_C(v) \qquad C = \{v \,:\, \|v_j\|\, 2 \leq g_j,\, \forall j \in \Omega\}$$

Now we have by using Moreau's identity in the second step

$$\operatorname{prox} \frac{1}{\tau}\Phi(x) = x - \frac{1}{\tau} \operatorname{prox} \frac{1}{\tau}\Phi^*(\tau x) = x - \frac{1}{\tau} \operatorname{prox}_{\mathbf{C}}(\tau x)$$

Point–wise this evaluates to

$$(\operatorname{prox} \frac{1}{\tau}\Phi(x))_i = \begin{cases} 0 & \|x_i\|\, 2 \leq \frac{g_i}{\tau} \\ x_i - \frac{g_i}{\tau} \frac{x_i}{\|x_i\| 2} & \text{else.} \end{cases}$$