Convexity in Image-Based 3D Surface Reconstruction

Thesis for obtaining the degree of doctor of natural sciences (Dr. rer. nat.)
of the faculty of computer science of the Technical University Munich

Kalin Kolev

Garching, Germany
December 7, 2011

Thesis Advisor: Prof. Dr. Daniel Cremers (TUM)
External Referee: Prof. Dr. Marc Pollefeys (ETH)
Committee Chairman: Prof. Dr. Nassir Navab (TUM)

Computer Vision Group, Department of Computer Science
Technical University Munich, 85748 Garching, Germany

`http://wwwcremers.in.tum.de/`

Foreword

The real voyage of discovery consists not
in seeking new landscapes, but in having new eyes.

Marcel Proust (1871-1922)

As the title of the current thesis suggests, convex modeling will play a central role throughout the work. Yet, convexity is utilized not as a construct to give a scientific form to the thesis. I will try to convince the skeptic readers that, along with mathematical elegance, it provides the capabilities to build robust and accurate approaches of high practical value.

One could argue that focusing on convex formulations limits the sight and does not allow to explore the whole potential of the application. In particular, it is debatable if the starting point of developing novel methods should be the optimization or the structure of the model. Accordingly, one can distinguish between two different research philosophies. While I was following "optimization-oriented" principles, there are of course many alternative approaches which may be equally valid. As always in research, there is not just one way to reach a goal.

Although special attention was attached to the intelligibility of the thesis, the reader needs some basic knowledge of analysis, variational calculus and epipolar geometry to understand it.

Even though I am the sole author of the thesis, it is written in first person plural. The first reason for that is that it is an established form in scientific publications. The second reason is that there are multiple people who supported me during my research and contributed to the work.

First of all, I want to thank the "usual suspect" – my supervisor Prof. Daniel Cremers. I would like to thank him for teaching me to think like a researcher, to write things down and to present my work in a convincing and understandable way. Actually, he was for me more than a supervisor by supporting me during all the time and never losing faith in me and my capabilities. Moreover, all the colleagues, who I worked with during my PhD period, should be acknowledged – Thomas Brox, Thomas Pock, Selim Esedoglu, Maria Klodt, Thomas Schoenemann and Mathieu Aubry. I thank them for the very fruitful collaboration. I also want to thank all the students who I supervised – Björn Sondermann, Haj Bensouda, Svetlana Matiouk and Pascal Wauer. I further want to thank the German research foundation DFG for funding my research.

Science is not everything in life, even for a scientist, and therefore my final thanks are addressed to those people who always supported me. These are, in particular, my parents, my sister and her family.

Garching, June 12, 2011 Kalin Kolev

Abstract

This work deals with the application of convex optimization techniques to the context of image-based 3D surface reconstruction, in particular shape from silhouettes, multiview stereo and 3D segmentation. The focus is thereby not only on the optimization, but also on the design of appropriate energy functionals to be minimized. Most of the chapters of this thesis are devoted to the construction of different energy models addressing various 3D reconstruction problems. In all cases, the modeling results in a convex optimization problem or involves handling convex subproblems. Numerical algorithms for solving them are presented and discussed in a separate chapter.

The current work starts with a brief introduction to the fields of multiview 3D reconstruction and 3D segmentation, two fundamental computer vision problems, and establishes the relation between them. In addition, it treats the main concept of convexity which forms the theoretical basis for all energy models considered in this manuscript. Special attention is paid to the conceptual differences between continuous and combinatorial optimization, in particular in terms of metrication accuracy, parallelizability and memory requirements. The demonstrated advantages of continuous techniques pose the main motivation for studying convex models.

Chapter 2 proposes an interactive framework for color-based 3D reconstruction from multiple images. For a user-provided input in the form of scribbles specifying foreground and background regions, corresponding color distributions are built as multivariate Gaussians and subsequently a 3D surface, that best fits to this data in a variational sense, is derived. The goal of the proposed probabilistic formulation is to enable robust 3D estimation by optimally taking into account the contribution of all views. Compared to classical methods for shape from silhouettes, the presented approach does not depend on initialization and enjoys significant resilience to violations of the model assumptions due to background clutter, specular reflections and camera sensor perturbations. Furthermore, we will see that exploiting a silhouette coherency criterion in a multiview setting allows for dramatic improvements of silhouette quality over independent 2D segmentations without any significant increase of computational efforts.

While color-based 3D reconstruction methods are fast and robust, their accuracy is limited. In particular, they are unable to recover surface concavities since these do not affect the projected object outlines. A remedy to this weakness offer multiview stereo approaches which aim at identifying view points corresponding to the same 3D geometry and allow for a much more precise localization of the imaged structure. Chapter 3 proposes three different energy models for multiview stereo amenable to convex optimization. They are based on a common variational template unifying regional volumetric terms and on-surface photoconsistency and use data measurements at increasing levels of sophistication. While the first two approaches are based on a classical silhouette-based volume subdivision, the third one introduces the concept of propagated photoconsistency, thereby addressing some of the shortcomings of traditional methodologies. Further, in Chapter 5, we will see how this model can be generalized by anisotropic regularization schemes and minimal ratio formulations. The proposed generalization entails additional useful properties like capabilities to integrate surface normal information, absence of a shrinking bias and scale invariance.

Unfortunately, multiview stereo as a paradigm also has its limitations and failure cases. First of all, the success of stereo-based approaches strongly depends on the viability of the matching process which, in turn, relies on Lambertian reflectance properties of the scene. Yet, lack of texture, unknown projective distortion and deviations from the Lambertian assumption make the matching of corresponding image points to a quite challenging task. These difficulties naturally lead to the idea of constructing algorithms that integrate multiview stereo and silhouettes aiming at reaching the robustness of shape-from-silhouette methods while retaining the accuracy of stereo-based ones. Chapter 4 deals with the design of such an approach and demonstrates that the fusion of stereo and silhouette information can be formulated as a problem of minimizing a convex functional over a convex domain, where the cost measure favors photoconsistent surfaces while the provided silhouettes are used to define a set of feasible shapes. Thus, the modeling boils down to solving a classical constrained convex optimization problem.

Chapter 6 extends the application field of the thesis by considering the problem of 3D segmentation in fluorescence imaging. Once again, convex optimization proves useful, allowing for accurate segmentations of the provided noisy volumetric measurements. An appropriate energy model fusing various cues like regional intensity subdivision, edge alignment and orientation information, tailored to the particular application at hand, is built and analyzed. Yet, despite this concrete scenario, the proposed approach is general enough to be applied to a variety of different segmentation tasks.

The work concludes with a summary of the main contributions and outlines directions for future research.

Contents

1

Introduction

A theory that fits all the facts is bound
to be wrong as some of the facts will be wrong.

Francis Crick (1916-2004)

As technology becomes more powerful, attention is being focused on the capability to virtually reproduce the surrounding three-dimensional real world. As a consequence, a great effort is being made to enable the creation and acquisition of high-quality 3D models. Application fields range from entertainment industry to robotics, cultural heritage and medicine.

A straightforward but very tedious and time-consuming way to obtain a 3D model of an existing object is to generate the observed 3D structure by hand, using a CAD[1] tool. Apart from its high costs, the precision of this technique is limited since it is based on visual perception rather than physical measurements. Even though manual modeling is widely-used in the movie and game industry, it is not interesting for application domains, where the physical accuracy of the 3D models is essential. Furthermore, the tremendous amount of human efforts needed makes this brute-force approach practically inapplicable for modeling large-scale scenes like cities and landscapes.

An alternative method for the automatic and highly precise acquisition of 3D models offer 3D scanning technologies. Among the most popular devices are laser range scanners [83] and structured-light scanners [146]. The functionality of a laser scanner is based on a laser light and a triangulation principle to probe the environment. The device shines a ray on the subject and exploits a camera to look for the location of the projected dot. Depending on how far away the laser strikes a surface, the laser dot appears at different places in the camera's field of view. In most situations, a single scan does not produce a complete model of the subject. Multiple scans, even hundreds, from many different directions are usually required to obtain information about all sides of the modeled object. Finally, all range scans have to be registered within the same coordinate system, which is generally not a trivial task. Structured-light 3D scanners project a pattern of light on the subject and look at the deformation of the pattern on the surface. This scanning technique is considerably faster since multiple points or the entire field of view is being scanned at once. While 3D scanners can produce highly accurate models, they typically require expensive hardware or need to deploy a specialized equipment. Moreover, they have a quite limited range and often require controlled illumination conditions, which poses a major difficulty in using them for modeling large-scale real-world scenes.

The aforementioned weaknesses of traditional 3D scanning techniques led to the development of time-of-flight (TOF) cameras.[2] A TOF camera has a variety of advantages over alternative scanning technologies [27]. It is an active sensor that measures the travel time of infrared

---

[1] Computer Aided Design

[2] Recently distributed Kinect cameras [90] can also be classified in this category.

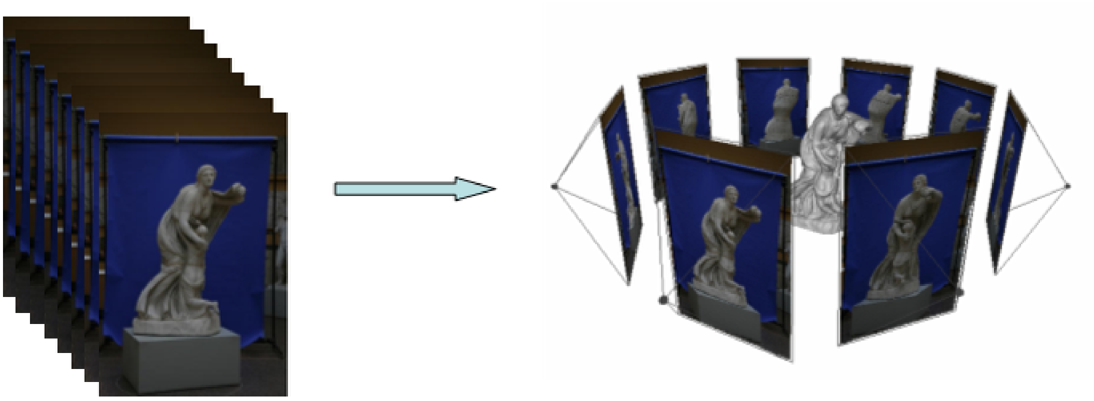


Fig. 1.1: Image-based 3D modeling. Image-based modeling approaches aim at estimating the 3D shape of a given object from a collection of images capturing its appearance from various viewpoints. The multiview reconstruction pipeline is usually split into two stages. In the first one, the input images are calibrated, i. e. the position and orientation of the utilized cameras as well as their intrinsic parameters are estimated. In the second stage, the imaged 3D geometry is derived, taking the particular camera setup into account.

light, and therefore it does not interfere with the scene in the visual spectrum. This makes TOF cameras quite practical for outdoor scenarios since they do not require a controlled illumination environment. Moreover, TOF cameras are capable of operating over very long distances, which strengthens their suitability for modeling of large-scale scenes. Finally, core components of such a device are a CMOS chip and an infrared light source which bears the potential for low cost manufacturing. However, TOF cameras have a major disadvantage – their accuracy is limited. In particular, they have a very low resolution, an adverse random noise behavior, and a substantial systematic measurement bias. Some attempts have been made to address these shortcomings by fusing TOF sensor measurements and traditional binocular stereo depth-maps [147], but it is unclear how to estimate the reliability of both sources in order to guarantee an adequate integration process.

The complicacies faced when using active scanning technologies focused the attention of many researchers on image-based techniques for 3D modeling. The formidable progress in multiview 3D reconstruction in recent years [112, 122], along with the rapid developments in digital photography, makes vision-based methods more and more attractive for generating fast, easily and reliably 3D content.

1.1 Image-Based Modeling

Problem Statement

Image-based modeling approaches use digital photographs to model the geometry of an observed 3D scene (see Fig. 1.1). The problem of reconstructing 3D shapes from a collection of images is one of the fundamental problems in computer vision since understanding the image formation process is at the core of many vision tasks. Yet, with the explosion of consumer digital photography and the revolution in multiview 3D reconstruction, the problem emerged to one of the most promising and extensively studied alternatives to active scanning. Nowadays, the interest in image-based modeling reaches far beyond the field of computer vision and attracts the attention of researchers in graphics, robotics, archeology and medicine.

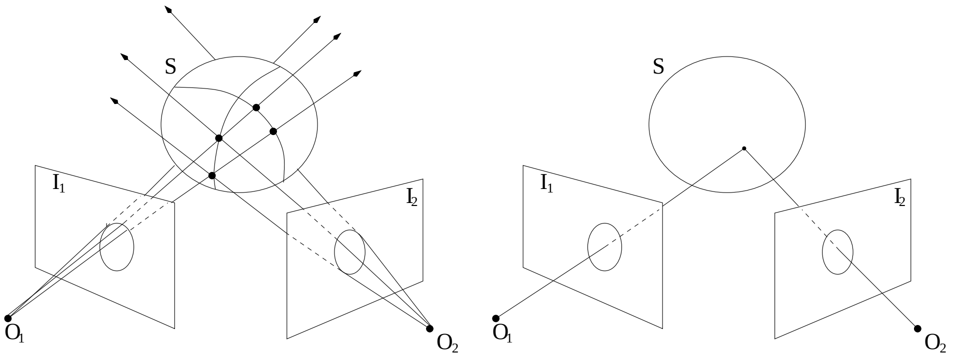
The multiview reconstruction pipeline is usually split into two stages. In the first one, the input images are calibrated, i. e. the position and orientation of the utilized cameras as well as their intrinsic parameters are estimated. In the second stage, the imaged 3D geometry is derived, taking the particular camera setup into account. The problem of automatic camera calibration from photographs has undergone an immense exploration in the last and the first half of the current decade and could, to some extend, be considered as geometrically solved. Several good books on that topic [50, 37], as well as publicly available software [1, 2, 3], exist. Yet, the 3D modeling still poses a major challenge. For that reason, in this work we assume that the input images are calibrated in a preprocessing step and concentrate on the modeling stage of the reconstruction pipeline. For some of the utilized image sequences calibration information was already provided. For the rest of the data sets we used the Bundler software [1] to estimate the calibration.

Image-based modeling approaches, also referred to as *shape-from-X*, can be classified according to the exploited visual cues. Some of the most popular domains are:

- Shape from silhouettes [79]
- Shape from stereo [112, 122]
- Shape from texture [138]
- Shape from shading / photometric stereo [53, 139]
- Shape from focus/defocus [96]

The exploration of all these cues is motivated by the investigation of the human perceptual system. It should be noted that some of these techniques are *passive* while others are *active*. For example, approaches exploiting shading information require controlled illumination conditions and approaches based on changing focus operate on pictures taken with different focus settings from the same viewpoint. On the other hand, shape-from-silhouette and multiview stereo methods use only the raw image data as input. Though classical shape-from-silhouette algorithms employ object outlines in binary form, others like the one described in Chapter 2 use only image color information. Shape from texture comprises a particular class of approaches that are passive in nature, as they do not require controlled environments, but they are restricted to a very narrow class of objects. While active approaches can achieve a high precision, comparable to that of laser scanners, they are not applicable to many real-world scenarios due to the specific imaging conditions required. In contrast, passive methods are very general and bear the potential for satisfying the growing demand for digital 3D models, but still suffer from limited accuracy. Closing the gap between active and passive image-based modeling, by reaching the accuracy of active techniques while retaining the generality of passive ones, is a central problem in the field.

One issue deserves more attention when dealing with image-based modeling. It is important to note that the problem is ill-posed since the projection mapping is not invertible. As a consequence, there are multiple possible solutions that could give rise to the provided observations. In order to make the reconstruction problem meaningful, certain assumptions about the imaged geometry have to be made. Each utilized image cue in shape-from-X approaches implies a particular assumption regarding the 3D scene. Nevertheless, some ambiguities remain. For example, shape-from-texture methods have to cope with a two-fold texture normal ambiguity, shape-from-shading methods – with a convex/concave ambiguity and shape-from-silhouette methods – with infinitely many solutions being consistent with a given set of silhouettes. Although different image cues could be combined to resolve these ambiguities [26, 137], an integration process could lead to contradictory shape estimation due to image noise or model inaccuracies and is not always preferable.



silhouette consistency

stereoscopic matching

Fig. 1.2: Silhouettes vs. stereo. *Left:* Silhouette consistency constraints enforce the surface projection to coincide with the observed object outlines. *Right:* Stereo-based constraints maximize the correlation of the visible projections for each point on the surface.

In this work we concentrate on generic image-based modeling, in particular based on silhouettes and multiview stereo. In the following, we give more details on these two paradigms.

Silhouette Consistency

Let $I_1, \dots, I_n$ be a collection of input color images. In particular, $I_i : \Omega_i \to \mathbb{R}^3$ for $i = 1, \dots, n$, where Ω_i denotes the domain of image i, usually a pixel grid, i. e. $\Omega_i \subset \mathbb{Z}^2$. Further, let $V \subset \mathbb{R}^3$ be a restricted volume area enclosing the scene of interest and $\pi_i : V \to \Omega_i$ the projection mapping of camera i. Since we assume precalibrated cameras, the projections π_i are given for all $i = 1, \dots, n$.

The goal of silhouette-based reconstruction approaches is to find a surface $S \subset V$ such that its projection onto the images coincides with the observed silhouettes of the object (see Fig. 1.2). More concretely, if $Sil_i \subset \Omega_i$ denotes the object silhouette in image i (which can be obtained through segmentation), the estimated surface S should fulfill the following silhouette constraints

$$\pi_i(S) = Sil_i \qquad \forall\, i = 1, \dots, n. \tag{1.1}$$

Obviously, this criterion does not lead to a unique solution in general. This issue has been investigated by Laurentini in [79], where he coined the term *visual hull* which is the largest shape that fulfills the silhouette constraints in (1.1). In fact, most of the existing silhouette-based approaches estimate the visual hull of the object or some approximation of it. Maybe, silhouettes pose the most intuitive image cue to infer 3D geometry from observed projections. Not surprisingly, the very first method for multiview 3D reconstruction is a silhouette-based one [8]. Due to their simplicity, shape-from-silhouette techniques feature a high degree of stability and efficiency and are still preferred in some applications like robot navigation and tracking. Their main disadvantage is the inability to reconstruct surface concavities since these do not affect the outlines of the imaged object. For that reason, they are not capable of delivering high-quality 3D models of arbitrary objects. Yet, silhouettes exhibit a very powerful image cue and are successfully used in combination with alternative cues like stereo [26] and shading [133]. In some cases, a rough silhouette-based reconstruction is used as an initialization. In others, the silhouette constraints in (1.1) are used in a hard or soft form to restrict the set of admissible shapes.

Multiview Stereo

The basic idea behind multiview stereo is to infer 3D geometry by matching image points corresponding to the same spatial structures (see Fig. 1.2). For given camera calibration, the underlying 3D points can be obtained via triangulation of the respective visual rays. It is important to note that the stereo concept is based on *Lambertian* reflectance properties of the imaged object, i. e. the reflected intensity at each point on the surface does not depend on the viewing direction. While this assumption is justified for a wide range of real-world objects, it breaks down for shiny or mirroring objects which follow a more complex reflectance model. Hence, such test cases pose a formidable challenge for multiview stereo approaches. Yet, stereo-based methods enjoy very high popularity due to their generality and fidelity.

To elucidate the principle of multiview stereopsis, let us assume that the observed surface is parametrized as $S : \Theta \subset \mathbb{R}^2 \to V$.[3] Essentially, all stereo-based methods aim at minimizing the variance of the following color samples

$$C(s) = \{ \, I_i(\pi_i(S(s))) \mid i \in Vis(s) \, \} \tag{1.2}$$

for all $s \in \Theta$. Thereby, $Vis(s) \subset \{1, \ldots, n\}$ denotes the *visibility* of point $S(s)$, i. e. the indices of cameras that have a direct visual contact to the point. There are two important aspects that should be tackled when designing a stereo-based method – visibility estimation and realization of the variance of the projection sets (1.2).

In multiview stereo, accurate visibility estimation is crucial. Different strategies were proposed in the literature to handle the problem. The most popular approach is the state-based one [125, 33]. The key idea is to integrate the visibility computations in a local surface evolution process. Thereby, the current surface estimate serves as a basis for visibility considerations in subsequent data term updates. While this simple technique works reliable if a good surface initialization is provided, they may fail in case of a poor initialization. A different straightforward approach is to exploit the usually known camera calibration to identify cameras which are geometrically close to each other and are likely to see approximately the same part of the scene. It is typically referred to as oriented visibility [81, 134]. This strategy gives good results for convex objects but could produce inaccuracies at concavities and locations of abrupt depth changes. Another methodology envisages to intertwine the reconstruction and the visibility estimation process [41, 84]. Thereby, visibility computations are based on a series of heuristics and are performed only for points on the surface. Even though such methods could lead to very accurate reconstructions, their reliability is doubtful in challenging situations. Maybe the most transparent way to solve the visibility problem is to explicitly incorporate it in the cost function [43, 28]. Despite rigorousness, this formulation entails multiple optimization difficulties. In particular, the arising minimization problems are typically highly non-convex and thus susceptible to undesired local solutions. Interestingly, there is one strategy that avoids direct confrontations with the visibility problem – the voting schemes proposed in [33, 132]. The basic idea is to let each camera give a vote for the state of a point in space without any visibility considerations. The fusion scheme of the various votes is implemented in a way that encourages occluded cameras to be discarded. In Chapter 3, we give more details on this approach since a variant of it is used.

At the core of multiview stereo is the minimization of the variance of the projection sets (1.2) for all points on the hypothetical surface. Traditionally, this is realized by means of the so called *photoconsistency* map $\rho : V \to (0, 1]$. It assigns each point within the volume V a cost

---

[3] This assumption is made here for the sake of simplicity. In fact, most of the existing multiview stereo approaches do not rely on an explicit surface parametrization. This issue will be discussed in more detail in the course of the current work.

measure proportional to the variance of the corresponding projection set. As a consequence, high values (close to 1) imply low probability for lying on the surface and low values (close to 0) – high probability. Hence, the multiview stereo problem consists in finding the surface with the best overall photoconsistency. Usually, the photoconsistency is defined by means of *similarity measures* which express the correlation of different points in different images.
In many practical cases, direct comparison of the colors for different image projections is not reliable due to color similarities of different objects, image noise or calibration inaccuracy. As a remedy, correlating entire local image neighborhoods instead of single points has become an established strategy. To this end, various similarity measures have been investigated. Here, we introduce two of the most popular tools – *sum of squared differences* (SSD) and *normalized cross correlations* (NCC). Let $p_i \in \Omega_i$ and $p_j \in \Omega_j$ be two pixels in two different images that should be tested on similarity. Let $\mathcal{P}(p_i, p_j) \subset \Omega_i \times \Omega_j$ be a set of pixel pairs describing corresponding patch neighborhoods around both pixels, in particular $(p_i, p_j) \in \mathcal{P}(p_i, p_j)$. For example, for simple square patches of size $(2m+1) \times (2m+1)$, centered at $(0,0)$, we have $\mathcal{P}((0,0),(0,0)) = \{-m,\dots,m\}^2 \times \{-m,\dots,m\}^2$.[4] More complex methodologies taking three-dimensional information into account are also possible. Now, the SSD score can be defined as

$$SSD(p_i, p_j) = \sum_{(o_i,o_j)\in\mathcal{P}(p_i,p_j)} |I_i(o_i) - I_j(o_j)|^2. \tag{1.3}$$

Here, not only the image points themselves are involved in the computation, but also points within the entire neighborhoods. The SSD-measure is very simple and works well in many cases. Yet, its robustness to image noise and deviations from the Lambertian assumption is limited. For example, specular reflections or inaccurate color calibration of the cameras lead to poor performance due to the requirement of a direct color accordance. This flaw is addressed by the NCC-measure which offers considerably more resilience to illumination variations. It is given by

$$NCC(p_i, p_j) = \sum_{(o_i,o_j)\in\mathcal{P}(p_i,p_j)} \frac{\langle I_i(o_i) - \bar{I}_i(p_i), I_j(o_j) - \bar{I}_j(p_j)\rangle}{\sqrt{\sum_{(o_i,o_j)\in\mathcal{P}(p_i,p_j)} |I_i(o_i) - \bar{I}_i(p_i)|^2}\sqrt{\sum_{(o_i,o_j)\in\mathcal{P}(p_i,p_j)} |I_j(o_j) - \bar{I}_j(p_j)|^2}}, \tag{1.4}$$

where $\langle .,. \rangle$ denotes the Euclidean scalar product and $\bar{I}_i(p_i)$, $\bar{I}_j(p_j)$ are the color means within both patches. It could be noted that the NCC score is invariant to additive and multiplicative illumination changes and thus compares the local image structure instead of the color occupancy. The first is due to the subtraction of the mean value and the latter – due to the normalization. Moreover, the NCC score always takes a value in the interval $[-1, 1]$, where 1 means perfect matching and -1 perfect mismatching, respectively. The NCC-measure is both robust and fast to evaluate, which is the reason for its enormous popularity in the multiview stereo community. In fact, it has become to a standard tool in dense stereo-based modeling.

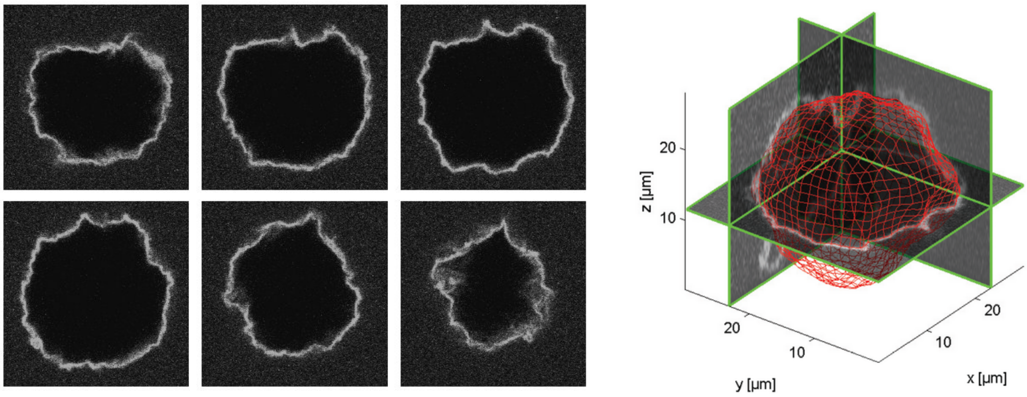
Challenges and Contributions

This work intensively deals with the image-based modeling problem. The goal is to increase the robustness and extend the range of applicability while retaining high accuracy. In particular, we focus on the following challenges:

- How can any dependence from an initial shape be removed in a way that prevents premature termination and always guarantees a reasonable result ?

---

[4] Boundary considerations are ignored here for the sake of simplicity.



input volume slices

superimposed segmentation

Fig. 1.3: Volumetric 3D segmentation. *Left:* Multiple input volume slices, *right:* superimposed segmentation result. The example shows a typical 3D segmentation problem arising in biological and medical applications. The volume data captures the deformation of a vesicle membrane under osmotic pressure, imaged by means of fluorescence microscopy.

- How should the methods be designed in order to make them more user-friendly (reducing the number of parameters and any user efforts needed to run the algorithms) ?
- How can the process of silhouette-based reconstruction and image segmentation be coupled with the objective of removing the necessity of preprocessing the images ?
- How can a multiview stereo system cope with objects violating the primary assumptions, e. g. exhibiting specular reflections, lack of texture and thin geometric structures ?
- How can complementary image cues like stereo and silhouettes be integrated in a robust and reliable way ?

In this thesis, we explore these issues and provide answers. We show that robust convex regularization schemes could provide the necessary mathematical machinery to handle objects with complex reflectance properties, e. g. shiny or weakly textured. Furthermore, approaches that involve solving a strictly convex problem at a primary or intermediate stage always lead to a unique solution, independent from initialization, which substantially increases their robustness. We show how complex geometrical structures like fine-scale details and thin protrusions can accurately be recovered by integrating stereo and silhouette information. Once again, we rely on the powerful convex optimization framework, whereas the photometric consistency affects the underlying cost function and the silhouette consistency enters in form of convex constraints restricting the domain of feasible solutions. Moreover, we pay special attention to the design of appropriate variational models amenable to convex analysis with focus on rigorousness, robustness and accuracy.

1.2 3D Segmentation

Problem Statement

The problem of 3D segmentation is closely related to the problem of multiview 3D reconstruction. It consists in partitioning a given three-dimensional domain into coherent, semantically homogeneous regions. In an alternative formulation, the problem transforms to one of finding a surface interface accomplishing the desired subdivision (see Fig. 1.3 for an example). More specifically, if $V \subset \mathbb{R}^3$ denotes a given volume, the goal is to find a partitioning $V_1, \dots, V_M$ with the following properties:

- $V = \bigcup_{i=1}^{M} V_i$.
- $V_i \cap V_j = \emptyset \qquad \forall\, i, j \in \{1, \dots, M\}, \quad i \neq j$.
- Each V_i is homogeneous. The homogeneity criterion depends on the particular segmentation task.

Obviously, one is confronted with similar optimization challenges as in the case of image-based modeling and thus the same theoretical tools could be applied.

The segmentation problem is one of the most classical and intensively investigated in computer vision. Various cues were explored to address it. In the example in Fig. 1.3, one can observe that characteristic for the interior of the vesicle are low homogeneous intensity values while the exterior region exhibits points of slightly higher intensity and substantially more noise. Hence, interior/exterior subdivision can be achieved by modeling the intensity statistics of both regions. In fact, *region-based* characteristics are an established cue in image segmentation [94, 22, 148]. Furthermore, the membrane surface in Fig. 1.3 is recognizable through its brightness due to the fluorescence properties of the building material. Thus, a reasonable segmentation approach should attract the estimated boundary towards such locations. This observation forms the basis of *edge-based* segmentation methods [59, 19, 62]. Finally, the membrane in Fig. 1.3 can be identified by estimating its orientation. In particular, the surface normal is aligned with the intensity gradient since the boundary features an abrupt intensity jump (from dark within the interior to bright at the boundary). The integration of orientation information is at the core of *flux-based* segmentation approaches [130]. The exploration of all discussed image cues is motivated by the human perceptual system. In real-world images, objects can often be identified through its regional characteristics like intensity/color or texture. On the other hand, this implies that the boundaries stand out as locations of remarkable image gradient, which can be used as edge information in the segmentation. Moreover, higher order shape characteristics like normal information are indispensable for the precise reconstruction of thin structures like blood vessels.

Challenges and Contributions

While all aforementioned image cues prove useful in particular application scenarios, their integration is not a trivial task. The construction of a general approach, applicable to a wide range of segmentation problems, remains an active research domain. Another crucial challenge is the choice of surface representation which is closely related to the underlying optimization technique. Desired are properties like topological flexibility, efficiency and robustness to local minima.

In Chapter 6 we tackle the problem of segmentation of vesicle membranes from fluorescence 3D imaging. The main contributions are summarized in the following:

- We unify various established image cues like regional statistics, edge attraction and orientation information into a continuous convex energy model.
- We generalize traditional regularization schemes taking the sparse slicing of the measured volume data in the z-direction, inherent in confocal microscopy, into account. We show that the generalized model retains convexity and thus all globality guarantees.

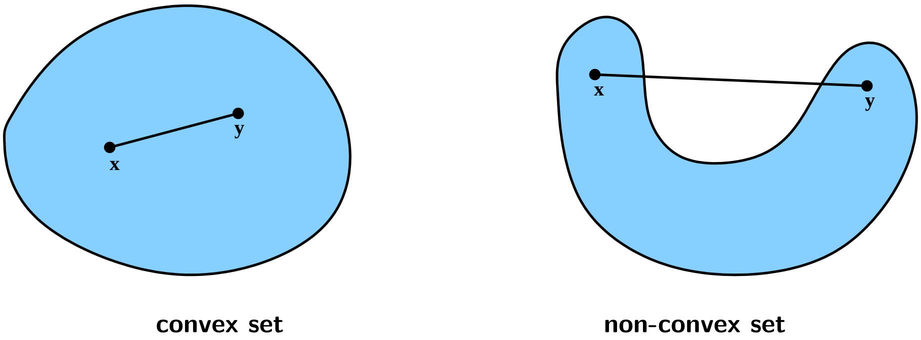


Fig. 1.4: Convex vs. non-convex sets. *Left:* A convex set C is characterized by the property that for each pair of elements $x, y \in C$ the line segment connecting them is also contained in C. *Right:* For a non-convex set at least one element pair exists that violates this condition.

- We adapt an efficient numerical scheme for solving the arising optimization problem.

As for image-based modeling, we make use of convex analysis to address the respective segmentation problem. It is important to note that although the proposed approach is designed with focus on a concrete application at hand, it is general enough to be applied to a variety of different segmentation tasks.

1.3 Convexity as a “Well-Posed” Mathematical Concept

As the title of the thesis suggests, convexity plays a central role in the mathematical modeling of the addressed problems. Convex optimization has become an appealing tool for solving a variety of vision tasks like image segmentation [15, 21], image denoising/deblurring [21, 58] and camera calibration [145]. An exhaustive exposition of convex analysis is out of the scope of the work. Here, we allude to the concepts which are relevant to the understanding of the proposed theory. We refer to [108, 10, 52] for a detailed introduction to convex optimization.

Convex Sets

We start with a definition of a convex set.

Definition 1. *A set C is said to be convex if $(1-\lambda)x+\lambda y \in C$ for all $x, y \in C$ and $\lambda \in (0,1)$.*

In other words, a given set is convex if for each pair of elements the line segment connecting them is also contained in the set (see Fig. 1.4). Half-spaces are important examples of convex sets. For any non-zero $d \in \mathbb{R}^N$ and any $\delta \in \mathbb{R}$, the closed half-spaces

$$\{\, x \mid \langle x, d\rangle \le \delta \,\} \qquad \{\, x \mid \langle x, d\rangle \ge \delta \,\}$$

as well as the open half-spaces

$$\{\, x \mid \langle x, d\rangle < \delta \,\} \qquad \{\, x \mid \langle x, d\rangle > \delta \,\}$$

are convex.

Theorem 1. *If $C_1, \dots, C_M$ is a collection of convex sets, the intersection $C = \bigcap_{i=1}^{M} C_i$ is also a convex set.*

Proof. For arbitrary $x, y \in C$ and $\lambda \in (0,1)$, we have $(1-\lambda)x+\lambda y \in C_i$ for all $i = 1, \dots, M$ due to the convexity of the sets. Now, we can deduce $(1-\lambda)x+\lambda y \in C$. $\square$

Corollary 1. *Let* $d_i \in \mathbb{R}^N$ *and* $\delta_i \in \mathbb{R}$ *for* $i \in I$, *where* I *is an arbitrary index set. Then the set*

$$C = \{\, x \in \mathbb{R}^N \mid \langle x, d_i \rangle \leq \delta_i, \forall i \in I \,\}$$

is convex.

Proof. Let $C_i = \{\, x \mid \langle x, d_i \rangle \leq \delta_i \,\}$. Then C_i is a closed half-space or $\mathbb{R}^N$ or $\emptyset$ and $C = \bigcap_{i \in I} C_i$. □

Of course, the conclusion would still be valid if some of the inequalities $\leq$ were replaced by $\geq$, $<$ or $>$. The above statement is significant in practice and is exploited in Chapter 4.
Convex sets have many useful properties. In the following, we summarize some of the most notable ones.

Property 1. *If* C *is a convex set in* $\mathbb{R}^N$, *then so is every translate* $C + a = \{\, x + a \,|\, x \in C \,\}$.

Property 2. *If* C *is a convex set in* $\mathbb{R}^N$, *then so is every scalar multiple* $\lambda C = \{\, \lambda x \,|\, x \in C \,\}$.

Property 3. *If* C_1 *and* C_2 *are convex sets in* $\mathbb{R}^N$, *then so is their sum* $C_1 + C_2$, *where*

$$C_1 + C_2 = \{\, x_1 + x_2 \,|\, x_1 \in C_1, x_2 \in C_2 \,\}.$$

Property 4. *If* C *is a convex set and* $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, *then*

$$(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C.$$

Each of these properties is easy to verify. Note that the last property is not valid for a general set C.

Convex Functions

In this paragraph, we introduce convexity for functions. Even though we consider only functions of the form $f : \mathbb{R}^N \to \mathbb{R}$ for the sake of simplicity, the generalization of the presented theory to a broader setting of functional analysis is straightforward. Note that a functional over a domain of continuous surfaces boils down to a N-dimensional function after discretization, where N gives the resolution of the underlying volume grid.

Definition 2. *A function* $f : \mathbb{R}^N \to \mathbb{R}$ *is called convex if for any* $x, y \in \mathbb{R}^N$ *and* $\lambda \in (0, 1)$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y).$$

If we have a strict inequality, i. e. the sign $\leq$ is replaced by $<$, the function is called *strictly convex.* Geometrically, the above definition states that the graph of a convex function lies below the line segment connecting each two points on it.
Now, we come to the central property of convex functions, which makes their exploration so attractive.

Theorem 2. *For a convex function* $f : \mathbb{R}^N \to \mathbb{R}$, *each local minimum is also a global minimum.*

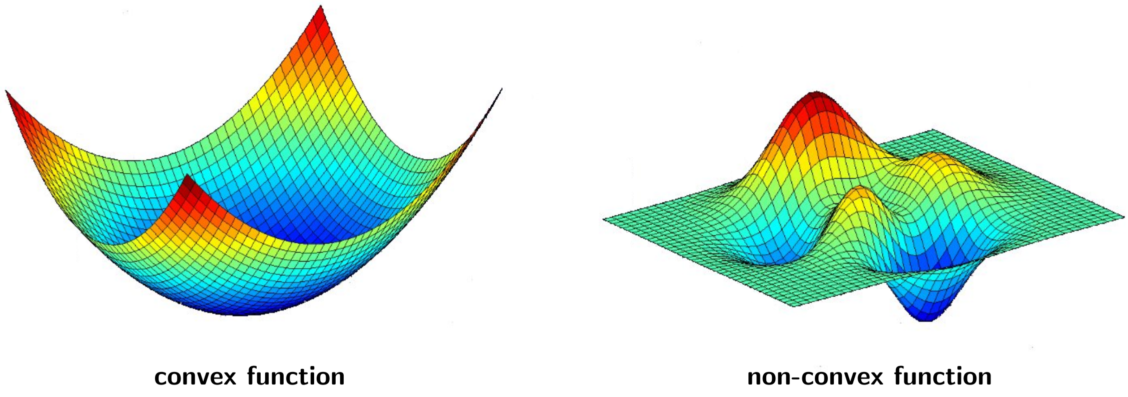


Fig. 1.5: Convex vs. non-convex functions. *Left:* For a convex function, each local minimum is also a global minimum. As a consequence, each local optimization procedure gives a globally optimal solution for a convex optimization problem. *Right:* A general non-convex function attains multiple different local minima. Hence, the result of an optimization method highly depends on the initialization.

Proof. Let $x \in \mathbb{R}^N$ be a local minimum of f. Then we have $f(x) \leq f(z)$ for any z in some neighborhood $U \subset \mathbb{R}^N$ of x. For any $y \in \mathbb{R}^N$, $\lambda x + (1-\lambda)y \in U$ for $\lambda < 1$ sufficiently close to 1, and since f is convex and x is a (local) minimizer

$$\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y) \geq f(x).$$

This implies $f(x) \leq f(y)$. □

The above theorem is crucial in convex analysis (see Fig. 1.5). Basically, it states that each convex optimization problem can easily be solved in a globally optimal manner. In particular, every local optimization procedure gives a globally optimal solution. This property has theoretical as well as practical consequences. It substantially increases the robustness of the underlying algorithms.

Next, we list some important properties of convex functions.

Property 1. *A function $f : \mathbb{R}^N \to \mathbb{R}$ is convex if and only if its epigraph*

$$epi f = \{\, (x, \mu) \in \mathbb{R}^{N+1} \mid x \in \mathbb{R}^N, \mu \geq f(x) \,\}$$

is a convex set.[5]

Property 2. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be twice continuously differentiable. Then f is convex if and only if the Hessian matrix*

$$H(x) = (h_{ij}(x)), \quad h_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_i, \dots, x_N)$$

is positive semi-definite for every $x \in \mathbb{R}^N$.

---

[5] Since this property is equivalent to the presented definition of convexity, it could be used as an alternative definition.

Property 3. *If* $f : \mathbb{R}^N \to \mathbb{R}$ *is a convex function and* $\phi : \mathbb{R} \to \mathbb{R}$ *is a non-decreasing convex function, then* $g(x) = \phi(f(x))$ *is convex on* $\mathbb{R}^N$.

Property 4. *If* $f_1, f_2 : \mathbb{R}^N \to \mathbb{R}$ *are convex functions, then so is* $f_1 + f_2$.

Property 5. *The pointwise supremum of an arbitrary collection of convex functions is convex.*

Property 6. *Let* $f : C \to \mathbb{R}$ *be a continuously differentiable function, defined on an open convex set* $C \subset \mathbb{R}^N$. *Then,* f *is convex on* C *if and only if for all* $x, y \in C$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

Most of the aforementioned assertions are straightforward to prove. Note that Property 1 provides a link between convex sets and convex functions. Note also that Property 3, 4 and 5 pose constructs to build convex functions by means of others and thus extend the class. Property 6 is very useful in practice. We will revert to it at a later stage.

Fenchel Duality

Definition 3. *The conjugate of a function* $f : D \subset \mathbb{R}^N \to \mathbb{R}$ *is the function* $f^\star : D \to \mathbb{R}$, *defined by*

$$f^\star(y) = \sup_{x \in D} \{ \, \langle y, x \rangle - f(x) \, \}.$$

Note that according to the above definition, some values of $f^\star$ may be infinite. We do not consider this case here for the sake of simplicity and assume $f^\star(y) < +\infty$. Again, we discuss some properties of the conjugate function.

Property 1. *The conjugate function is always convex.*

Property 2. *For a function* $f : \mathbb{R}^N \to \mathbb{R}$ *and any* $x, y \in \mathbb{R}^N$, *the following inequality holds*

$$\langle x, y \rangle \leq f(x) + f^\star(y).$$

Property 3. *For a function* $f : \mathbb{R}^N \to \mathbb{R}$, $f^{\star\star} = f$ *if and only if* f *is convex and semi-continuous.*

Property 4. *The conjugacy operation is order-reversing. For functions* $f, g : \mathbb{R}^N \to \mathbb{R}$, *the inequality* $f \leq g$ *implies* $f^\star \geq g^\star$.

Based on the above theory, we formulate the following

Theorem 3. *For given functions* $f : E \subset \mathbb{R}^N \to \mathbb{R}$ *and* $g : Y \subset \mathbb{R}^N \to \mathbb{R}$ *and a linear map* $A : E \to Y$ *let* $p, d \in \mathbb{R}$ *be primal and dual values defined, respectively, by the Fenchel problems*

$$\begin{array}{rcl} p & = & \inf_{x \in E} \{ \, f(x) + g(Ax) \, \} \\ d & = & \sup_{y \in Y} \{ \, -f^\star(A^\star y) - g^\star(-y) \, \}. \end{array}$$

These values satisfy the weak duality inequality $p \geq d$. *If, furthermore,* f *and* g *are convex and satisfy some technical conditions, then the values are equal, i. e.* $p = d$.

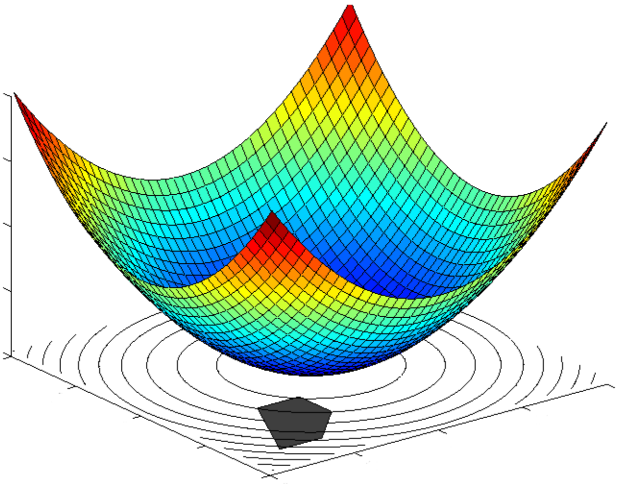


Fig. 1.6: Convex function over a convex domain. A constrained convex optimization problem involves minimizing a convex function over a convex domain. A globally optimal solution to such a problem can always be obtained.

This theorem is of great practical importance. It covers a broad range of cost functions, frequently encountered in computer vision. Basically, the statement allows to convert a convex minimization problem to an equivalent maximization one with the same extremal value. As a consequence, the gap between both values provides information about the accuracy of a given solution and could be used to formulate a stopping criterion in the optimization. In practice, such a convergence condition is more reliable than traditional ones regarding the evolution step or the function decay since it doesn't depend on discretization and thus overcomes potential numerical difficulties.

Constrained Convex Optimization

So far, we have considered general functions of the form $f : \mathbb{R}^N \to \mathbb{R}$. However, typical cost functions encountered in practice are usually accompanied by a set of constraints which restrict the domain of feasible arguments, i. e. $f : C \subset \mathbb{R}^N \to \mathbb{R}$. A minimization problem of the form

$$\min_{x \in C} f(x),$$

where $f : C \to \mathbb{R}$ is a convex function and $C \subset \mathbb{R}^N$ is a convex set, is called a *constrained convex optimization* problem (see Fig. 1.6). The globality properties previously discussed are also valid in this case. It is easy to see that the generalization of Theorem 2 to functions over a convex domain is straightforward. Yet, finding a globally optimal solution is more challenging in this case since the minimizer may lie on the domain boundary.

There two aspects of paramount importance when dealing with an optimization problem – *existence* and *uniqueness* of solutions. In the context of convex optimization, existence is warranted if the domain set C is compact and the underlying function f is lower semi-continuous, i. e. for each $x \in \mathbb{R}^N$

$$\lim_{y \to x} \inf f(y) \geq f(x).$$

According to the definition of convexity, uniqueness is provided if f is strictly convex.

In order to handle constrained minimization problems, we need the term *feasible direction*.

Definition 4. *A vector* $d \in \mathbb{R}^N$, $d \neq \vec{0}$, *is a feasible direction at* $x \in C$ *if there exists* $\alpha_0 > 0$ *such that* $x + \alpha d \in C$ *for all* $\alpha \in [0, \alpha_0]$.

Now, we can formulate a sufficient condition for a global solution to a constrained convex optimization problem

Theorem 4. *Let* $f : C \to \mathbb{R}$ *be a convex continuously differentiable function defined on the convex set* $C \subset \mathbb{R}^N$. *Suppose the point* $x^\star \in C$ *is such that for any feasible direction* d *at* $x^\star$ *we have*

$$d^T \nabla f(x^\star) \geq 0.$$

Then, $x^\star$ *is a global minimizer of* f *over* C.

Proof. Let $x \in C$, $x \neq x^\star$. By convexity of C

$$x^\star + \alpha(x - x^\star) = \alpha x + (1 - \alpha)x^\star \in C$$

for all $\alpha \in (0, 1)$. Hence, the vector $d = x - x^\star$ is a feasible direction at $x^\star$. By assumption, we have

$$\nabla f(x^\star)^T (x - x^\star) = d^T \nabla f(x^\star) \geq 0.$$

By using Property 6 of convex functions (see Section 1.3), we can conclude

$$f(x) \geq f(x^\star) + \nabla f(x^\star)^T (x - x^\star) \geq f(x^\star).$$

□

The above condition is quite general but not practical for verifying a potential minimizer due to the necessity for an exhaustive search.

In practice, a convex domain C is usually specified by a set of equality and inequality constraints. In such cases, we are confronted with the following minimization problem

$$\begin{array}{rrcl} & f(x) & \to & \min \\ \text{s. t.} & h_j(x) & = & 0, \quad j = 1, \ldots, m \\ & g_i(x) & \leq & 0, \quad i = 1, \ldots, p, \end{array}$$

where $h_j : \mathbb{R}^N \to \mathbb{R}$, $j = 1, \ldots, m$ and $g_i : \mathbb{R}^N \to \mathbb{R}$, $i = 1, \ldots, p$.[6] In this case, more specific conditions can be posed, known as Karush-Kuhn-Tucker (KKT) conditions.

Theorem 5. *Let* $f : \mathbb{R}^N \to \mathbb{R}$ *be a convex continuously differentiable function on the set of feasible points*

$$C = \{\, x \in \mathbb{R}^N \mid h_j = 0,\ g_i \leq 0,\ j = 1, \ldots, m,\ i = 1, \ldots, p \,\},$$

where $h_j : \mathbb{R}^N \to \mathbb{R}$, $j = 1, \ldots, m$ *and* $g_i : \mathbb{R}^N \to \mathbb{R}$, $i = 1, \ldots, p$ *are also continuously differentiable and* C *is convex. Suppose there exist* $x^\star \in \mathbb{R}^N$ *and* $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p \in \mathbb{R}$ *such that*

- $\mu_i \geq 0$, $i = 1, \ldots, p$
- $\nabla f(x^\star) + \sum_{j=1}^{m} \lambda_j \nabla h_j(x^\star) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^\star) = \vec{0}$

---

[6] Note that the righthand side of all equality and inequality constraints is zero. This is not restrictive since it can always be achieved by rearranging terms.

- $\sum_{i=1}^{p} \mu_i \, g_i(x^\star) = 0$.

Then, $x^\star$ is a global minimizer of f over C.

Proof. Suppose $x \in C$. By convexity of f, we have (see Property 6 in Section 1.3)

$$f(x) \geq f(x^\star) + \nabla f(x^\star)^T (x - x^\star).$$

We can rewrite this inequality by using the second condition in the theorem. We get

$$f(x) \geq f(x^\star) - \sum_{j=1}^{m} \lambda_j \nabla h_j(x^\star)^T (x - x^\star) - \sum_{i=1}^{p} \mu_i \nabla g_i(x^\star)^T (x - x^\star).$$

Now, we can observe that $\nabla h_j(x^\star)^T (x - x^\star) = 0$ for all $j = 1, \ldots, m$ since each h_j is constant on C by definition. Hence, the second term in the above expression vanishes. We now claim that $\sum_{i=1}^{p} \mu_i \nabla g_i(x^\star)^T (x - x^\star) \leq 0$. To see this, note that because C is convex, $(1-\alpha)x^\star + \alpha x \in C$ for all $\alpha \in (0, 1)$. Thus,

$$g_i(x^\star + \alpha(x - x^\star)) = g_i((1-\alpha)x^\star + \alpha x) \leq 0$$

for all $\alpha \in (0, 1)$ and $i = 1, \ldots, p$. Summing over i, premultiplying each term by $\mu_i \geq 0$, subtracting $\sum_{i=1}^{p} \mu_i \, g_i(x^\star) = 0$, and dividing by α, we get

$$\frac{\displaystyle\sum_{i=1}^{p} \mu_i \, g_i(x^\star + \alpha(x - x^\star)) - \sum_{i=1}^{p} \mu_i \, g_i(x^\star)}{\alpha} \leq 0.$$

If we take the limit $\alpha \to 0$, we obtain $\sum_{i=1}^{p} \mu_i \nabla g_i(x^\star)^T (x - x^\star) \leq 0$.
Finally, we can conclude

$$\begin{aligned} f(x) &\geq f(x^\star) - \sum_{j=1}^{m} \lambda_j \nabla h_j(x^\star)^T (x - x^\star) - \sum_{i=1}^{p} \mu_i \nabla g_i(x^\star)^T (x - x^\star) \\ &\geq f(x^\star) \end{aligned}$$

for all $x \in C$, and the proof is completed. □

Remark. The constants λ_j, $j = 1, \ldots, m$ and μ_i, $i = 1, \ldots, p$ are called KKT multipliers. In the special case $p = 0$, i. e. without inequality constraints, the KKT conditions boil down to the well-known Lagrange condition and the KKT multipliers are equivalent to the Lagrange multipliers.

In the following, we briefly present some basic numerical algorithms for solving constrained convex optimization problems of the form $f : C \to \mathbb{R}$, where f is a continuously differentiable convex function and

$$C = \left\{ x \in \mathbb{R}^N \;\middle|\; \begin{array}{c} h_j(x) = 0, \; j = 1, \ldots, m \\ g_i(x) \leq 0, \; i = 1, \ldots, p \end{array} \right\}$$

is a convex set. It should be noted that if the minimizer to be found is situated within the interior of the domain, the problem can be handled by any method for unconstrained optimization. Thus, appealing are only the cases when the minimizer lies on the boundary.

Gradient Descent

Gradient descent is one of the simplest and most popular approaches for optimization. It is based on the observation that the gradient of a continuously differentiable function points in the direction of the steepest ascent. Hence, an iterative procedure following the opposite direction leads to a minimizer. Yet, in the case of constrained optimization, we have to ensure that all iterations are feasible while pursuing function decay in each step. It turns out that in this case this can be achieved by simply backprojecting each intermediate estimate onto the domain set C. Thus, we obtain the following gradient descent procedure for constrained convex optimization: We start with some $x^{(0)} \in C$ and iterate for $k \geq 0$

$$\alpha_k = \arg\min_{\alpha > 0} f(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

$$x^{(k+1)} = \Pi_C(x^{(k)} - \alpha_k \nabla f(x^{(k)})),$$

where Π_C denotes projection onto the set C.[7] The main strength of the above approach is its simplicity which implies numerical stability and low memory requirements. Note that the above scheme is completely iterative and no auxiliary variables, that need to be stored, are involved. Its main weakness is, as in the case of unconstrained optimization, the slow speed of convergence. Gradient descent is a typical first-order method.

Quasi-Newton Methods

Quasi-Newton methods, also called variable metric methods, can be considered as a generalization of the gradient descent method. They follow the iterative scheme

$$x^{(k+1)} = x^{(k)} + \alpha_k d_k,$$

where $\alpha_k \in \mathbb{R}$ is a time-step whose value is chosen such that a reduction in $f(x)$ is obtained. The main feature of quasi-Newton methods is that the calculation of the evolution direction d_k depends on a matrix B_k that is forced to be positive-definite. When $x^{(k)}$ is close to the required solution, B_k is usually set to an approximation of the Hessian of f. Away from the solution, however, it is more beneficial to choose a matrix B_k that gives a good search direction. In fact, quasi-Newton methods differ in the determination of the matrices B_k. In the unconstrained case, the vectors d_k are defined as

$$d_k = -B_k^{-1} \nabla f(x^{(k)}).$$

The positive definiteness[8] of B_k implies that d_k always gives a descent direction for f due to

$$\nabla f(x^{(k)})^T d_k = -\nabla f(x^{(k)})^T B_k^{-1} \nabla f(x^{(k)}) < 0$$

unless $\nabla f(x^{(k)}) = 0$. One can observe that the value of d_k is the one that minimizes the quadratic function

$$Q(d) = f(x^{(k)}) + d^T \nabla f(x^{(k)}) + \frac{1}{2} d^T B_k d.$$

Yet, we should keep in mind that we consider a constrained optimization problem and the evolution direction d_k, defined above, may not be a feasible direction. In this case, d_k is

---

[7] Note that the projection operator Π_C yields a unique result for a convex set C.

[8] Recall that the inverse of a positive-definite matrix is also positive-definite.

computed as the vector d that minimizes the quadratic function $Q(d)$ subject to the linear constraints

$$\begin{aligned} h_j(x^{(k)}) + d^T \nabla h_j(x^{(k)}) = 0, &\quad j = 1, \ldots, m \\ g_i(x^{(k)}) + d^T \nabla g_i(x^{(k)}) \leq 0, &\quad i = 1, \ldots, p, \end{aligned}$$

which are approximations of the original constraints at the point $x = x^{(k)} + d$. Thus, finding a feasible evolution direction boils down to solving a convex quadratic programming problem. It turns out that, also in the constrained case, requiring the matrices B_k to be positive-definite is sufficient to guarantee correct convergence.

The main advantage of quasi-Newton methods is their superlinear rate of convergence which makes them substantially faster than the classical gradient descent approach. Unfortunately, they suffer from a notable limitation. They involve solving a constrained quadratic problem in each minimization step, which could be as difficult as solving the original convex problem. However, we will see that in certain cases there are possibilities to circumvent this computational burden. We refer to [107] and the references therein for more details on quasi-Newton methods for constrained optimization.

Penalty Methods

The key idea behind penalty methods is to replace the constrained optimization problem by a unconstrained one. The associated unconstrained problem is solved and the solution is used as an approximation to the minimizer of the original problem. More precisely, the modified minimization problem has the form

$$f(x) + \gamma P(x) \;\rightarrow\; \min,$$

where $\gamma \in \mathbb{R}$ is a positive constant, called penalty parameter, and $P : \mathbb{R}^N \rightarrow \mathbb{R}$ is called a penalty function. Formally, a penalty function can be defined as follows.

Definition 5. *A function $P : \mathbb{R}^N \rightarrow \mathbb{R}$ is called a penalty function for the above constrained optimization problem if it satisfies the following three conditions:*

- *P is continuous.*
- *$P(x) \geq 0$ for all $x \in \mathbb{R}^N$.*
- *$P(x) = 0$ if and only if x is feasible, i. e. $x \in C$.*

Clearly, for the above unconstrained problem to be a good approximation to the original problem, the penalty function P must be appropriately chosen. Its role is to "penalize" points outside the feasible set. Therefore, it should be defined in terms of the constraint functions $h_1, \ldots, h_m, g_1, \ldots, g_p$. One standard choice is given by the so-called *Courant-Beltrami* penalty function

$$P(x) = \sum_{j=1}^{m} (h_j(x))^2 + \sum_{i=1}^{p} (g_i^+(x))^2,$$

where

$$g_i^+(x) = \max(0, g_i(x)) = \begin{cases} 0 & \text{if } \; g_i(x) \leq 0 \\ g_i(x) & \text{if } \; g_i(x) > 0. \end{cases}$$

Note that the above function is differentiable. Hence, derivative-based minimization schemes can effortlessly be applied.

Recall that the derived unconstrained optimization problem provides only an approximate solution to the original constrained one. The precision of the approximation depends not only on the penalty function P, but also on the penalty parameter γ. One would expect that the larger the value of γ, the closer the approximated solution will be to the true solution as points that violate the constraints are penalized more heavily. Ideally, in the limit $\gamma \to \infty$, the penalty method should yield the exact solution to the constrained problem. Yet, for large values of γ some numerical difficulties occur since this leads to large gradients in the optimization. This dilemma, which is hard to resolve, exhibits the main drawback of penalty methods. On the positive side, penalty methods do not involve enforcing the associated constraints in each iteration step, which is the reason for their high efficiency. Yet, the rate of convergence is a delicate issue since the original problem is not solved directly but only in an approximative manner. We refer to [23] for more details on penalty methods.

Primal-Dual Methods

Primal-dual methods have been devised for a specific class of convex functions, commonly appearing in computer vision. While a variety of recent developments in this field exist [20], here we concentrate on a general primal-dual formulation proposed in [102].

Before introducing the specific optimization problem, we start with some definitions. Let $A : X \subset \mathbb{R}^N \to Y \subset \mathbb{R}^M$ be a continuous linear mapping and $C \subset X$, $K \subset Y$ closed convex sets. We consider the following saddle-point problem

$$\min_{x \in C} \max_{y \in K} \langle Ax, y\rangle + \langle g, x\rangle - \langle h, y\rangle,$$

where $g \in \mathbb{R}^N$ and $h \in \mathbb{R}^M$ are constant vectors. We assume that the problem has at least one solution $(\hat{x}, \hat{y}) \in C \times K$ and for any $(x, y) \in C \times K$, we have

$$\langle A^\star y + g, x - \hat{x}\rangle - \langle Ax - h, y - \hat{y}\rangle \geq 0,$$

where $A^\star$ denotes the adjoint of A. Then, the algorithm has the following form: We choose $(x^{(0)}, y^{(0)}) \in C \times K$ and let $\bar{x}^{(0)} = x^{(0)}$. We choose two time-steps $\tau, \sigma > 0$ and iterate for $k \geq 0$

$$\begin{aligned} y^{(k+1)} &= \Pi_K(y^{(k)} + \sigma(A\bar{x}^{(k)} - h)) \\ x^{(k+1)} &= \Pi_C(x^{(k)} - \tau(A^\star y^{(k+1)} + g)) \\ \bar{x}^{(k+1)} &= 2x^{(k+1)} - x^{(k)}. \end{aligned}$$

Note that the values of $\bar{x}^{(k)}$ might not necessarily be in C. We refer to [102] for more details on the choice of the time-step parameters τ and σ and a correctness proof. The variables $x^{(k)}$ are called primal variables, whereas the $y^{(k)}$'s are referred to as dual variables. Yet, the term “duality”, used here, should not be mixed up with the term “Fenchel duality” defined in Section 1.3. In essence, the above optimization scheme realizes gradient descent with respect to the primal variable and gradient ascent with respect to the dual variable. The extrapolation variable $\bar{x}^{(k)}$ is used for correctness purposes and doesn't have any impact on the speed of convergence. This implies that primal-dual methods are also first-order methods.

We will encounter constrained convex optimization problems throughout this thesis and will use and compare different numerical schemes to solve them. We will see that there is no general recipe for solving constrained convex optimization problems. The various methods exhibit different grade of suitability to the particular problems.

1.4 Continuous vs. Discrete Optimization

Continuous and discrete optimization are branches in applied mathematics that have appeared and emerged independently from each other. The continuous optimization approach expresses the space of candidate solutions as *real*-valued combinations of a (finite or infinite) number of basis solutions. In contrast, discrete optimization relies on a domain of *integer*-valued such combinations. Moreover, the ranges of these integers must be finite. It is important to note that continuous optimization can also be applied to discrete problems. In this context, the obtained result has to be subsequently converted to the desired quantized form.

Obviously, the term "convexity" makes sense only in a continuous setting. Thus, the presented convex optimization poses a subdomain of the field of continuous optimization. In this section, we extend our motivation for the exploration of convex formulations for 3D shape reconstruction by conducting a detailed comparison between a convex relaxation technique and an established combinatorial counterpart in terms of discretization accuracy, computational and memory requirements, and potential for parallel computing. It should be emphasized that the goal here is not to highlight continuous optimization in favor of discrete optimization, in general, but to show its better suitability to the problems handled in this thesis. The main results of this evaluation are published in [65] and [72].

Let us consider the minimization of the following functional[9]

$$E(S) = \lambda \int_{int(S)} f(x)\, dx \;+\; ||S||, \tag{1.5}$$

where $S : \Theta \subset \mathbb{R}^2 \to V$ is a certain surface estimate lying within a volume $V \subset \mathbb{R}^3$, $int(S) \subset V$ denotes the surface interior and $||S||$ – the surface area with respect to some norm $||.||$. Furthermore, $f : V \to \mathbb{R}$ signifies a provided volume map reflecting the desired subdivision and $\lambda \in \mathbb{R}$ is a weighting parameter. Conventionally, the first term in the above formulation is called data term and the second one – smoothness term. The functional in (1.5) is of a quite general form and covers a broad range of variational problems frequently encountered in shape optimization. In fact, all energy models proposed in this work can be regarded as special cases of (1.5) or involve solving subproblems that fit to this general form. Now, let's concentrate on the minimization of (1.5). There are two ubiquitous techniques for minimizing energy functionals of this form in a globally optimal manner – *graph cuts* and *convex relaxation.*

Before introducing these two methods in more detail, we observe that the optimization of (1.5) can be expressed as a binary labeling problem. In particular, we get the following equivalent functional

$$\begin{aligned} E(u) &= \lambda \int_V f(x)\, u(x)\, dx \;+\; \int_V ||\nabla u||\, dx \\ \text{s. t.} \quad & u : V \to \{0, 1\}. \end{aligned} \tag{1.6}$$

In the above formulation, the surface S is represented implicitly by means of a binary function u, i. e. $u = \mathbf{1}_{int(S)}$, where $\mathbf{1}_{int(S)}$ denotes the indicator function of the surface interior. In other words, u takes on the value 1 within the surface interior and 0 within the exterior region. Hence, the gradient ∇u vanishes almost everywhere in V, except for the surface boundary, where u jumps from 0 to 1 (or vice versa). Thus, the integral over the length of the gradient ∇u can be regarded as a measure for the surface area, which justifies the last term in (1.6). Note that u is not differentiable at the surface boundary. In this respect, the gradient ∇u

---

[9] Throughout this thesis, we will denote by $E(.)$ the functionals that we minimize so as to emphasize their physical interpretation as "energy". Moreover, the terms "cost functional" and "energy functional" will be used as synonyms.

should be interpreted in a distributional sense. Generally, the last term in (1.6), known in the literature as *total variation (TV)*, has some very nice and useful properties which are discussed in Appendix B in more detail. A function $u : V \to \mathbb{R}$ is said to be of bounded variation, and write $u \in BV(V)$, if its total variation is finite. Naturally, in practice we always assume $u \in BV(V)$ in order to make the respective minimization problems meaningful.
Implicit representations are more flexible and tractable than explicit counterparts since they don't rely on a particular parametrization, but incorporate directly geometric properties of the underlying shape. Moreover, they offer topological resilience and considerable numerical stability. Implicit representations have attracted formidable attention since the popularization of level set methods [29, 98]. The goal is now to minimize the functional in (1.6) on a discretized volume grid $\tilde{V}$ sampling V. Let

$$V = [v_{11}, v_{12}] \times [v_{21}, v_{22}] \times [v_{31}, v_{32}] \subset \mathbb{R}^3 \tag{1.7}$$

with boundary values $v_{lm} \in \mathbb{R}$.[10] Then, we can define

$$\tilde{V} = \left\{ \left(\begin{array}{c} v_{11} + i \cdot \dfrac{v_{12} - v_{11}}{N_1} \\ v_{21} + j \cdot \dfrac{v_{22} - v_{21}}{N_2} \\ v_{31} + k \cdot \dfrac{v_{32} - v_{31}}{N_3} \end{array} \right) \middle| \begin{array}{l} i = 0, \ldots, N_1 - 1 \\ j = 0, \ldots, N_2 - 1 \\ k = 0, \ldots, N_3 - 1 \end{array} \right\} \tag{1.8}$$

as a discretized version of V of resolution $N_1 \times N_2 \times N_3$. Obviously, we have the relation $\tilde{V} \subset V$. If $N = N_1 \cdot N_2 \cdot N_3$ denotes the overall number of voxels, we end up with a binary N-dimensional labeling problem.

Graph Cuts

The graph cut approach addresses the above labeling problem by constructing a directed graph in form of a regular lattice representing $\tilde{V}$ with nodes $\tilde{x}$. Thereby, each voxel corresponds to a node in the lattice. Neighboring nodes are connected so as to approximate the metric $||.||$ measuring the boundary size of the surface S. A crucial point in this construction process is that the degree of connectivity determines the accuracy of the metric approximation. Additionally, a source node s and a sink node t are introduced. They allow to include the unary term $f(\tilde{x})\, u(\tilde{x})$ for the voxel corresponding to $\tilde{x}$: If $f(\tilde{x}) \geq 0$, an edge to the source is created, weighted with $f(\tilde{x})$, otherwise – an edge to the sink weighted with $-f(\tilde{x})$. Thus, the binary labeling problem of minimizing (1.5) boils down to finding a minimal s/t-cut in the generated graph. An s/t-cut is a partitioning of the nodes into two sets S and T, where S contains the source s and T the sink t. Nodes $\tilde{x} \in S$ are assigned the label $u(\tilde{x}) = 0$, nodes $\tilde{x} \in T$ the label $u(\tilde{x}) = 1$. The weight of such a cut is the sum of the weights of all edges starting in S and ending in T. It is important to note that the minimal cut can be found in polynomial time with respect to the number of nodes in the graph, i. e. the volume resolution N for the particular problem at hand.
Graph cuts have gained remarkable attention in recent years due to their generality and ability to deliver globally optimal solutions (to discrete problems). Yet, a thorough introduction is out of the scope of the current work. We refer to [47, 75, 11, 76] for more details.

---

[10] For simplicity's sake, we assume that the volume is axis-aligned. However, this is not restrictive since it can always be achieved by applying an appropriate transformation in a pre-processing step.

Convex Relaxation

The convex relaxation approach is inspired by the key observation that the functional in (1.6) is convex, but it is defined over the non-convex domain of binary functions. This property is exploited by extending the feasible set to functions taking on also intermediate values between 0 and 1, i. e. $u : V \to [0,1]$, an operation called *relaxation*. This leads to a constrained convex optimization problem that can efficiently be solved in a globally optimal manner based on the theory in Section 1.3.

Proposition 1. *The minimization of*

$$\begin{aligned} &E(u) = \lambda \int_V f(x)\, u(x)\, dx \, + \, \int_V ||\nabla u||\, dx \\ &\text{s. t.} \quad u : V \to [0,1] \end{aligned}$$

poses a constrained convex optimization problem, i. e. we have a convex functional over a convex domain.

Proof. For arbitrary $u_1, u_2 : V \to [0,1]$ and $\alpha \in (0,1)$, we obtain

$$\begin{aligned} E(\alpha u_1 + (1-\alpha)u_2) \;&=\; \lambda \int_V f(x)\, \left(\alpha u_1(x) + (1-\alpha)u_2(x)\right)\, dx \\ &+\; \int_V ||\alpha \nabla u_1 + (1-\alpha)\nabla u_2||\, dx \\ &\leq\; \alpha \left(\lambda \int_V f(x)\, u_1(x)\, dx\right) + (1-\alpha)\left(\lambda \int_V f(x)\, u_2(x)\, dx\right) \\ &+\; \alpha \int_V ||\nabla u_1||\, dx + (1-\alpha) \int_V ||\nabla u_2||\, dx, \end{aligned}$$

where the last expression follows from the triangle inequality for the norm $||.||$. Now, we can conclude

$$E(\alpha u_1 + (1-\alpha)u_2) \leq \alpha E(u_1) + (1-\alpha)E(u_2).$$

The convexity of the domain $\{\, u \mid u : V \to [0,1] \,\}$ is due to the fact that it can be defined by means of the inequality constraints $u(x) \geq 0$ and $u(x) \leq 1 \;\; \forall\, x \in V$ (see Corollary 1). □

The above statement implies global optimizability of the relaxed problem. Yet, facing the original "binary" minimization problem, it is not obvious if we could benefit from this fact. Of course, the real-valued solution of the relaxed problem can easily be binarized via thresholding, but it remains unclear if this would lead to a good estimate. Surprisingly, the simple thresholding procedure turns out to be genuinely powerful in this context.

Theorem 6. *Let* $u^\star : V \to [0,1]$ *be a global minimizer of the functional in* (1.6) *over the relaxed domain* $\{\, u \mid u : V \to [0,1] \,\}$. *Then, for almost any threshold* $\mu \in (0,1)$ *the binary function* $\mathbf{1}_{\Sigma_\mu(u^\star)} : V \to \{0,1\}$ *with* $\Sigma_\mu(u) = \{\, x \in V \mid u(x) > \mu \,\}$ *is also a global minimizer.*

Proof. Using the layer cake representation of the function $u^\star : V \to [0,1]$ (see [21])

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

we obtain for the first term of $E(u^\star)$

$$\begin{aligned}
& \lambda \int_V f(x)\, u^\star(x)\, dx \\
= \;& \lambda \int_V f(x) \int_0^1 \mathbf{1}_{\Sigma_\mu(u^\star)}(x)\, d\mu\, dx \\
= \;& \lambda \int_0^1 \int_V f(x)\, \mathbf{1}_{\Sigma_\mu(u^\star)}(x)\, dx\, d\mu.
\end{aligned}$$

Furthermore, the coarea formula [121] can be used to express the TV-term of $E(u^\star)$ as an integral over the length of all level lines of $u^\star$, measured with respect to the norm $||.||$

$$\begin{aligned}
& \int_V ||\nabla u^\star||\, dx \\
= \;& \int_0^1 ||\partial \Sigma_\mu(u^\star)||\, d\mu \\
= \;& \int_0^1 \int_V ||\nabla \mathbf{1}_{\Sigma_\mu(u^\star)}||\, dx\, d\mu,
\end{aligned}$$

where $\partial \Sigma_\mu(u^\star)$ denotes the boundary of $\Sigma_\mu(u^\star)$. Finally, plugging all together yields

$$E(u^\star) = \int_0^1 E(\mathbf{1}_{\Sigma_\mu(u^\star)})\, d\mu.$$

Clearly, the functional is now merely an integral over the binary characteristic functions of the upper level sets of $u^\star$. Now, we can deduce

$$E(u^\star) = \int_0^1 E(\mathbf{1}_{\Sigma_\mu(u^\star)})\, d\mu \geq \min_{\mu \in (0,1)} E(\mathbf{1}_{\Sigma_\mu(u^\star)}).$$

Since $u^\star$ is assumed to be a global minimizer of E, this implies that the equality holds in the above expression and $\mathbf{1}_{\Sigma_{\mu_0}(u^\star)}$ is also a global minimizer, where

$$\mu_0 = \arg \min_{\mu \in (0,1)} E(\mathbf{1}_{\Sigma_\mu(u^\star)}).$$

We can conclude that up to a null set the function $E(\mathbf{1}_{\Sigma_\mu(u^\star)})$ is constant with respect to μ, i. e. $\mathbf{1}_{\Sigma_\mu(u^\star)}$ is a global minimizer of the functional in (1.6) for almost any $\mu \in (0,1)$. ☐

Remark. The theorem doesn't preclude the existence of a discrete set in $(0,1)$ for which the thresholding property is violated. Yet, if we assume that such a constant exists, this would imply that the function $\mathbf{1}_{\Sigma_\mu(u^\star)}$, regarded with respect to μ, jumps at this isolated point and takes on its previous value again. In turn, this would imply the existence of different binary functions $u_1, u_2 : V \to \{0,1\}$, i. e. $u_1 \neq u_2$, with the same energy – $E(u_1) = E(u_2)$. In fact, this is extremely unlikely to occur in practice.

The above theorem, referred to as *thresholding theorem* in the literature, is central since it allows for global optimizability of (1.6) via relaxation. In essence, it states that the minimization of (1.6) boils down to solving a constrained convex optimization problem. Note that any binary global minimizer of the relaxed problem is also a global minimizer of the original problem (1.6) because the domain of binary functions is a subset of the domain of relaxed functions. Finally, we end up with the following algorithm for globally optimizing (1.6) on the given discrete volume grid $\tilde{V}$

(1) Minimize $E(\tilde{u})$ on $\tilde{V}$ subject to $\{\ \tilde{u} \mid \tilde{u} : \tilde{V} \to [0,1]\ \}$.

(2) Threshold the result $\tilde{u}_{rel}$ with some $\mu \in (0,1)$ to obtain $\Sigma_\mu(\tilde{u}_{rel}) = \{\ \tilde{x} \in \tilde{V} \mid \tilde{u}_{rel}(\tilde{x}) > \mu\ \}$.

(3) Derive a minimum of (1.6) as $\tilde{u}_{bin} = \mathbf{1}_{\Sigma_\mu(\tilde{u}_{rel})}$.

Obviously, step (1), which involves solving a constrained convex minimization problem, is crucial. In order to avoid shifting the focus from the current comparative evaluation, we defer details on the numerical optimization to subsequent chapters. Even though the thresholding theorem is valid for (almost) any $\mu \in (0,1)$, it is advisable to choose a threshold within $[0.1, 0.9]$ in step (2) to avoid possible numerical imprecisions.

Based on the seminal works [55] and [121], convex relaxation approaches were developed by following two at first sight very different research directions – the continuous generalization of graph cuts [5] and the generalization of level set methods [15]. Recently, these techniques have attracted considerable attention and have been applied to a variety of computer vision problems like image denoising [21], image segmentation [15], multiview 3D reconstruction [73] and depth-map fusion [144].

Quantitative Comparison

A direct comparison of graph cut and convex relaxation methods introduced above reveals the fundamental difference between both methodologies. Graph cuts follow a discrete philosophy and quantize the cost functional at the beginning of the optimization process. Subsequently, the obtained discrete problem is solved exactly. Convex relaxation techniques administrate a continuous philosophy and postpone the discretization step as much as possible so as to achieve high degree of accuracy. In fact, the underlying constrained convex optimization problem could be solved in a continuous setting by using any of the approaches discussed in Section 1.3. These conceptual differences entail certain practical consequences. In this paragraph, we present a detailed quantitative comparison of the two techniques. In particular, we focus on metrication accuracy, runtime and memory requirements.

Metrication Accuracy and Consistency. Critical for the minimization of a cost functional of the form (1.6) is the realization of the smoothness term. In particular, important is the way the utilized optimization scheme approximates the underlying metric $||.||$. If the approximation is poor, some geometric artifacts known as *metrication errors* appear in the reconstruction. In contrast, an accurate approximation yields visually pleasant results and is genuinely effective in suppressing noise.

We compare the metrication accuracy of both minimization methods in a synthetic experiment shown in Figure 1.7. In order to promote quantitative evaluations, we rely on a scenario with a known analytic solution – a bounded catenoid defined by

$$S(u,v) = \left(2\cosh\left(\frac{v}{2}\right)\cos u, 2\cosh\left(\frac{v}{2}\right)\sin u, v\right)^T \tag{1.9}$$

with $(u,v) \in [0,2\pi] \times [-1,1]$. This minimal surface problem with given boundary constraints can be simulated in the variational framework (1.6) by using the regional map f to determine the base circles, i. e. the corresponding boundary slices in $\tilde{V}$, and defining $f(\tilde{x}) = 0$ for all other voxels. In order to enforce the boundary conditions, we set $\lambda = \infty$. For the metric $||.||$, we use the classical Euclidean metric, i. e. $||.|| \equiv |.|$. The result of the convex relaxation technique at a volume resolution of $180 \times 180 \times 60$ is depicted in Fig. 1.7 (c) and the graph

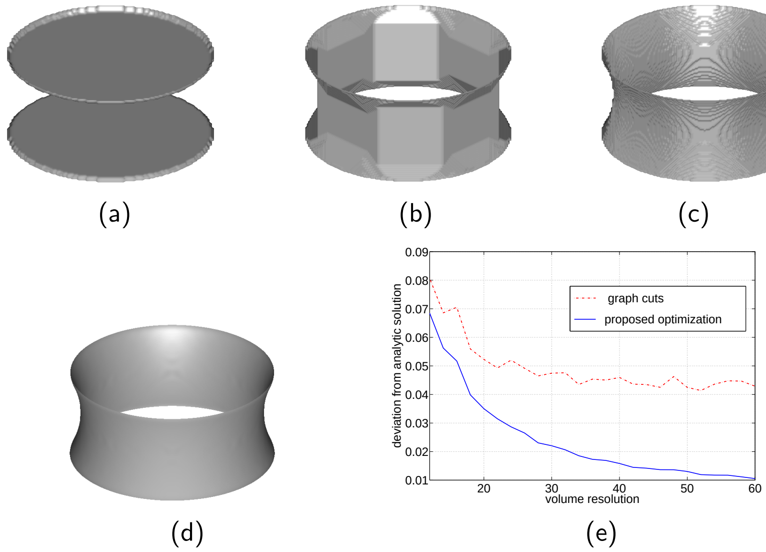
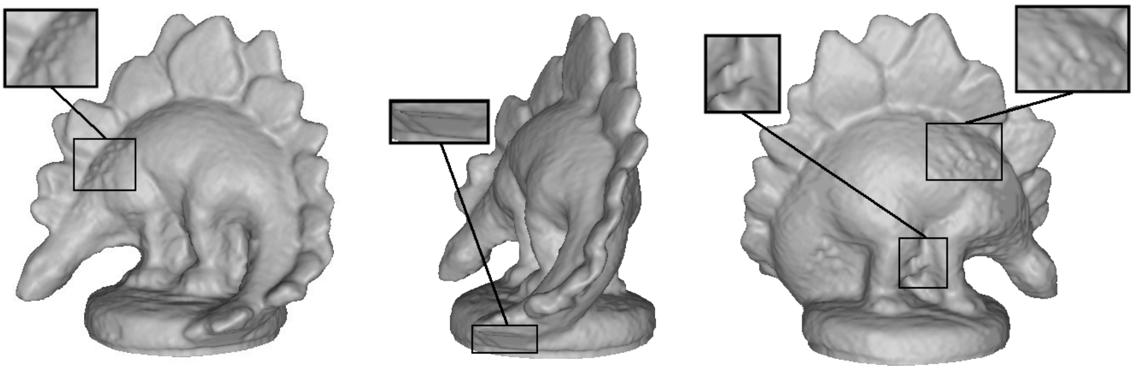


Fig. 1.7: Continuous vs. discrete minimal surfaces. (a) Graph cut reconstruction with a 6-neighborhood system at a volume resolution of $180 \times 180 \times 60$ (the highest in the plot). (b) Graph cut reconstruction with a 26-neighborhood system at the same volume resolution. (c) Surface produced by the convex relaxation technique. (d) Known analytic solution. (e) Deviation of the recovered surface from the analytic ground-truth for increasing volume resolutions. The experiment demonstrates that graph cut solutions can indeed be improved by reverting to larger neighborhood connectivity (26 instead of 6 neighbors). Yet, for any connectivity there is a metrication error which persists with increasing resolutions. The continuous convex relaxation method, on the other hand, is spatially consistent as the discretization error decays to zero.

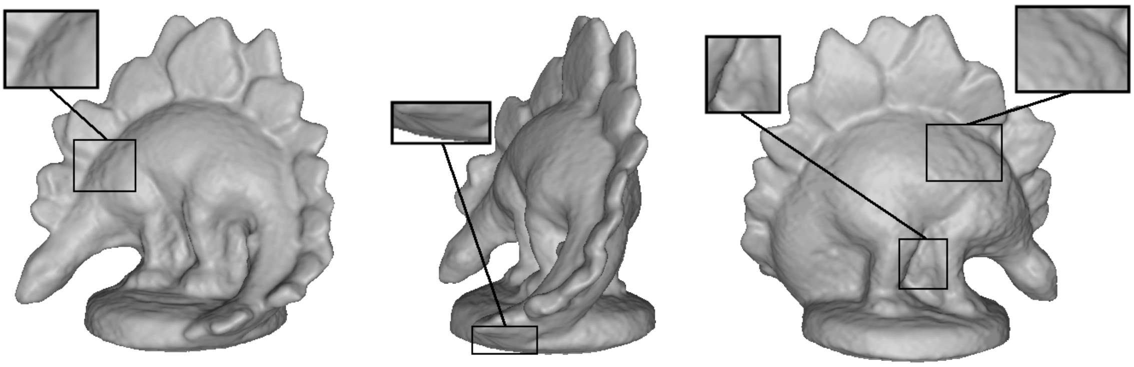
cut solutions are illustrated in Fig. 1.7 (a) for the 6-connectivity system and in Fig. 1.7 (b) for the 26-connectivity, respectively, and the same volume grid. The 6-connectivity involves joining neighboring graph nodes only along the three coordinate axes while 26-connectivity requires the inclusion of diagonal edges. Note that for rendering purposes as well as for further processing, we convert the obtained binary labelings to continuous surface estimates. As can be seen in the figure, the 6-neighborhood system completely fails to reconstruct the correct surface topology in contrast to the full 26-neighborhood. Indeed, it can be shown that increasing the graph connectivity leads to better approximations of the Euclidean metric [11]. Yet, even with the 26-neighborhood system discretization artifacts are clearly visible in terms of polyhedral blocky structures. Furthermore, we can say that for a fixed connectivity structure the computed graph cut solution is not spatially consistent with respect to the volume resolution in contrast to the solution of the convex relaxation method. This is demonstrated in Fig. 1.7 (e), where for both optimization approaches the deviation of the estimated surface from the analytic ground-truth is plotted for increasing spatial resolutions. The surface error is measured in terms of the Hausdorff metric

$$\epsilon = \int_{S_{true}} d(S_{true}(s), S_{num}) ds, \tag{1.10}$$

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



graph cut reconstruction



reconstruction via convex relaxation

Fig. 1.8: Comparison between graph cuts and convex relaxation on the "dinoRing" data set. *First row:* Graph cut reconstruction at volume resolution 256^3. *Second row:* Surface generated with the convex relaxation method at the same resolution. Both reconstructions are obtained by minimizing the same instance of (1.6) (see Chapter 3 for details). The convex relaxation result exhibits visual improvements in areas of noisy data due to the lack of texture or occlusions.

optimization	**connectivity**	**completeness**	**accuracy**
graph cuts	6	99.2 %	0.44 mm
relaxation	6	99.4 %	0.43 mm

Tab. 1.1: Quantitative evaluation of the reconstructions in Fig. 1.8.

structure.

The above observations are further confirmed on a practical test scenario. In particular, we provide an additional comparison on the "dinoRing" data set, that will frequently appear throughout the thesis, shown in Fig. 1.8. The data set is part of an established multiview stereo benchmark [112], allowing for quantitative evaluations, and will be presented in the course of the work in more detail. We ran both optimization techniques for the same instance of (1.6) at resolution 256^3 obtained with the multiview stereo approach described in Chapter 3. At this point, we focus on the optimization and the particular definition of f as well as the metric $||.||$ can be considered as a black box. Only a graph structure of 6-connectivity was used for the graph cuts due to memory restrictions. Note, however, that the convex relaxation method also relies on a 6-neighborhood system in its numerical implementation to impose surface smoothness.[11] At first glance, both reconstructions look similar. However, a closer look reveals that the convex relaxation approach is more successful in suppressing noise. This leads to visual improvements in areas of inaccurate data due to lacking texture or erroneous occlusion handling. A quantitative evaluation of both reconstructions is shown in Table 1.1. The numbers give accuracy (in mm) and completeness (in %). The completeness score measures the percentage of points in the provided ground truth model that are within $1.25mm$ of the reconstructed model. The accuracy metric shown is the distance d that brings 90% of the reconstructed surface within d from some point on the ground truth. As expected, the convex optimization manifests itself with some minor improvements. Recall that the basic difference between both techniques consists in the realization of surface regularization. Hence, it is not surprising that the overall quality of the reconstructions is determined by the accuracy of the volume map f and the particular choice for the metric $||.||$. The dilemma continuous vs. discrete minimization becomes more relevant when the data term is underweighted, i. e. for small values of λ. In case of noisy data for example, when surface smoothing becomes crucial, a continuous PDE-based approach provides more resilience and accuracy than a discrete counterpart.

Computational Time, Parallelizability and Convergence. In Table 1.2, we list the computational times for both optimization techniques for the two data sets. The graph cut method is evaluated for the 26-neighborhood system only on the “catenoid” data set and not on the “dinoRing” data set due to memory limitations. In our experiments, we used the publicly available implementation of [75]. It is evident that the convex relaxation approach manifests computational times similar to these of the graph cut optimization with 26-connectivity, but substantially higher than those obtained with a 6-neighborhood. The CPU runtimes were measured on a 2.66 GHz Intel QuadCore architecture.

Following recent progress in parallel computing and general purpose GPU programming, it is intriguing if the compared techniques bear the potential to make use of it. The parallelization of the convex relaxation method is straightforward and allows for an optimal exploitation of all computational capabilities since it involves iteratively solving a PDE on a discrete volume grid $\tilde{V}$. The parallelization of graph cut methods entails more difficulties. While a lot of efforts were made to parallelize graph cut algorithms [119, 86, 131], there is usually no theoretical guarantee that the overall runtime will be reduced for every problem instance [46]. Moreover, even if an acceleration is achieved, the ratio between a single- and a multi-thread implementation is far from optimal. Since there is no ultimate strategy for graph cut parallelizations, we demonstrate the parallelizability of convex relaxation methods only. Table 1.2 shows the respective GPU runtimes for the two data sets measured on a PC equipped with a NVIDIA Tesla C2070 graphics card. Note the tremendous savings in computational time. The notable deviation of the numbers for both data sets is due to the difference of the optimized energy models, which requires different numerical schemes.

Along with the measured computational time, it is important to determine the convergence properties of both optimization methods in a rigorous manner. Generally, graph cut algorithms have an exact termination criterion and a guaranteed polynomial running time. On the other hand, the convex relaxation approach is based on an iterative procedure whose termination is hard to predict. While the required number of iterations is typically size-independent (leading to a computation time which is linear in the number of voxels), one cannot speak of a guaranteed polynomial time complexity.

---

[11] In fact, the relaxed problem can be solved in different ways (see Section 1.3). Yet, in all cases, the underlying differential operators involve only partial derivatives along the canonical coordinate axes.

data set	**optimization**	**connectivity**	**runtime**
catenoid	graph cuts (CPU)	6	13 s
catenoid	graph cuts (CPU)	26	755 s
catenoid	relaxation (CPU)	6	576 s
catenoid	relaxation (GPU)	6	4 s
dino	graph cuts (CPU)	6	41 s
dino	relaxation (CPU)	6	577 s
dino	relaxation (GPU)	6	1 s

Tab. 1.2: Runtimes for the reconstructions in Fig. 1.7 and 1.8.

Memory Consumption. Finally, we compare the graph cut and the convex relaxation method with respect to their memory consumption. According to this criterion, the continuous approach is the clear winner: It requires only one floating point value for each voxel in $\tilde{V}$. In contrast, graph cut methods require an explicit storage of edges as well as one float value for each edge. This issue becomes especially important for high volume resolutions. For the two demonstrated experiments the measured difference in memory consumption was about a factor of 20 (for 6-connected graph cuts).

Conclusion. In summary, we briefly list the discussed strengths and weaknesses of both optimization techniques. Graph cuts are relatively fast and entail a polynomial time guarantee, but suffer from metrication errors, lack of optimal parallelizability and high memory requirements. Convex relaxation methods could be a little slower, but offer perfect parallelizability. Moreover, they always compute an accurate and spatially consistent solution. Yet, a small drawback is the absence of any rigorous convergence guarantees, even though this is not critical in practice. Generally, continuous convex relaxation methods seem to bear more potential in the long run for time- and memory-consuming applications involving shape optimization, which makes them particularly suitable for the problems tackled in this work.

1.5 Thesis Outline

The thesis is focused on applying the theory of convex optimization to the fields of image-based modeling and 3D segmentation. As previously mentioned, both problems are closely related since, in a volumetric formulation, both of them boil down to a voxel labeling problem. Yet, image-based modeling entails more difficulties regarding data term calculations which involve, for example, visibility estimation and matching of different image points or patches. For that reason, we pay more attention to multiview 3D reconstruction considered in Chapters 2, 3, 4 and 5. Chapter 6 is devoted to 3D segmentation.

- Chapter 2 presents an interactive approach for color-based multiview 3D reconstruction. Although it can be classified as a shape-from-silhouette method, the proposed formulation significantly differs from that of classical ones. Instead of segmenting each image separately in order to construct a 3D surface consistent with the extracted silhouettes, we compute the most probable 3D shape that gives rise to the observed color information. The probabilistic framework, based on Bayesian inference, enables robust 3D estimation by optimally taking into account the contribution of all views. Subsequently, the derived MAP estimation is cast as a convex energy minimization problem.

- In Chapter 3, we show that multiview stereo can be formulated in a convex manner. In particular, we study three energy models which are based on a common variational template unifying regional subdivision terms and on-surface photoconsistency. The three models use data measurements at increasing levels of sophistication. While the first two approaches are based on a classical silhouette-based volume subdivision, the third one relies on stereo information to define regional costs. Furthermore, the proposed propagating scheme is exploited to compute a precise photoconsistency measure as opposed to the classical estimation. The third model, which gives the most accurate results, is explored qualitatively and quantitatively in more detail.

- Chapter 4 addresses the problem of multiview stereo and silhouette integration. Thereby, the stereo information affects the underlying cost functional and silhouette consistency is imposed by means of convex constraints restricting the domain of feasible shapes. After relaxation, a constrained convex optimization problem is obtained whose solution is binarized via appropriate thresholding retaining the silhouette consistency of the result. Even though the original problem is not solved in a globally optimal manner, we show that the estimated surface lies within an energetic bound within the optimal one.

- Chapter 5 extends the class of convex energy models by including an anisotropic metric, which allows to integrate surface normal information, and by replacing the traditional linear variational formulation by a ratio model avoiding some of its shortcomings. The derived anisotropic generalization retains all globality guarantees of isotropic counterparts. The proposed ratio functional is minimized in a globally optimal manner by solving a sequence of convex optimization problems. Moreover, we discuss and demonstrate some important properties of minimal ratio models – absence of a shrinking bias and scale invariance.

- Chapter 6 is devoted to the 3D segmentation problem. We propose an energy model fusing various cues like regional intensity subdivision, edge alignment and orientation information. Although the presented approach is designed with focus on a particular application at hand – vesicle membrane reconstruction from fluorescence imaging – it is general enough to be applied to a variety of different segmentation tasks.

- Chapter 7 summarizes the main contributions of the thesis and outlines future research directions.

2

Interactive Color-Based Multiview Reconstruction

Not all who wander are lost.

John Tolkien (1892-1973)

In this chapter, we start the exposition of image-based modeling with one of the simplest and most popular modalities – shape from silhouettes.

2.1 Introduction

Motivation

The earliest approaches for multiview 3D reconstruction, dating back to the 1970's, use outlines to infer geometrical structure. While silhouette-based methods are not capable of retrieving surface concavities, since these do not affect the image projections, they come along with some important advantages and are often preferred in applications like robot navigation and tracking. Firstly, they enjoy significant stability and efficiency, which allows them to operate in challenging imaging conditions. Secondly, they seem to be the only reasonable alternative for recovering textureless or homogeneous objects. Thirdly, they usually do not require exact visibility estimation. This is a great advantage over multiview/photometric stereo and shading techniques, where visibility reasoning leads to a chicken-and-egg problem. Moreover, as we will see later in the thesis, silhouette-based methods can provide useful initial solutions that can be refined with other techniques.

Usually, silhouettes are used to infer surfaces in a two step process: an individual decision about pixel occupancy is made on a per-view basis, then geometrical structure is inferred from all estimated segmentations. Unfortunately, the automatic segmentation of individual images is in many cases not feasible, especially in the presence of noise, illumination variations and background clutter. A straightforward strategy to address this difficulty is to rely on interactive input from the user to guide the process by manually labeling image regions. While two scribbles marking foreground and background are usually sufficient for simple images, the extent of required user interaction increases significantly in case of cluttered or camouflaged environments. This problem becomes more relevant if we consider a collection of input images, where even a modest amount of user interaction on an individual image basis entails significant efforts. Applying image segmentation methods leads to a two-step silhouette fusion procedure, where binary image labelings are first computed separately and combined subsequently to build a unified 3D model. Yet, this simple scheme is suboptimal in the sense that the segmentation of each individual image does not take into account information from the remaining imagery. Thus, it is beneficial to design an interactive approach which exploits the fact that all input views capture the same scene in order to counteract possible inaccuracies in single observations (see Fig. 2.1).

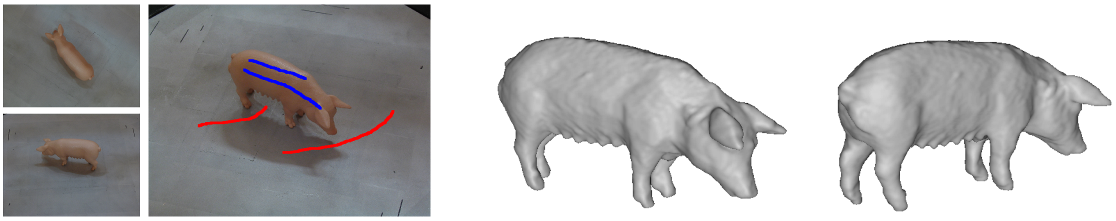


Fig. 2.1: Surprisingly accurate reconstructions can be obtained from a few scribbles marking foreground and background in only one of the input images of a sequence. Such an approach is in contrast to classical shape-from-silhouette methods, where all silhouettes are extracted explicitly in advance, which entails considerable interactive efforts.

Previous Work

Historically, the main strategy for computing a silhouette-consistent shape was to directly implement the intersection of visual cones corresponding to different silhouettes [8]. Such techniques aim at estimating the object's *visual hull*, i. e. the largest shape that yields the same silhouettes as the observed ones [79]. An important class within this domain exhibit volumetric approaches. The key idea is to discretize the space by a fixed voxel grid and label each voxel as opaque or transparent according to its projections onto the images. An early paper reporting a volumetric approximation of the visual hull is due to Martin and Aggarwal [91]. Subsequently, octree-based representations were employed by [106, 123] in order to increase the efficiency.

Probabilistic methods for multiview silhouette fusion were proposed in the context of model-free tracking [38, 48]. However, since they are based on background subtraction, they require special environmental conditions and are not directly applicable to the problem of joint color-based segmentation and reconstruction from real-world image sequences.

Along with shape from silhouette techniques, researchers advocated the use of theoretically more transparent energy minimization methods which compute directly a 3D shape consistent with all images [141, 117, 17, 18]. In [141] the sought-after 3D surface is modeled in a variational sense by minimizing the reprojection error between estimated object texture and observed colors. Thus, the following energy functional is being minimized

$$E(S) \;=\; \sum_{i=1}^{n} \int_{\pi_i(S)} |f - I_i(z)|^2 \, dz \;+\; \sum_{i=1}^{n} \int_{\Omega_i \setminus \pi_i(S)} |g - I_i(z)|^2 \, dz, \tag{2.1}$$

where $f, g \in \mathbb{R}^3$ denote the colors of foreground and background, respectively. The values of f and g can be estimated in advance or updated during the minimization of S. $\pi_i(S) \subset \Omega_i$ signifies the projection of S in image i. Note that rather than regularizing the boundary of individual segmentations, the variational formulation in (2.1) allows to directly impose regularity of the estimated 3D model. Yet, one of the difficulties with such energy minimization methods is that respective functionals are not convex. Therefore, the proposed gradient descent optimization is likely to get stuck in a local minimum, especially in case of a complex object topology. The aspect that typically prevents global optimizability of functionals like (2.1) is the fact that the observed projections cannot be inverted and therefore do not allow a direct inference of voxel occupancy. They merely allow statements about the collection of voxels along respective lines of sight. This difficulty can be circumvented by measuring costs in 3D space rather than on the image plane [117, 17, 18]. This leads to the following formulation

$$E(S) = \lambda \int_{int(S)} \rho_{obj}(x) \, dx \;+\; \int_{S} ds, \tag{2.2}$$

where $\lambda \in \mathbb{R}$ is a weighting parameter, $int(S) \subset V$ denotes the interior of S and $\rho_{obj} : V \to \{-1, 1\}$ is defined as

$$\rho_{obj}(x) = \begin{cases} -1, & \text{if} \quad \pi_i(x) \in Sil_i \qquad \forall i = 1, \dots, n \\ 1, & \text{otherwise.} \end{cases} \tag{2.3}$$

The energy model in (2.2) can be regarded as a regularized version of classical visual hull formulations, but it still requires binary view segmentations, which makes the approach susceptible to noise in individual observations. A more general algorithm, that operates directly on image color information, was proposed in [17, 18]. It involves minimizing a cost functional of the form

$$E(S) = \lambda \int_{int(S)} (\phi - P_{obj}(x)) \; dx \; + \; \int_S e^{-\beta \cdot Z(s)} \, ds, \tag{2.4}$$

where $\lambda, \beta \in \mathbb{R}$ and $\phi \in [0, 1]$ are parameters, and the function $Z : V \to \mathbb{R}$ is defined as

$$Z(x) = \max_{i \in \{1, \dots, n\}} \sum_{c \in \{r,g,b\}} |\nabla I_i^c(\pi_i(x))|^2. \tag{2.5}$$

Thereby, I^r, I^g and I^b denote the three color channels of image I. Moreover, the probability map P_{obj} assigns each point in space $x \in V$ a likelihood for lying in the interior of a silhouette-consistent shape. It is computed by estimating color models for object and background and averaging over all provided observations. Although the model in (2.4) offers significant improvements over classical silhouette fusion techniques, the applicability of its implementation proposed in [17, 18] is limited. Firstly, the method does not allow for efficient user interaction on a single image. It either runs without any user intervention, which is unreliable in many cases, or it requires interaction in all views in order to build separate background models. Secondly, the method is quite slow (computational times up to a couple of hours) and sequential in nature, which entails the lack of parallelization potential. In this respect, additional difficulties are caused by the employment of graph cut optimization, the parallelization of which is also not straightforward as we saw in Section 1.4.

User interaction is an established tool for segmenting individual real-world images. The pioneering work [13] addresses the foreground/background interactive segmentation in still images via max-flow/min-cut energy minimization. The energy balances between likelihood of pixels belonging to the foreground and the edge contrast imposing regularization. The user-provided scribbles collect statistical information on pixels and serve additionally as hard constraints. The GrabCut [9, 109] framework further simplifies the user interaction required. It allows for interactively adding scribbles to improve the initial segmentation. Full color statistics are used, modeled as mixtures of Gaussians, and these are updated as the segmentation progresses. Further developments led to the utilization of weighted geodesic distances to the information supplied by the user [35, 7]. Recent advances in convex optimization initiated the appearance of [128], where total variation minimization is adopted to interactive image segmentation. Yet, all of these methods are restricted to individual image segmentation. They are unable to adequately process collections of images capturing the same scene, where the interdependence between different observations is crucial. This drawback motivated the development of image cosegmentation [111] - a framework exploiting the overlapping in content of two or more images with the goal of improving the segmentation results. While the current work is inspired by a similar incentive, there is one important difference - in our case the images are calibrated, which is an additional source of information and allows for a more accurate modeling of the interconnection between different observations.

Furthermore, researchers adapted interactivity to video segmentation [135, 85]. In this context, additional improvements are obtained by imposing time coherency, under the assumption that the changes between successive frames are minor. However, the generalization of such approaches, to make them applicable to the problem of segmenting collections of still photographs imaging the same scene, remains an open challenge.

Contribution

In this chapter, we propose a probabilistic treatment of the shape-from-silhouette problem. Instead of processing the input images independently and subsequently fusing the resulting information, we compute the most probable surface that gives rise to the given observations. To this end, we adopt a volumetric approach, where we assign to each voxel probability costs for being inside or outside the imaged shape. Color distributions for foreground and background are estimated from user interactions in the form of a few scribbles in only *one* of the input images. We avoid explicit visibility reasoning by initially neglecting the interdependence of voxels and reintroducing it in a probabilistic manner at a later stage of the modeling. The consequence of this approximation is that the resulting Bayesian inference problem can be optimized globally. In particular, we employ convex relaxation techniques to find the exact solution. In numerous experiments, we demonstrate that the proposed probabilistic formulation provides far more robust reconstructions than the classical silhouette fusion method [8, 79]. Furthermore, we show dramatic improvements of individual image segmentations by exploring multiview coherency criteria. It is important to notice that the proposed framework is general and can be used in combination with any probabilistic model for image inference. The main results in this chapter are published in [66, 67].

2.2 Probabilistic Volume Intersection

3D Shape Modeling via Bayesian Inference

We consider the problem of probabilistic voxel labeling from a series of calibrated images of a scene. The relationship between image observations and surface estimation is established in terms of Bayesian inference, which allows to derive a MAP estimate for the sought-after 3D surface by modeling the process of image formation. The probabilistic framework covers a wide range of noise sources like camera sensor perturbance, surface reflections, erroneous camera calibration etc. All these effects have as a result that observed colors deviate from the expected ones. In the following, the proposed probabilistic formulation is explained in more detail.

Here, we revert to the notations in Chapter 1. In particular, $V \subset \mathbb{R}^3$ denotes a continuous volume encompassing the object of interest and $\tilde{V} \subset V$ a discretized version defined in (1.8). Given the set of views, we are looking for the most probable surface $\hat{S}$ that gives rise to the provided image observations, that is

$$\hat{S} = \arg\max_{S \in \Lambda} P(S \mid \{I_1, \dots, I_n\}), \tag{2.6}$$

where $\Lambda = \{\, S \mid S : \Theta \subset \mathbb{R}^2 \to V \,\}$ is the set of all closed surfaces lying inside the volume V. By means of the Bayes formula, we obtain

$$P(S \mid \{I_1, \dots, I_n\}) \propto P(\{I_1, \dots, I_n\} \mid S) \cdot P(S), \tag{2.7}$$

where the a priori probability $P(S)$ allows to introduce preference to a certain class of surfaces possessing desired properties like smoothness, simple topology etc. It should be noted that the constant term

$$\frac{1}{P(\{I_1,\dots,I_n\})}$$

has been omitted in the above expression since it does not influence the shape retrieval process. A crucial issue in this formulation is the modeling of the likelihood $P(\{I_1,\dots,I_n\} \mid S)$. It reflects the image formation process in terms of the probability for observing images $I_1,\cdots,I_n$, provided a surface estimate S. To this end, we could rely on the simple and straightforward assumption that observations of separate voxels are independent from each other and only their projections onto the images influence their state. This leads to factorization over the entire volume

$$P(\{I_1,\dots,I_n\} \mid S) \approx \left[\prod_{\tilde{x}\in\tilde{V}} P(\ \{I_i(\pi_i(\tilde{x}))\}_{i=1,\dots,n} \mid S)\right]^{d\tilde{x}}, \tag{2.8}$$

where the exponent $d\tilde{x}$ denotes the discretization step and plays the role of a normalizer. It is introduced to ensure the correct continuum limit and make the expression invariant to refinement of the grid. In practice, the probability values are usually smaller than 1. Hence, the above product will tend to zero for increasing volume resolutions as the number of multipliers will grow. For example, if we double the number of voxels, the product will be generally raised to the power of 2. The effect of this modification will be neutralized by the exponent $d\tilde{x}$ which will be halved.

In fact, the independence assumption is not fulfilled since the appearance of a voxel can be affected by other voxels in the line of sight. However, we neglect this interdependence at this point and reintroduce it at a later stage of the modeling process.[1]

According to a certain surface estimate S, the voxels can be divided into two classes: lying inside an object or belonging to the background. Hence, the volume V can be expressed as $V = int(S) \cup ext(S)$, where $int(S)$ denotes the surface interior and $ext(S)$ the exterior region, respectively. Analogously, we obtain for the discrete counterpart $\tilde{V} = \tilde{int}(S) \cup \tilde{ext}(S)$, where $\tilde{int}(S)$ and $\tilde{ext}(S)$ are discretized versions of $int(S)$ and $ext(S)$. Considering this partitioning, we can proceed with

$$\begin{aligned} P(\{I_1,\dots,I_n\} \mid S) \quad &\approx \\ \left[\prod_{\tilde{x}\in\tilde{int}(S)} P(\ \{I_i(\pi_i(\tilde{x}))\}_{i=1,\dots,n} \mid \tilde{x}\in\tilde{int}(S))\right]^{d\tilde{x}} &\cdot \\ \left[\prod_{\tilde{x}\in\tilde{ext}(S)} P(\ \{I_i(\pi_i(\tilde{x}))\}_{i=1,\dots,n} \mid \tilde{x}\in\tilde{ext}(S))\right]^{d\tilde{x}} &. \end{aligned} \tag{2.9}$$

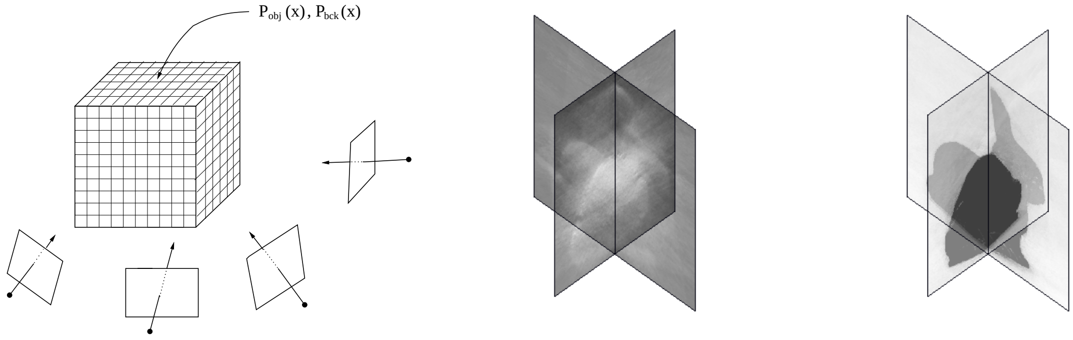
To simplify the notation, we denote

$$\begin{aligned} P_{obj}(\tilde{x}) &= P(\ \{I_i(\pi_i(\tilde{x}))\}_{i=1,\dots,n} \mid \tilde{x}\in\tilde{int}(S)) \\ P_{bck}(\tilde{x}) &= P(\ \{I_i(\pi_i(\tilde{x}))\}_{i=1,\dots,n} \mid \tilde{x}\in\tilde{ext}(S)) \end{aligned} \tag{2.10}$$

for $\tilde{x} \in \tilde{V}$ (see Fig. 2.2). Now, plugging the results in (2.7) and (2.9) into (2.6) gives the

---

[1] The weaker assumption of a factorization not over all voxels but merely over all lines of sight gives rise to a cost functional with integrals over all image domains as suggested in [141]. While this approximation is more faithful, it does not lead to a globally optimizable cost functional and does not entail uniqueness of solutions.



probabilistic volume intersection

foreground probability map P_{obj}

background probability map P_{bck}

Fig. 2.2: Probabilistic volume intersection. *Left:* Two probabilities P_{obj}, P_{bck} are assigned to each voxel explaining its projections onto the images with respect to the provided color models for foreground and background, respectively. *Right:* Slices through the probability maps P_{obj} and P_{bck} for the "bunny" sequence (see Fig. 2.6).

following expression

$$\hat{S} = \arg\max_{S\in\Lambda} \left(\prod_{\tilde{x}\in\widetilde{int}(S)} P_{obj}(\tilde{x}) \right)^{d\tilde{x}} \cdot \left(\prod_{\tilde{x}\in\widetilde{ext}(S)} P_{bck}(\tilde{x}) \right)^{d\tilde{x}} \cdot P(S). \tag{2.11}$$

Note that $P_{obj}(\tilde{x})$ and $P_{bck}(\tilde{x})$ defined in (2.10) do not represent the probability that $\tilde{x}$ is part of object or background, but rather the probability for observing certain colors in respective projections given that $\tilde{x}$ is part of the object or the background. In particular, this implies that for an arbitrary $\tilde{x} \in \tilde{V}$ these probabilities will generally not sum to 1. This is an important point in the modeling process since it allows to use two different color distribution models for foreground and background instead of a single one.

Joint Probabilities

Now, we are confronted with the question of how to compute the joint probabilities given in (2.10). Such a computation involves fusing hypotheses stemming from different views. A straightforward way to accomplish this task is to assume again independence of the image observations. Taking visibility into account, we note that the probability of a voxel being part of the foreground is equal to the probability that all cameras observe this voxel as foreground, whereas the probability of background membership describes the probability of at least one camera seeing background. This formulation can be regarded as the probabilistic analog to classical silhouette carving techniques, where a voxel is set transparent if it projects on background in at least one of the input images. Note that this is a conceptual difference to explicit visibility estimation, where the current surface determines the state of each voxel [141]. Following this train of thoughts, we obtain the formulation

$$\begin{aligned} P_{obj}(\tilde{x}) &= \prod_{i=1}^{n} P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \widetilde{int}(S)) \\ P_{bck}(\tilde{x}) &= 1 - \prod_{i=1}^{n} \left[1 - P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \widetilde{ext}(S)) \right]. \end{aligned} \tag{2.12}$$

The asymmetry in both expressions is due to the fact that they describe different types of events. The expression for $P_{obj}(\tilde{x})$ relies on the assumption that the observed object is completely visible in all of the images, i. e. no obstacles block the field of view of the cameras to it. The overall foreground score can then be obtained by simple multiplication of all image votes. The term $P_{bck}(\tilde{x})$ requires more care regarding the fact that a background voxel could be occluded in some of the images by the object itself. Hence, a simple multiplication of the single probabilities will not work. Instead, we revert the foreground evidence of the individual image responses with respect to the background model. In this sense, the interdependence of voxels neglected in (2.8) is now reintroduced.

A closer look at equation (2.12) reveals that it contains a bias with respect to the number n of images. Since the individual observation probabilities $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S))$ and $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S))$ are both bounded by 1 and typically smaller than 1 for realistic scenarios, $P_{obj}(\tilde{x})$ (and $P_{bck}(\tilde{x})$) would tend to zero (or one) for $n \to \infty$. This bias disappears if we consider each camera separately to approximate $P_{obj}(\tilde{x})$ and $P_{bck}(\tilde{x})$

$$\begin{aligned} P_{obj}(\tilde{x}) &\approx P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S)) \quad \forall\, i \\ 1 - P_{bck}(\tilde{x}) &\approx 1 - P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S)) \quad \forall\, i, \end{aligned} \tag{2.13}$$

and subsequently compute the geometric mean as an average score, yielding

$$\begin{aligned} P_{obj}(\tilde{x}) &= \sqrt[n]{\prod_{i=1}^{n} P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S))} \\ P_{bck}(\tilde{x}) &= 1 - \sqrt[n]{\prod_{i=1}^{n} \big[1 - P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S))\big]}. \end{aligned} \tag{2.14}$$

A direct comparison to (2.12) shows that the proposed model results in a normalizing root being introduced, which makes both expressions invariant to the number of cameras. The use of geometric mean is motivated by the nature of the fusion process. For example, if one camera supplies a weak evidence for foreground membership (i. e. $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S)) \approx 0$), this will immediately decrease the overall product and therewith the final value for P_{obj}. Analogously, a strong background response (i. e. $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S)) \approx 1$) of one of the cameras will drastically bring the value of P_{bck} closer to 1. This coincides with the classical visual hull computation, where a voxel is classified as background if at least one of its projections is inside a background region.

The probability for observing a certain color value in a given image can be modeled by a parametric distribution such as multivariate Gaussian

$$\begin{aligned} P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S)) &\sim \mathcal{N}(\mu_{obj}, \Sigma_{obj}) \\ P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S)) &\sim \mathcal{N}(\mu_{bck}, \Sigma_{bck}) \end{aligned} \tag{2.15}$$

in sRGB color space. Here, μ_{obj}, μ_{bck} denote the mean vectors and Σ_{obj}, Σ_{bck} the covariance matrices of both regions. As previously mentioned, the parameters of the color distributions are determined interactively by requiring the user to mark object and background regions via scribbles in one of the input images (see Section 2.4). Note that $\mathcal{N}(\mu_{obj}, \Sigma_{obj})$ and $\mathcal{N}(\mu_{bck}, \Sigma_{bck})$ stand for continuous density functions. In order to derive corresponding probability values, a normalization over the entire discretized color space has to be performed. This step is important as it guarantees that the values for $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{int}(S))$ and $P(I_i(\pi_i(\tilde{x})) \mid \tilde{x} \in \tilde{ext}(S))$ are within the unit interval $[0, 1]$ and validates the formulation in (2.14). Example probability maps P_{obj} and P_{bck} are depicted in Fig. 2.2. Note that the probability for foreground evidence is quite blurry while that of the background region is more distinct. This is due to the nature of the silhouette fusion scheme (see (2.14)). As

the foreground probability map P_{obj} is estimated by simply averaging the single observation probabilities, the obtained values are diluted. In contrast, one high observation probability with respect to the background model would immediately result in a high value for P_{bck}.
It should be noticed that the proposed probabilistic framework is quite general and does not rely on any inherent assumptions about particular modeling of the observation probabilities like (2.15).

2.3 MAP Estimation via Energy Minimization

Variational Formulation

Now, we come to the question of how the MAP estimation problem in (2.11) can be solved. It can be converted to an equivalent energy minimization problem and can be solved exactly by means of established convex relaxation techniques.
A standard approach to achieve that is to apply the negative logarithm, which converts the maximization problem in (2.11) to a minimization one. In a continuous setting, this yields

$$\begin{aligned} E(S) &= -\int_{int(S)} \log P_{obj}(x)\,dx \\ &\quad -\int_{ext(S)} \log P_{bck}(x)\,dx - \log P(S) \\ \hat{S} &= \arg\min_{S\in\Lambda} E(S). \end{aligned} \tag{2.16}$$

Minimizing the above energy functional is equivalent to maximizing the total a posteriori probability of all voxel assignments. In the spirit of energy minimization, the first two terms can be interpreted as external costs and measure the discrepancy between image observations and projections predicted by the model. The last term exhibits internal energy costs and summarizes prior knowledge on the surface geometry. In order to handle image perturbances like sensor noise, imprecise camera calibration and background clutter, this term is usually used to impose spatial smoothness of the recovered surface. From a theoretical point of view, a regularization term is often needed to guarantee uniqueness of solutions [95]. This can be achieved by setting

$$P(S) = e^{-\nu|S|}, \tag{2.17}$$

where ν is a weighting constant and $|S|$ denotes the Euclidean surface area. The Euclidean metric could be replaced by a more general Riemannian metric so as to impose image edge alignment for example [61, 17]. Yet, edge responses are provided separately by individual observations and could degrade the reconstructions, especially in case of noisy or cluttered image data. For that reason, we relied on the simple Euclidean metric in our regularization criterion. It should be mentioned that the a priori model in (2.17) introduces a minimal surface bias even though it achieves a high degree of smoothness. Alternatively, higher-order shape characteristics like curvature could be used instead, but this would make the optimization much more challenging. To the best of our knowledge, up to date there is no approach allowing global minimization of curvature in 3D. The choice of the minimal surface model in (2.17) is motivated by its simplicity and global optimizability, as well as its high efficiency in suppressing noise. By plugging (2.17) into the functional in (2.16), we finally obtain

$$\begin{aligned} E(S) &= -\int_{int(S)} \log P_{obj}(x)\,dx \\ &\quad -\int_{ext(S)} \log P_{bck}(x)\,dx + \nu|S|. \end{aligned} \tag{2.18}$$

Our goal is to minimize this functional.

Numerical Optimization

At the core of the minimization of the functional in (2.18) is the observation that it can be recast equivalently in a convex form.
The first step is to represent the surface S implicitly by the characteristic function $u : V \to \{0, 1\}$ of $int(S)$, i. e. $u = \mathbf{1}_{int(S)}$ and $1 - u = \mathbf{1}_{ext(S)}$. In this way, we obtain the following constrained non-convex energy minimization problem corresponding to (2.18)

$$\begin{gathered} E(u) = \int_V \log \frac{P_{bck}(x)}{P_{obj}(x)}\, u(x)\, dx \;+\; \nu \int_V |\nabla u|\, dx \to \min, \\ \text{s. t. } u \in \{0, 1\}\,. \end{gathered} \tag{2.19}$$

It is easy to see that the minimization problem in (2.19) is equivalent to

$$\begin{gathered} E(u) = \frac{1}{\nu} \int_V \log \frac{P_{bck}(x)}{P_{obj}(x)}\, u(x)\, dx \;+\; \int_V |\nabla u|\, dx \to \min, \\ \text{s. t. } u \in \{0, 1\}\,. \end{gathered} \tag{2.20}$$

Now, it is essential to observe that the problem in (2.20) is a special case of the one in (1.6). Hence, the convex relaxation technique proposed in Section 1.4 can be leveraged. In particular, we obtain the following

Theorem 7. *Let* $u^\star : V \to [0, 1]$ *be a global minimizer of the functional in* (2.20) *over the relaxed domain* $\{\, u \mid u : V \to [0, 1] \,\}$. *Then, for almost any threshold* $\mu \in (0, 1)$, *the binary function* $\mathbf{1}_{\Sigma_\mu(u^\star)} : V \to \{0, 1\}$ *with* $\Sigma_\mu(u) = \{\, x \in V \mid u(x) > \mu \,\}$ *is also a global minimizer.*

Proof. The above statement can be inferred directly from Theorem 6 and the fact that the functional in (2.20) is a special case of the one in (1.6). □

Thus, solving the minimization problem in (2.20) boils down to minimizing the respective functional within the domain of relaxed functions $u : V \to [0, 1]$, which poses a constrained convex optimization problem, and subsequently thresholding the result. In the sequel, we focus on numerical considerations. To this end, we adapt a primal-dual approach.
One can notice that the energy functional in (2.20) can be written in the form

$$E(u) = \frac{1}{\nu} \int_V f u\, dx \;+\; \int_V |\nabla u|\, dx, \tag{2.21}$$

where $f : V \to \mathbb{R}$ summarizes the constant part not dependent on u, i. e.

$$f = \log \frac{P_{bck}(x)}{P_{obj}(x)}.$$

We proceed by switching to a dual formulation of the total variation regularizer by means of an auxiliary variable $\xi : V \to \mathbb{R}^3$, which allows for the following conversion

$$E(u) = \frac{1}{\nu} \int_V f u\, dx \;+\; \left(\sup_{|\xi| \le 1} \int_V \langle \xi, \nabla u \rangle\, dx \right). \tag{2.22}$$

Now, we obtain a new functional

$$E(u, \xi) = \frac{1}{\nu} \int_V f u\, dx \;+\; \int_V \langle \xi, \nabla u \rangle\, dx \tag{2.23}$$

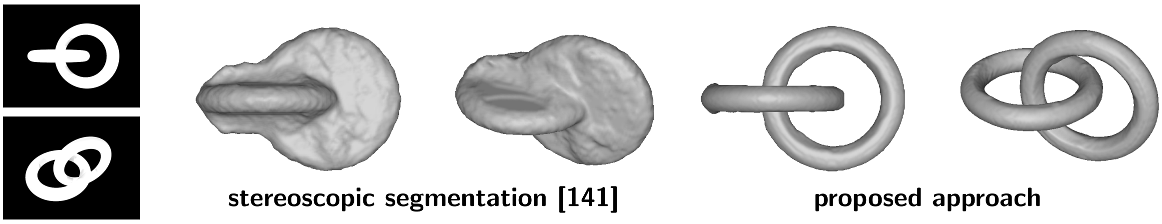


Fig. 2.3: Tori sequence. *Left:* 2 out of 20 synthetic input images of resolution 640 × 480. *Middle:* Multiple views of the reconstruction with stereoscopic segmentation [141]. *Right:* Corresponding views of the reconstruction obtained by the proposed approach. Both methods were initialized with a sphere enclosing the objects. While stereoscopic segmentation gets stuck in a local minimum and completely fails to capture the correct topology, the presented probabilistic fusion scheme accurately recovers the imaged geometry.

that should be minimized with respect to u and maximized with respect to ξ under the constraints $u \in [0,1]$ and $|\xi| \leq 1$. This states a typical saddle point problem that can be solved by a projected gradient descent/ascent strategy. Denoting by $C_{rel} = \{u \mid u : V \to [0,1]\}$ the set of relaxed labeling functions and by $K = \{\, \xi : V \to \mathbb{R}^3 \mid |\xi(x)| \leq 1 \,\}$ a vectorfield mapping within the unit ball, the primal-dual optimization scheme can be described as follows. We choose $(u^0, \xi^0) \in C_{rel} \times K$ and let $\bar{u}^0 = u^0$. We choose two time-steps $\tau, \sigma > 0$. Then, we iterate for $k \geq 0$

$$\begin{aligned} \xi^{(k+1)} &= \Pi_K(\xi^{(k)} + \sigma\, \nabla \bar{u}^{(k)}) \\ u^{(k+1)} &= \Pi_{C_{rel}}(u^{(k)} + \tau(\,\mathrm{div}(\xi^{(k+1)}) - \tfrac{1}{\nu} f)) \\ \bar{u}^{(k+1)} &= 2u^{(k+1)} - u^{(k)}, \end{aligned} \tag{2.24}$$

where Π_K and $\Pi_{C_{rel}}$ denote projections onto the corresponding sets. Both projections can easily be realized by simple normalization and clipping, respectively. In our implementation, the ∇-operator was discretized by means of forward differences on $\tilde{V}$ and the div-operator - with backward differences so as to ensure correct integration by parts.

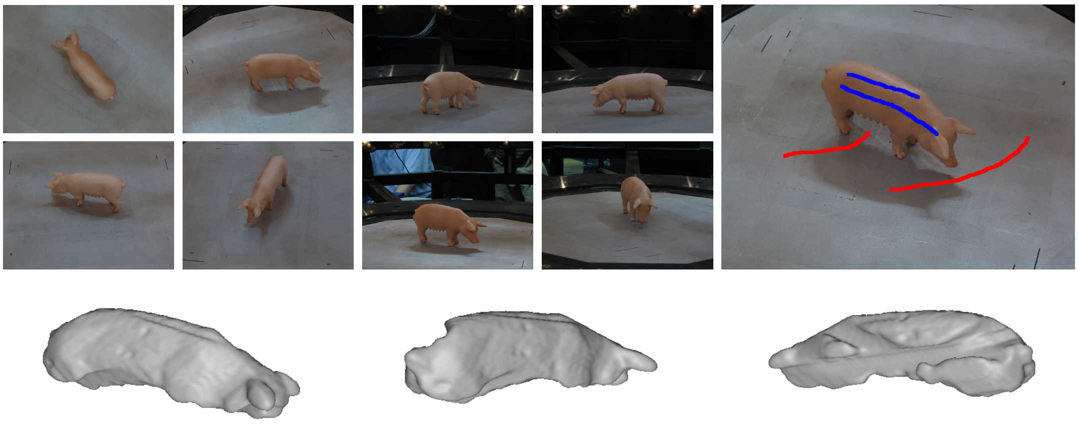
For sufficiently small time-step parameters convergence of the above iterative procedure can be proven. In our experiments, we observed stable behavior for $\tau = \sigma = 0.1$. Moreover, the parameter ν balancing the weighting between data fidelity term and smoothness was fixed to 1.8 throughout all our experiments.

2.4 Experiments

We demonstrate the viability of the proposed approach on multiple challenging synthetic and real-world image sequences. In particular, we show that the suggested probabilistic fusion scheme can handle objects of arbitrary topology independent from initialization and offers significant robustness to shading effects, camera sensor noise and background clutter, as frequently encountered in real scenarios.

Insensitivity to Object Topology

To validate the importance of global optimization, we start with a synthetic image sequence of two coupled tori (see Fig. 2.3). Although the data set is not interesting from a photometric point of view due to the contrasting appearance of objects and background, it is intriguing from a geometric point of view due to the complex topology of the objects. We compare the proposed approach to stereoscopic segmentation [141] which is an alternative local multiview fusion scheme. As can be expected, the local optimization procedure in [141], involving surface



independent silhouette fusion

proposed probabilistic fusion

Fig. 2.4: Sow sequence. *First row:* 9 out of 27 input images of resolution 1024×768. The utilized user input is highlighted, whereas blue scribbles mark foreground and red - background. *Second row:* Multiple views of the visual hull obtained with the classical independent silhouette fusion technique. *Third row:* Corresponding views of the reconstruction result produced by the proposed approach. The numerous shading effects like shadows and light reflections on the object's surface as well as the bad color calibration of the cameras lead to relatively poor independent segmentations of individual images (see Fig. 2.5). This, in turn, results in overcarving of the subsequently computed visual hull. In contrast, the proposed probabilistic fusion method produces a very accurate 3D model under these challenging conditions.

evolution at current contour generators only, is highly sensitive to initialization. As stated in [141], the method requires the initial surface to intersect each of the holes of the final one in order to converge to an accurate result. However, finding such an initialization is not a trivial task since it implies knowledge of the imaged objects. It is not surprising that stereoscopic segmentation completely fails to recover the correct topology starting from a sphere enclosing the two tori. In contrast, the proposed approach, which does not depend on initialization and always guarantees convergence to a global optimum for the provided user input, quite accurately captures the imaged geometry.

Robustness to Shading Effects and Camera Sensor Noise

The next two experiments, illustrated in Fig. 2.4 and 2.6, show the effect of shading effects like shadows and illumination highlights on the 3D reconstruction process.
The first image sequence, depicted in Fig. 2.4, captures a sow figurine. The data set is relatively challenging even though it does not create such an impression at first glance. While the figurine is rosy and well distinguishable from the surrounding gray environment, the numerous shading effects like shadows and light reflections adulterate the color and significantly diminish this discrepance. Furthermore, the images exhibit relatively bad color calibration as they were acquired by different camera devices. Such effects usually cause misclassification of

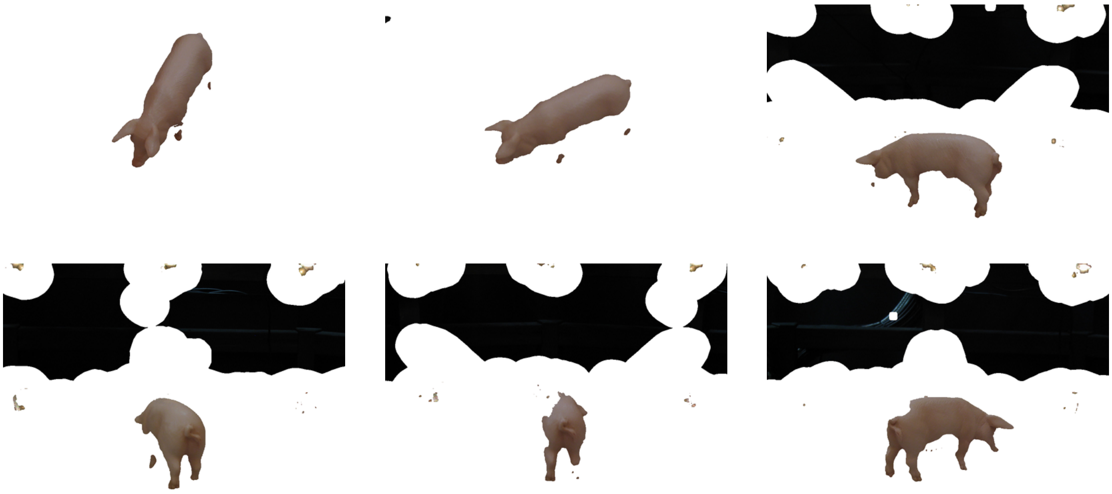


Fig. 2.5: Independent image segmentations for the sow sequence (6 out of 27) and the user interaction in Fig. 2.4. False negatives are mainly caused by shading effects like shadows and light reflections, whereas false positives are due to variations in the background color. Expectedly, the poor segmentation results produce a poor 3D model (see Fig. 2.4). Note however that while false positives do not lead to reconstruction inaccuracies in most cases, false negatives have a direct influence due to overcarving along the respective viewing rays.

respective foreground pixels when performing individual image segmentation (see Fig. 2.5), which, in turn, leads to overcarving of the subsequently computed visual hull. The proposed probabilistic fusion scheme, which avoids premature hard labeling decisions by exploiting the entire amount of available image information, is designed as a remedy to similar frequently appearing difficulties. We emphasize the benefits of the utilized outline coherency criteria by showing a direct comparison to the classical two-step silhouette integration method [8, 79] (see Fig. 2.4). In particular, we employed the approach in [128] to perform individual image segmentations. It should be mentioned that in addition to regional color cues this method relies on image edge information to increase the precision of the segmentations. In both cases, we used the same user input in *one* of the images, displayed in Fig. 2.4, to build the underlying color models. Moreover, in both cases foreground/background distributions were modeled by multivariate Gaussians. Note that even though individual user interaction per view helps to overcome color calibration problems, it doesn't give any substantial improvements in case of shading effects and considerably increases the interactive efforts required. Expectedly, the independent silhouette fusion technique produces a rather poor reconstruction. This is confirmed by individual image segmentations (see Fig. 2.5). In contrast, the proposed probabilistic fusion method produces a quite accurate 3D model under these challenging conditions. Even some of the small-scale surface details are recognizable.

Similar conclusions can be drawn from the experiment depicted in Fig. 2.6. The image sequence displays a red ceramic bunny figurine. This time, illumination variations cause less problems due to the diffuse reflectance properties of the material. For that reason, the independent silhouette fusion approach already gives a satisfactory result (see Fig. 2.6). Most of the small inaccuracies are due to unclean locations on the figurine. We use this data set to investigate the behavior of both methods (the proposed probabilistic fusion and the independent silhouette fusion) in case of camera sensor noise. To this end, noise within certain range was added randomly to the image color data. We plotted the deviation of the computed 3D model from a given ground truth at increasing noise range (see Fig. 2.7) measured in units of

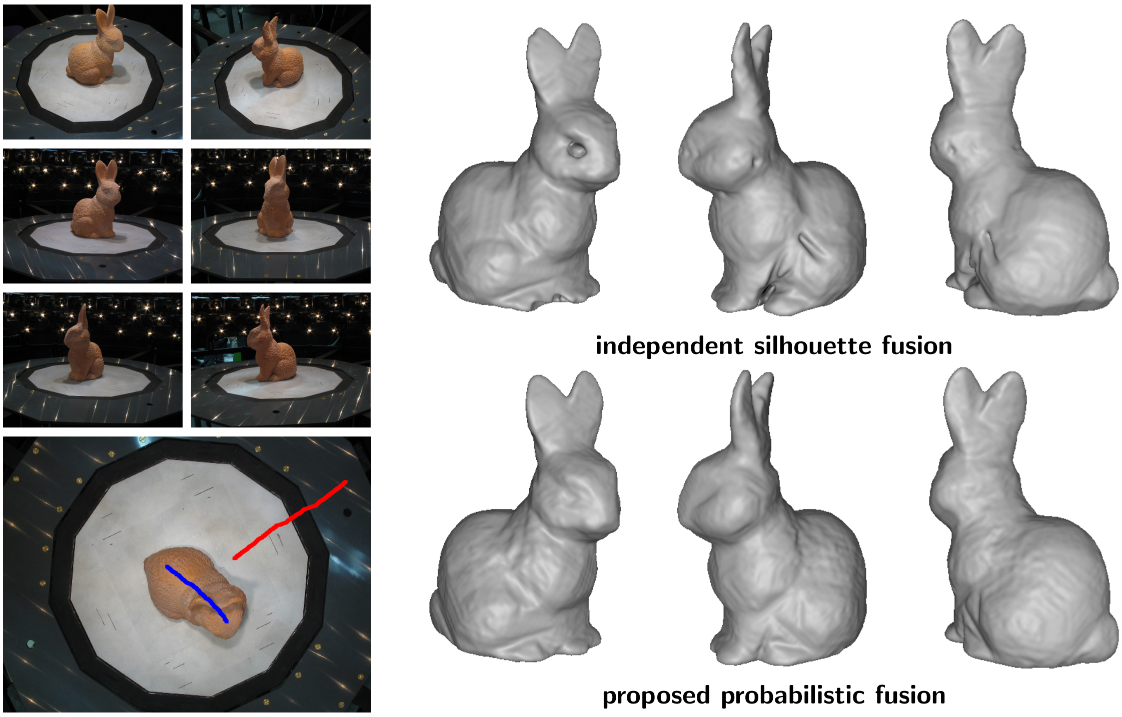


Fig. 2.6: Bunny sequence. *Left:* 7 out of 36 input images of resolution 640×480. The utilized user input is highlighted, whereas blue scribbles mark foreground and red - background. *Right, first row:* Multiple views of the visual hull obtained with the classical independent silhouette fusion technique. *Right, second row:* Corresponding views of the reconstruction result produced by the proposed approach. Although the traditional silhouette fusion approach gives a relatively accurate reconstruction in this case, some small imprecisions are still notable in particular caused by unclean locations on the figurine. The concurrent probabilistic fusion scheme produces an impeccable 3D model.

sRGB color space (color values are within $[0, 255]$). As a ground truth we used a visual hull of the object computed from manually obtained segmentations. Note that the visual hull is only an approximation of the physical object. Yet, it serves as a ground truth in this case since it exhibits the case of perfect data. If $u_{gt} : V \to \{0, 1\}$ denotes an implicit labeling representing this ground truth surface (being 1 within the interior region and 0 within the exterior) and $u : V \to \{0, 1\}$ the obtained 3D labeling, we measure the misalignment between them as

$$\epsilon = \frac{\displaystyle\int_V |u_{gt}(x) - u(x)|\, dx}{\displaystyle\int_V u_{gt}(x)\, dx + \int_V u(x)\, dx}. \tag{2.25}$$

In particular, we have $\epsilon \in [0, 1]$ with $\epsilon = 0$ if and only if both reconstructions are identical and $\epsilon = 1$ if u is the empty set. Two important observations can be made when analyzing the graphs in Fig. 2.7. First, it is evident that the noise levels, at which both compared approaches start to degrade, are quite different. While the independent silhouette fusion method shows a notable deviation at noise range of 20 color space units, the accuracy of the probabilistic one is unaffected up to noise range of 50 units. The superior resilience to camera sensor noise of the proposed probabilistic formulation is additionally emphasized by its generally smooth behavior for ascending noise levels, which is in contrast to the jumpy performance of its opponent.

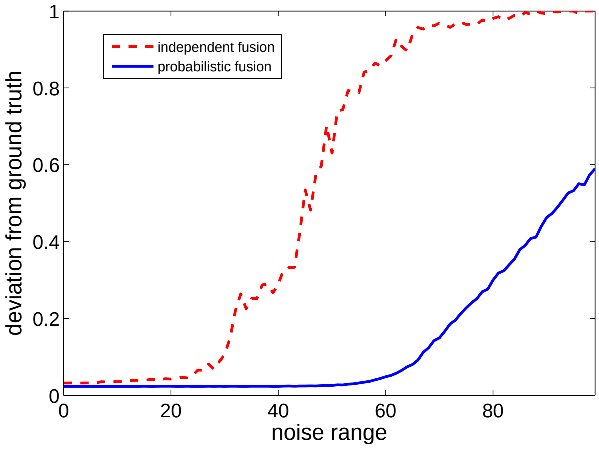


Fig. 2.7: Robustness to camera sensor noise. The accuracy of the proposed probabilistic approach and the traditional independent silhouette fusion procedure for the image sequence in Fig. 2.6 is investigated for ascending levels of image noise. The noise is added randomly and measured in terms of its application range in units of sRGB color space. The precision of the reconstruction is computed as the deviation from a provided ground truth surface. See text for more details.

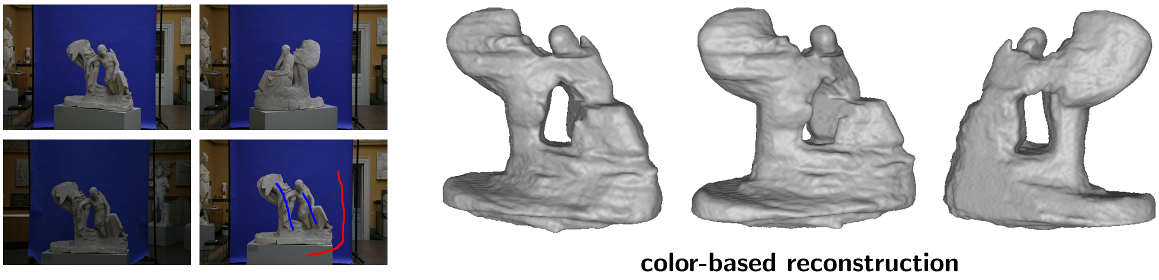
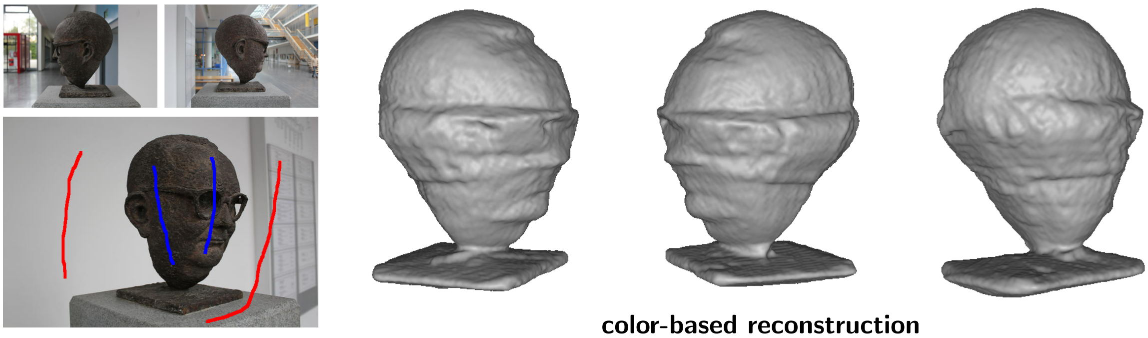


Fig. 2.8: Statue sequence. *Left:* 4 out of 36 images of resolution 1536×1024. The utilized user interaction is superimposed in the last image (blue scribbles mark foreground and red - background). *Right:* Multiple views of the estimated color-based reconstruction. Note the color similarity between the object and the pedestal as well as the severe intensity variations.

Robustness to Background Clutter

While the image sequences considered so far capture a more or less homogeneous background, the next two data sets take a further step and increase the degree of difficulty by picturing typical real-world backgrounds spanning a wide range of colors.

The first sequence, depicted in Fig. 2.8, illustrates a statue imaged in front of a blue poster in the Academic Art Museum in Bonn, Germany. Although the poster helps to separate the captured statue from the others in the background, the object is not completely separable in color space due to its similarity to the pedestal. Since the goal is the precise reconstruction of the statue, the pedestal underneath was marked as background by the provided user interaction (see Fig. 2.8). Expectedly, this diminishes the discriminative power of both color distributions. An additional challenge pose the severe intensity variations which are due to the fact that the photographs were taken at different times of the day. Despite all of these difficulties, the proposed approach produces a relatively accurate reconstruction result, even though it exhibits some small imprecisions (e. g. at the basement). In fact, the estimated 3D model turns out to be precise enough to initialize a stereo-refinement process and obtain a highly accurate 3D model (see Fig. 3.15).



color-based reconstruction

Fig. 2.9: Bust sequence. *Left:* 3 out of 36 input images of resolution 1296×864 and superimposed user interaction (blue scribbles mark foreground and red - background). *Right:* Multiple views of the 3D reconstruction obtained with the proposed approach. Note the wide range of background colors as well as the complex reflectance properties of the material.

The second image sequence, depicted in Fig. 2.9, displays a bronze bust sculpture of Robert Sauer. As can be seen from the example pictures, the background continually changes including the surrounding building interior and hence a very wide range of colors. This significantly exacerbates the separability of the sculpture in individual images, even though most of the objects in the background are relatively far apart from it. Additional difficulties cause the complex reflectance properties of the material. Once again, the proposed approach produces a quite accurate result under these challenging conditions. Even though the reconstruction exhibits some small-scale artifacts (e. g. at the basement) and some oversmoothing effects (e. g. the spectacle frame), the shape of the bust is clearly recognizable.
The accuracy of the computed 3D models is confirmed by the image segmentations obtained by projecting them onto the input views (see Fig. 2.10). In case of background clutter, this leads to dramatic improvements over the naive isolated segmentation approach and clearly demonstrates the potential of the proposed probabilistic silhouette coherency criteria. This observation is additionally emphasized by a quantitative evaluation over the entire image sequences, shown in Fig. 2.11. To this end, ground truth segmentations were obtained by labeling the images manually. The segmentation error was computed as

$$err = \frac{p_{false}}{p_{true} + p_{false}}, \tag{2.26}$$

where p_{true} and p_{false} denote the number of correctly classified and misclassified pixels in all views, respectively. Note that $err \in [0, 1]$. The independent segmentation method demonstrates poor performance for all data sets except for the "bunny" sequence due to shading effects, illumination variations and background clutter. In contrast, the proposed probabilistic fusion approach shows clear superiority and gives accuracy improvements ranging from factor 3 (for the "bunny" sequence) to factor 46 (for the "statue" sequence). Note that while the segmentation error is negligible for the "sow" and "bunny" sequences, acquired in lab conditions, it increases for the "statue" and "bust" sequences, generated in more complex environments, but to an acceptable extent. These results provide an explicit justification for the exploration of various coherency criteria in the context of multiview segmentation.

User Interaction

The provided user interaction is visualized for all real experiments in Fig. 2.4, 2.6, 2.8 and 2.9. It is evident that the proposed approach gets by with only a few roughly specified scribbles

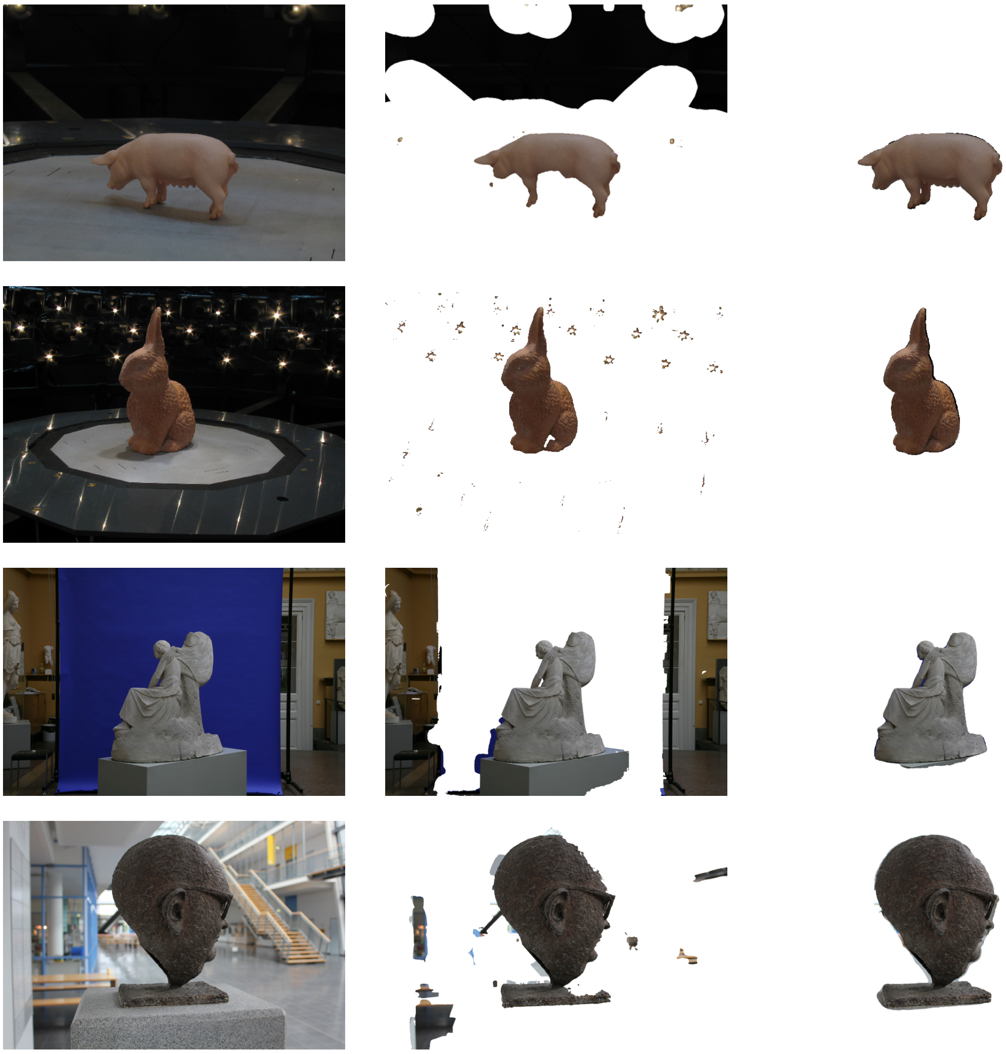


Fig. 2.10: Segmentation of individual images of the sequences in Fig. 2.4, 2.6, 2.8 and 2.9. *First column:* One of the input images. *Second column:* Interactive segmentation with the method in [128]. *Third column:* Interactive segmentation with the proposed approach obtained by projecting the computed 3D model onto the image. Even though the estimated silhouette-coherent segmentations are not pixel precise due to the use of 3D regularization and the discrepance between image resolution and volumetric resolution, the silhouettes are registered accurately and offer dramatic improvements over independent 2D segmentations.

in *one* of the input images. This suggests that the method is not very sensitive to the user intervention, which was confirmed in our experiments.

As previously mentioned, we relied on single Gaussians in our modeling as all of our test objects are single-colored. Multivariate Gaussians minimize the interaction efforts while achieving a substantial degree of robustness to model deviations. We also experimented with Gaussian mixture models. However, we observed that the results gradually degrade for more than two mixture modes due to over-fitting effects. Note that the user-specified scribbles occupy only a small portion compared to the entire amount of pixel data. Yet, Gaussian mixture models or kernel density estimation can still be preferable in case of multi-colored objects.

It should be emphasized that all demonstrated data sets could successfully be handled e. g. by the independent silhouette fusion scheme or the method in [17] with the appropriate amount of user interaction on a per-view basis. In contrast, the proposed approach stands out by its

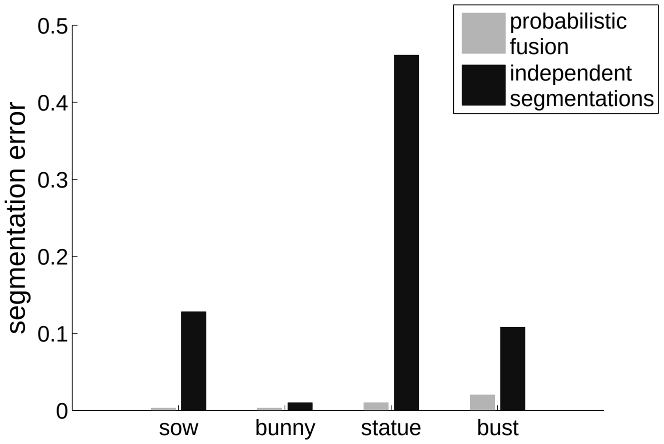


Fig. 2.11: Accuracy of individual image segmentations. The proposed probabilistic approach is compared to the naive independent segmentation method. The probabilistic fusion scheme has been evaluated by projecting the final 3D model onto the image planes. Ground truth segmentations have been obtained by labeling the images manually. Note the tremendous improvement in segmentation accuracy achieved by exploiting probabilistic silhouette coherency criteria.

	# images	image resolution	runtime GPU
tori	20	640 × 480	3.54 s
sow	27	1024 × 768	2.08 s
bunny	36	640 × 480	4.41 s
statue	36	1536 × 1024	3.95 s
bust	36	1296 × 864	4.26 s

Tab. 2.1: Data sets and GPU runtimes for all demonstrated experiments.

capability to produce an accurate reconstruction from only a few scribbles in *one* of the input images. This property reveals its high practical value, especially in case of long sequences containing multiple hundreds or thousands of photographs.

Computational Time

As previously mentioned, the proposed approach has been designed with focus on not only robustness but also computational efficiency. In particular, we make use of recent progress in parallel computing with a GPU implementation of the method. Note that its ingredients enable parallelization over the volume grid since all involved computations are at a voxel basis. Moreover, it can be observed that the overall computational time scales linearly with both the number of input images and the volume resolution. Runtimes for all demonstrated experiments, measured on a NVIDIA Tesla C2070, can be found in Table 2.1. In our GPU implementation, we used exclusively global memory to store all input images and volumetric data. Even though we tried to employ shared memory in the optimization step, exploiting the neighboring structure of the underlying PDEs (2.24), this didn't lead to a notable runtime reduction. Note that the computational time of the presented method does not depend on the image resolution (ignoring the time for loading the images) but only on the number of views. In all test cases, volumetric resolution is in the range between 8 and 21 million voxels.

It should be recalled that the input of the proposed approach consists not only in the image sequence and the provided user interaction but also a specification of a bounding box containing the object of interest. Although a tight specification is not necessary for the method to work, it influences the precision of the computed 3D model and hence the computational time (a loose bounding box requires a high volume resolution). One way to obtain a bounding box estimate is to use the 3D point cloud, produced by classical structure-from-motion techniques, which are needed to calibrate the input views.

2.5 Discussion

We presented a novel energy minimization approach for interactive joint silhouette extraction and 3D reconstruction from a number of calibrated 2D camera views. The energy model is derived from a probabilistic setting via Bayesian inference and is optimized globally using convex relaxation. The probabilistic formulation avoids making hard decisions about silhouette occupancy based on single views and allows to optimally take into account color information from all input images. In addition, it provides a novel decoupling scheme to account for the interdependence between voxels, which gives rise to a Bayesian inference problem and allows to compute the globally optimal reconstruction. We experimentally demonstrated that the proposed method compares favorably to state-of-the-art silhouette-based reconstruction methods in that it is more robust to noise, background clutter, shading effects and camera sensor perturbations. Moreover, it does not require initialization and therefore easily handles 3D shapes of complex topology. Making use of a GPU implementation, robust interactive reconstructions were computed with runtimes of up to 4.41 seconds.

3

Multiview Stereo as a Convex Problem

If everybody is thinking alike, then somebody isn't thinking.

George Patton (1885-1945)

In this chapter, we focus on one of the most prominent modalities for image-based modeling, which also plays a central role in the human perceptual system – multiview stereo.

3.1 Introduction

Motivation

In the previous chapter, we considered the reconstruction of a 3D shape from its image outlines. While the silhouette cue gives rise to genuinely robust 3D modeling approaches, it suffers from limited accuracy due to its inability to capture surface indentations. Multiview stereo overcomes this difficulty by relying on the process of matching corresponding points in different views (see Fig. 3.1). Although there are no inherent assumptions about the 3D geometry involved, there is a primary assumption about the reflectance properties of the observed object justifying the matching procedure – the so called Lambertian assumption. It states that the appearance of a point on the surface should be independent from the viewing direction and thus consistent throughout the entire image sequence. Even though many real-world objects possess Lambertian reflectance properties, there are also cases, where the respective assumption is severely violated, e. g. metallic objects, shiny objects, translucent objects etc. Besides, the reliability of the matching process strongly depends on the degree of textureness of the imaged object. For these reasons, the development of robust regularization schemes is of paramount importance for the design of practical multiview stereo approaches. Fortunately, as we show in this chapter, the multiview stereo problem can be cast in a convex setting.

Previous Work

The earliest dense multiview stereo algorithms use carving techniques to obtain a volumetric representation of the scene by repeatedly eroding inconsistent voxels [113, 78]. Thus, the reconstruction problem is posed as a volume labeling problem, where each voxel is labeled as opaque or transparent depending on its photoconsistency score. Thereby, a greedy strategy is applied to infer an occupancy decision for all voxels separately. Yet, the space carving framework manifests some weaknesses. In particular, it introduces a bias towards maximal photoconsistent shapes and does not enforce smoothness, which often results in rather noisy reconstructions. Moreover, the greedy nature of the methodology could produce erroneous overcarving or undercarving effects. Even though subsequent generalizations [16, 87] man-

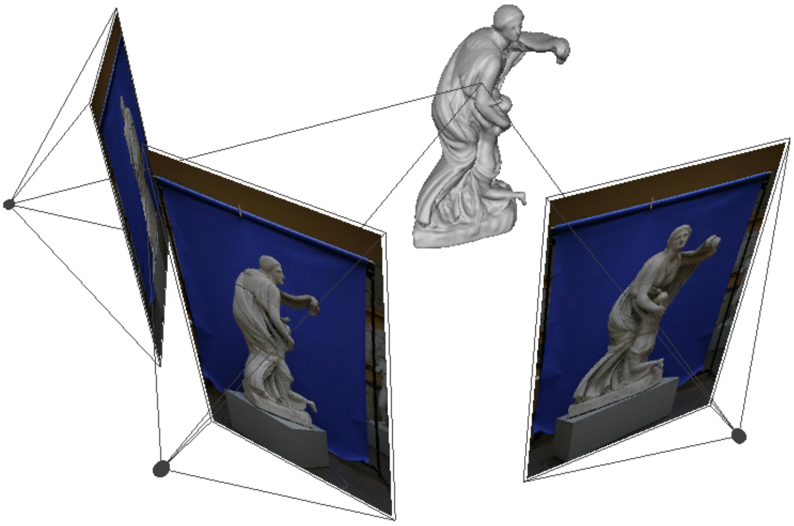


Fig. 3.1: Stereo-based matching. Stereovision relies on the process of matching corresponding points in different views. Assuming Lambertian reflectance properties of the observed object, a 3D point on the surface should project to image points of similar appearance.

age to mitigate this drawback, it still persists, which strongly affects the robustness of the respective approaches. For this reason, mathematically more elegant energy minimization techniques became more popular.

Variational methods for multiview stereo inherit the active contour framework proposed originally for image segmentation [59]. They pose the problem as one of modeling a continuous two-dimensional surface in space by minimizing an appropriate energy functional. This methodology allows to combine a data fidelity criterion on the unknown surface with desired properties like regularity, thus achieving a considerable increase in robustness to image noise. The first approaches are based on the geodesic active contour model [19, 62] by measuring weighted surface area, where weights reflect local photoconsistency. This leads to the following energy functional

$$E(S) = \int_S \rho(s)\, ds, \tag{3.1}$$

where $\rho : V \rightarrow [0,1]$ denotes a photoconsistency map encoding the agreement of the respective image projections. The corresponding flow acts as a smoothness term while at the same time attracting the evolving shape towards photoconsistent locations. Different techniques were applied to model the surface: level sets [36], triangle meshes [33, 31] and graph cuts [134, 54]. A generalization of this approach was developed in [105], which allows to replace the classical pointwise photoconsistency estimation with a global matching score on the entire image domain. A major drawback of the minimal surface model is that it couples data fidelity and regularization. As a result, it is difficult to adjust the regularizing behavior [118]. In particular, the global minimum of the underlying functional is always the empty set, which makes local optimization schemes indispensable. Although graph cut optimization could be applied to minimize (3.1) [134, 54], it has to be performed only within a restricted band so as to avoid the trivial solution and to obtain a meaningful result. As a consequence, this strategy is still local in nature and does not entail any globality guarantees.

In order to react on the shrinking behavior of the minimal surface model and to make global optimization feasible, researchers explored the possibility to extend the model in (3.1) by including additional regional terms over the surface interior and exterior, respectively. This leads to minimizing the following functional

$$E(S) = \lambda \left(\int_{int(S)} \rho_{obj}(x)\, dx \; + \int_{ext(S)} \rho_{bck}(x)\, dx \right) \; + \int_S \rho(s)\, ds, \tag{3.2}$$

where $int(S), ext(S) \subset V$ signify interior and exterior region, and $\rho_{obj}, \rho_{bck} : V \to [0,1]$ define assignment costs associated with both regions. $\lambda \in \mathbb{R}$ is a weighting parameter. Note that, in general, the empty set is no longer a global minimum of (3.2). Yet, a major challenge poses the construction of the regional maps $\rho_{obj}, \rho_{bck} : V \to [0,1]$. Different strategies were proposed to address this task. One of the simplest is to build a ballooning constraint preferring shapes of larger volumes [134, 81]. However, although the empty set can be excluded as a solution, oversmoothing effects still persist making it difficult to reconstruct simultaneously thin protrusions and deep concavities [51, 71]. A straightforward alternative to the ballooning model is the incorporation of data-aware volumetric terms instead of a constant formulation. This could be achieved for example by merging precomputed depth maps to label voxels as interior or exterior with respect to the estimated surface [51, 144, 45]. Although this technique could produce very high-quality reconstructions, it is suboptimal in the sense that the process of 3D modeling is split into two stages. Erroneous decisions in the first stage could propagate to the final estimate, especially at locations of specular reflections or weak texture. Alternative formulations were proposed as building blocks of adaptive 3D modeling schemes. Yet, their accuracy could be unpredictable due to the underlying non-uniform spatial sampling.

A different method for finding a closer specification for the observed shape is to complete the classical photoconsistency information identifying locations, where the surface is likely to pass, by orientation information reflecting the local object normal.[1] In essence, the following cost functional is being minimized

$$E(S) = \; -\lambda \int_S \langle N_S(s), F(s) \rangle\, ds \; + \int_S \rho(s)\, ds, \tag{3.3}$$

where N_S denotes the local surface normal of S and $F : V \to S^2$, where $S^2 \subset \mathbb{R}^3$ is the unit sphere, represents an estimate of the unit outward orientation of the observed shape. Obviously, minimizing (3.3) brings the normals N_S is accordance with the provided vectorfield F while encouraging the surface to pass through photoconsistent locations in space. The question, which arises now, is how to define a vectorfield F with the desired properties. In [12], the authors propose to use the gradient of the photoconsistency map for that purpose. While this simple method improves on the constant ballooning model, it still does not provide state-of-the-art reconstructions. An interesting approach for estimating and integrating a normal field was described in [49]. Thereby, starting from an initial patch, the surface is built progressively based on geometric and photometric criteria. [41, 140] proposed to compute the local scene orientation directly via an optimization procedure over the distortion of the resulting image projections. Optimal oriented patches in space are determined based on the agreement of their projections onto the images, where a direct visibility contact is provided. Another established strategy for obtaining object normals is to utilize precomputed depth-maps [88, 84]. In fact, it could be regarded as an alternative to classical depth-map fusion methods. Once the surface normals are available, the 3D model is usually computed by applying a local optimization procedure like [60]. In contrast, [12] constructs a graph cut framework to minimize (3.3) in a globally optimal manner.

---

[1] The problem of integrating given normal information is considered in Chapter 5 in more detail.

Contribution

The main contribution in the context of multiview stereo is the development of appropriate variational formulations amenable to convex optimization. In particular, we investigate three different energy models. They all consist of a combination of a photoconsistency-based discontinuity term and regional labeling terms, and thus fit into the general form (3.2). The difference between them is in the way the photoconsistency function is computed as well as in the definition of regional costs. The first model follows a classical formulation – the photoconsistency is estimated by just averaging the matching scores of the corresponding projections, and the regional subdivision terms are computed with the approach described in Chapter 2. In the second energy model, the photoconsistency estimation is performed by applying the voting scheme of [33], which results in more precise photoconsistency maps. The regional terms are constructed in the same way as in the first model. In contrast, the third model replaces the silhouette-based foreground/background subdivision of the volume by a more sophisticated one using stereo information, which allows to capture also surface indentations not "visible" in the image silhouettes. This approach is related to depth-map fusion methods. However, the proposed formulation is entirely volumetric and does not involve any preprocessing on the image domain. This entails a series of advantages.

- It avoids discretization problems that could arise in a per-pixel visual ray determination since a ray through a pixel will generally not be adapted to the volume subdivision.
- A crucial issue when measuring photoconsistency along viewing rays is the sampling rate of the discretization. A too dense sampling leads to high computational costs, whereas a too sparse sampling could result in a miss of the maximizing location. In the volumetric framework, the sampling is naturally given by the volume resolution.
- The computational time of the proposed mechanism does not depend on the resolution of the input images but only on the volume resolution.

Additionally, we present a convex relaxation scheme for converting the constructed energy functional into a convex one. Two different numerical schemes, a linearized fixed-point iteration method and a primal-dual method, are examined and compared to each other on their suitability for solving the arising numerical problem.
The main results in this chapter are published in [73, 71, 72].

3.2 Continuous Energy Models for Multiview Stereo

In this section, we present and discuss three different energy models for multiview reconstruction. All three functionals have the same structure combining on-surface photoconsistency and regional costs.
Building the basic template, we revisit the variational formulation in (3.2) and consider the following minimization problem

$$\begin{aligned} E(S) &= \lambda \left(\int_{int(S)} \rho_{obj}(x)\, dx \;+\; \int_{ext(S)} \rho_{bck}(x)\, dx \right) \;+\; \int_S \rho(s)\, ds \\ \hat{S} &= \arg\min_{S \subset V} E(S). \end{aligned} \tag{3.4}$$

The first two terms of the functional impose correct subdivision of the volume into interior/exterior according to the respective regional costs. The last term acts as a constraint

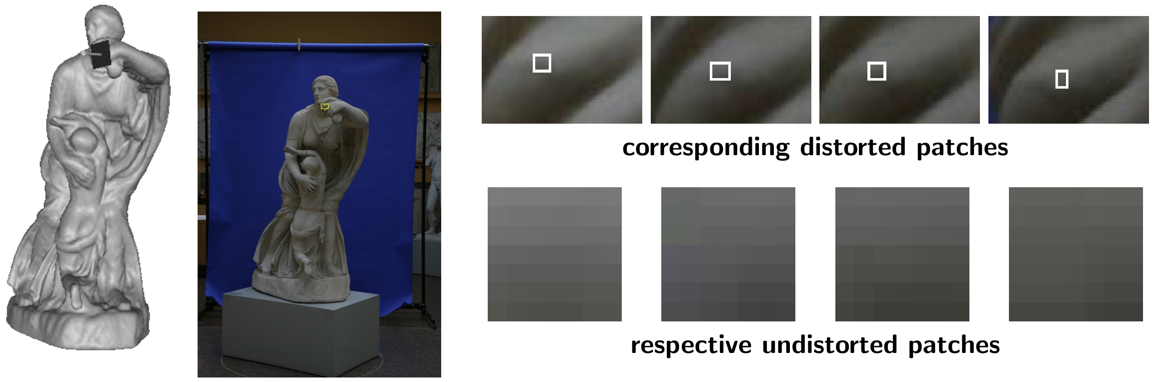


Fig. 3.2: Estimating patch distortion. Corresponding patches are matched via a homography mapping. The homography is defined by means of a local planar approximation of the imaged surface at each point. *From left to right:* Tangent plane and corresponding normal vector for a particular point on the observed object; a region, marked in one of the input images, used for the zoomings on the right; estimated distorted patches of size 7 × 7 pixels to be matched (first row) and respective undistorted versions (second row).

both for smoothness and photoconsistency by seeking for the minimal surface with respect to a Riemannian metric. Hence, it can be considered as a weighted smoothness term. The underlying metric is defined in a way that favors photoconsistent locations and encourages the surface to pass through them. It should be noted that the cost functions $\rho_{obj}, \rho_{bck}, \rho$ may also depend on the surface estimate S so as to estimate explicit visibility or correct image patch distortion. Yet, such a dependency is suppressed here for simplicity since it would considerable exacerbate convex modeling.

In the sequel, we discuss different strategies to define the maps $\rho_{obj}, \rho_{bck}, \rho$, which give rise to different formulations and fairly different results.

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

This model relies on a foreground/background subdivision of the 3D space based on silhouette cues. In particular, we use the framework, introduced in Chapter 2, to derive corresponding regional maps. We reconsider the conditional probabilities for observing colors $I_l(\pi_l(x))$ in images $1, ..., n$

$$\begin{aligned} P_{obj}(x) &= P(\ \{I_l(\pi_l(x))\}_{l=1,\dots,n} \mid x \in int(S)\ ) \\ P_{bck}(x) &= P(\ \{I_l(\pi_l(x))\}_{l=1,\dots,n} \mid x \in ext(S)\ ). \end{aligned} \tag{3.5}$$

While describing joint probabilities for observing particular colors provided given region assignment, the above values lend themselves to formulating respective regional maps by setting

$$\begin{aligned} \rho_{obj}(x) &= -\log P_{obj}(x) \\ \rho_{bck}(x) &= -\log P_{bck}(x). \end{aligned} \tag{3.6}$$

Both values could be additionally normalized to lie within $[0, 1]$. We refer to Chapter 2 for more details on the probabilistic modeling.

Now, we concentrate on the definition of the photoconsistency function ρ in the last term of (3.4). A basic requirement for computing photoconsistency is that camera visibility information is available. To this end, we apply a state-based approach (see Section 1.1). Specifically, we minimize the energy with the classical Euclidean regularizer $\rho(x) = 1$ first, which boils down to the method in Chapter 2. With the resulting surface, one can compute a signed distance function $\phi : V \rightarrow \mathbb{R}$ which in turn allows for normal estimation $N_x = \frac{\nabla\phi}{|\nabla\phi|}$ to each $x \in V$. In this way, the notion of surface normal, which is a local entity, is generalized to the entire volume V. Then, visibility is determined by front-facing cameras according to the estimated normal direction. In particular, photoconsistency is computed in terms of the normalized cross-correlations (see (1.4)) by averaging over front-facing cameras

$$c(x) = \frac{1}{|\mathcal{N}(x)|} \sum_{(i,j)\in\mathcal{N}(x)} NCC(\pi_i(x), \pi_j(x)), \tag{3.7}$$

where $\mathcal{N}(x)$ denotes the set of all front-facing camera pairs according to the normal direction N_x. Specifically, we could define

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

where V_k is the viewing direction of camera k. Yet, the above formulation takes only local visibility criteria into account. The presence of self-occlusions or obstacles, blocking the field of view, is not explicitly modeled. For this reason, we rely on the following generalized formulation

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

where $Vis : V \rightarrow \mathcal{P}(\{1,\ldots,n\})$ is a global visibility map and $\mathcal{P}(X)$ denotes the power set of X. In Appendix A, we give more details on how such a volumetric visibility map can efficiently be obtained. In our experiments, we set $\gamma_{max} = 60°$.

Another important issue associated with the computation of the matching score concerns the estimation of patch distortion. A square patch in one of the images does not in general correspond to a square patch in the other images due to the non-linear nature of the projection mapping. In order to take this aspect into account, we locally approximate the surface by its tangent plane according to [36] (see Fig. 3.2). As a result, distortion can be estimated via a homography mapping. For a camera pair (i,j) and a 3D point $x \in V$ it is given by

$$H_{ij,x} = R_{ij}^T - \frac{R_{ij}^T T_{ij} N_x^T}{N_x^T x}, \tag{3.8}$$

where $R_{ij} \in \mathbb{R}^{3\times3}$ and $T_{ij} \in \mathbb{R}^3$ denote the relative rotation and translation between the local coordinate systems of both cameras. All involved entities are defined in the coordinate frame of the reference camera i. Note that the matrix $H_{ij,x} \in \mathbb{R}^{3\times3}$ induces a projective transformation between the two image planes, which gives rise to a more accurate model than traditionally used affine transformations. In particular, it is only determined up to a scale factor. Now, we can compute a NCC score based on the proposed local distortion model

$$NCC(\pi_i(x), \pi_j(x)) = \frac{1}{c_1 c_2} \sum_{p\in\mathcal{P}} \langle I_i(p) - \bar{I}_i(\pi_i(x)), I_j(H_{ij,x}(p)) - \bar{I}_j(\pi_j(x))\rangle, \tag{3.9}$$

where $\mathcal{P}$ stands for a square patch around $\pi_i(x)$ in the reference image i, $\bar{I}_i$ and $\bar{I}_j$ are the corresponding mean values, and c_1, c_2 are normalization constants defined in analogy with (1.4). The expression $H_{ij,x}(p)$ should be evaluated by rewriting the vector p in homogeneous coordinates (in the local frame of camera i) and reverting back to non-homogeneous coordinates after applying $H_{ij,x}$. Note that the size of each patch is determined according to its projection on the reference image plane rather than being set to a fixed size on the tangent plane. This avoids sampling problems on the image domain. For all our experiments we used 7×7 pixel windows.

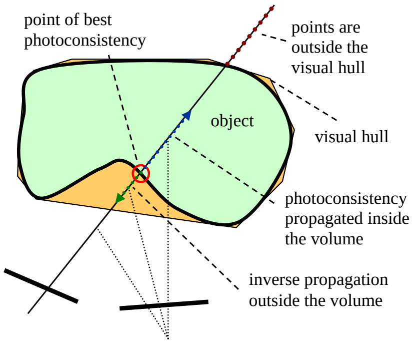


Fig. 3.3: Propagation of photoconsistency. Illustration of the proposed approach to spread stereo information inside a volume. Spatial labeling of the volume as interior (blue arrow) or exterior (green arrow) is derived based on the location of maximal photoconsistency (red circle) along the depicted viewing ray.

For each 3D point $x \in V$ we get some score $c(x)$ between -1 and 1, where 1 implies perfect correlation. This value is then mapped to the unit interval $[0, 1]$ using the following function proposed in [134]

$$f(s) = 1 - \exp\left(-\tan\left(\frac{\pi}{4}(s-1)\right)^2 / \sigma^2\right). \tag{3.10}$$

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

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

The classical photoconsistency estimation used by the previous model generally yields noisy measures due to homogeneity or repeatability of the texture pattern, which could result in noisy reconstructions. For that reason, this model relies on a more elaborate approach to increase the accuracy of the photoconsistency computation. The basic idea is to ask each camera to give a vote to a point in space. The vote is accepted only if the optimum is reached at the current point. This methodology leads to a considerable increase in the precision of the corresponding photoconsistency maps (see Fig. 3.4). This scheme is further generalized by the next model and described in more detail in Section 3.2.

Thus, the current model combines the silhouette-based volume subdivision used by energy model I and the denoised photoconsistency estimation proposed in the next paragraph.

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

A major limitation of both previously presented energy models is that they use silhouette-based regional terms which do not capture surface indentations, as these do not affect the observed silhouettes. As a result, the functionals introduce a bias towards the maximal silhouette-consistent shape, i. e. the visual hull. In order to address this shortcoming, the current model replaces previously proposed regional terms by more accurate ones (see Fig. 3.4). The basic idea is to propagate classical *on-surface* photoconsistency within the volume and thereby define confidence values for lying *inside* or *outside* the observed object (see Fig. 3.3). In the following, this approach is explained in more depth.

The main difficulty in defining volume subdivision terms ρ_{obj} and ρ_{bck} is the fact that the state of each voxel in space (inside/outside the object) is affected by potentially distant points. We address this problem by measuring photoconsistency along visual rays and exploiting the following property of silhouette-consistent shapes.

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

The above statement is quite obvious. In fact, if there exists a visual ray through a point x, which does not intersect the real surface S_{gt}, then x does not project within the silhouette of the respective image. Hence, this point cannot lie in the interior of a silhouette-consistent shape. Note that the above property is fulfilled for the maximal consistent shape as well as for any subset of it. This naturally leads to the following idea. We can compute photoconsistency along each visual ray and take the position, where its maximum is reached, as a potential intersection with the real surface S_{gt}; see Figure 3.3. Of course, a viewing ray could intersect S_{gt} more than once, but only the first intersection is expected to be photoconsistent according to a certain set of neighboring cameras. Based on this observation, we can convert classical photoconsistency measures ρ, describing the likelihood for each point in space to lie *on* the surface, into regional terms ρ_{obj}, ρ_{bck} representing an *interior/exterior* assignment.

The details of this formulation are exposed in the sequel. We start with an initial silhouette-based surface approximation S_I computed as described in Chapter 2. Due to the above property, we consider all points x lying in the interior $int(S_I)$ of S_I and corresponding visual rays passing through x. Let $r_j(x,t)$ be the visual ray to camera j parametrized by t starting at the camera position. Let t_{cur} be the position of x along the ray. We measure photoconsistency along the ray with respect to another camera i as

$$C_i^j(x,t) = NCC(\pi_i(r_j(x,t)), \pi_j(r_j(x,t))). \tag{3.11}$$

Once again, patch distortion is approximated by a local homography mapping defined by the normal direction N_x. Since it is expected that the orientation at the surface intersection point of a viewing ray corresponding to a front-facing camera is similar to that measured at x, the same normal direction N_x can be used to estimate distortion along the entire ray r_j to simplify computations. Note that the second term in (3.11) stays constant for varying t as points on the ray r_j always project onto the same location in image I_j. This formulation can easily be extended to multiple cameras

$$C^j(x,t) = \sum_{i=1}^{m} w_i^j(x) C_i^j(x,t). \tag{3.12}$$

Thereby, we sum only over neighboring cameras according to the normalized viewing direction $V_j(x)$ of camera j. That is, camera i is excluded if $\alpha_i^j(x) = \angle(V_i(x), V_j(x)) > \alpha_{max}$ for some bounding angle α_{max}. Moreover, global visibility criteria are additionally taken into account, i. e. camera i is embraced only if $i \in Vis(x)$. The weights w_i^j are computed as

$$w_i^j(x) = \frac{\alpha_{max} - \alpha_i^j(x)}{\sum_{k=1}^{m} \left(\alpha_{max} - \alpha_k^j(x)\right)} \tag{3.13}$$

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

$$\begin{aligned} C^j_{max}(x) &= \max_t C^j(x,t) \\ t_{max} &= \arg\max_t C^j(x,t). \end{aligned} \tag{3.14}$$

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

$$\begin{aligned} \rho^j_{obj}(x) &= H(t_{max} - t_{cur}) \cdot (1 - f(C^j_{max})) \\ &\quad + (1 - H(t_{max} - t_{cur})) \cdot f(C^j_{max}) \\ \rho^j_{bck}(x) &= H(t_{max} - t_{cur}) \cdot f(C^j_{max}) \\ &\quad + (1 - H(t_{max} - t_{cur})) \cdot (1 - f(C^j_{max})), \end{aligned} \tag{3.15}$$

where H is the Heaviside function

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

and f is defined in (3.10). The computed values depend on whether the maximal photoconsistency location t_{max} lies before or behind the current point t_{cur}. If for example $t_{max} < t_{cur}$, ρ^j_{obj} decreases and ρ^j_{bck} increases with the maximal measured photoconsistency C^j_{max} accounting for uncertainties as a result of mismatches. In effect, the Heaviside function H realizes this case differentiation. The final regional costs can be computed by simple averaging over single rays r_j, which yields

$$\begin{aligned} \rho_{obj}(x) &= \frac{1}{l} \sum_{j=1}^{l} \rho^j_{obj}(x) \\ \rho_{bck}(x) &= \frac{1}{l} \sum_{j=1}^{l} \rho^j_{bck}(x). \end{aligned} \tag{3.17}$$

In practice, only visual rays of front-facing cameras with respect to the normal N_x, which see the current point x, are considered as described in 3.2. Note that $\rho_{obj}(x) + \rho_{bck}(x) = 1$ for all $x \in V$. In case of photometrically challenging scenes contaminated by noise and shading effects, more sophisticated fusion strategies could be used. For example, we experimented with a weighting procedure based on the variance of the measured photoconsistency values along viewing rays, but we could not observe any visible improvements in the reconstructions. One could note that the process of maximization of photoconsistency along visual rays can also be exploited in the computation of on-surface costs $\rho(x)$. In particular, identifying locations of optimal photoconsistency allows to "clean up" the usually quite noisy photoconsistency map by means of the voting scheme described below. The basic idea is to treat all potential causes of mismatches like occlusion, image noise, lack of texture etc. uniformly as outliers in the matching process. Specifically, the photoconsistency value $\rho(x)$ for a given 3D point x is computed by asking every image j to give a vote for that location and subsequently fusing the votes to a final score

$$\rho(x) = \exp\{-\mu \sum_{j=1}^{l} \text{VOTE}_j(x)\}, \tag{3.18}$$

where VOTE_j denotes the vote of camera j and $\mu \in \mathbb{R}$ is a rate-of-decay parameter which in our experiments was set to 0.15. Now, the idea is to permit a camera j to give a vote to the 3D location x only if the correlation score along the corresponding viewing ray takes its maximum at x, i. e.

$$\mathrm{VOTE}_j(x) = \begin{cases} C^j_{max}(x) & \text{if } t_{max} = t_{cur} \\ 0 & \text{otherwise.} \end{cases} \tag{3.19}$$

Note that the presented voting scheme accounts for outliers due to occlusions, noise or shading effects as well as matching ambiguities. It should be noted that the formulation in (3.19) could be generalized by taking the presence of a local extremum (instead of a global one) of the photoconsistency score along the ray as a criterion for a vote. While this method could lead to some minor accuracy improvements in certain cases, its implementation could be very tricky due to jaggedness of the photoconsistency map.

In order to accelerate the computation of the data terms of the three energy models, we use a banded multiresolution scheme by starting with a coarse volume resolution and subsequently restricting the computations at finer levels. We carry out 1-4 iterations at each level and update the data terms iteratively. The surface estimate from the previous step provides orientation information, needed to obtain correct patch distortion, and also serves as a basis for computing the global visibility map Vis (see Appendix A). In this way, useful geometric information is propagated to the final result.

Comparison of the Three Energy Models

Fig. 3.4 shows a comparison between the data terms used by the three energy models on two data sets which will serve throughout the thesis for further evaluations. Visualized are cross-sections through all data volumes at the lowest resolution: ρ_{obj}, ρ_{bck} and ρ, respectively. The traditional silhouette-based subdivision technique and photoconsistency estimation (upper row) are opposed to the more elaborate stereo-based approach and voting scheme (lower row). As expected, the naive silhouette-based method fails to produce accurate regional terms at concavities like the legs of the dino figurine or the back of the temple model in contrast to the stereo-based one. As a result, the corresponding volume subdivision and on-surface discontinuity costs compete each other in such areas, which makes the task of finding an appropriate weighting very challenging or even unfeasible. Moreover, the voting scheme used by the second and third energy models yields notably more precise photoconsistency maps by removing the influence of repeated texture patterns or accidental matching. However, as a side effect, this approach could erroneously suppress photoconsistency in case of occlusions or ambiguous texture (see for example the vertical inside wall of the temple model), which in turn lets the regional terms play the decisive role. See Fig. 3.6 and 3.7 for the corresponding reconstructions.

Comparing the runtimes of the three approaches for computing the data terms, it is evident that the first one requires substantially less computational efforts than the other two. Indeed, it involves projecting each point in space onto a set of images and matching the respective projections while the voting scheme used by energy models II and III envisages considering different viewing rays to each particular 3D point. With the applied banded strategy the runtime difference is in the order of the band width. In practice, the particular factor is even larger, since at grazing viewing angles the number of considered voxels within the band along the respective viewing ray is usually larger than the thickness of the band. Computational times for the data sets in Fig. 3.4 are given in Section 3.4.

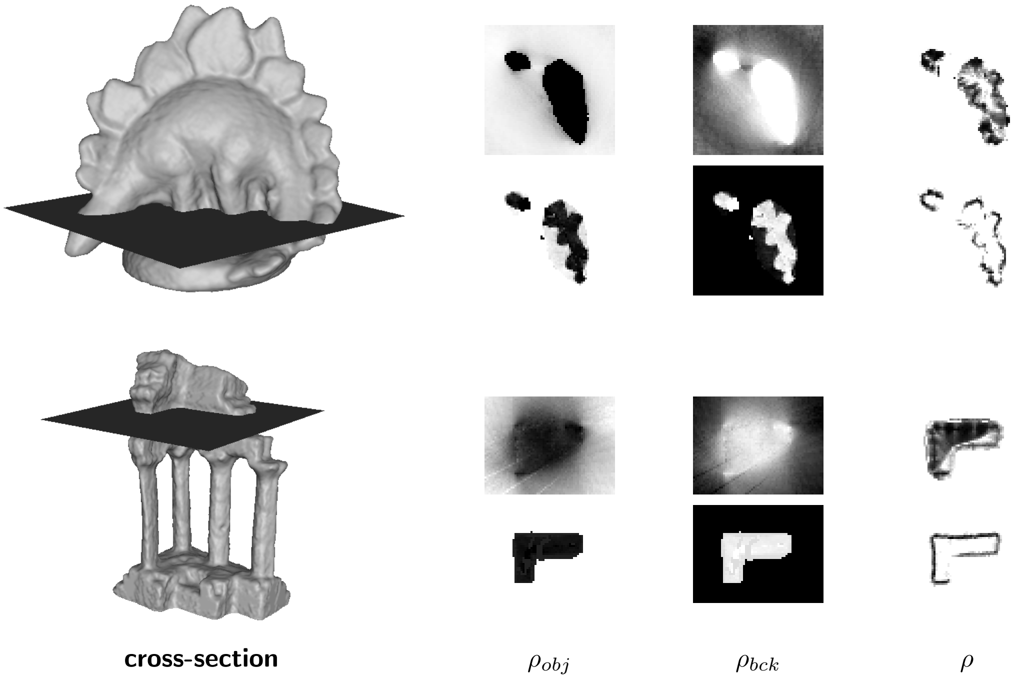


Fig. 3.4: Comparison between the volumetric data terms used by energy models I, II and III for two data sets. Visualized are cross-sections through all data volumes: ρ_{obj}, ρ_{bck} and ρ, respectively. The silhouette-based subdivision technique and photoconsistency estimation used by energy models I and II (upper row) are opposed to the more elaborate stereo-based approach and voting scheme used by models II and III (lower row). Intensity values correspond to estimated costs. Note that stereo information allows to capture surface indentations in contrast to silhouettes and thus produces more accurate regional terms. Note also that the voting scheme generally yields more precise photoconsistency maps but could fail in case of occlusions or ambiguous texture.

3.3 Convex Relaxation and Numerical Optimization

In the previous section, we considered different continuous energy models for multiview stereo sharing the same variational structure (3.4). In this section, we focus on their optimization. In particular, we show how the formulation in (3.4) can be cast in a convex relaxation framework and discuss the optimization of the arising numerical problem.

Convex Relaxation

The first step towards a convex formulation of (3.4) is the switch to an implicit representation. To this end, we introduce an implicit binary function $u = \mathbf{1}_{int(S)}$ as the characteristic function of the surface interior $int(S)$. This implies $1 - u = \mathbf{1}_{ext(S)}$. Now, we can rewrite the minimization problem in (3.4) in the form

$$\begin{aligned} E(u) &= \lambda \left(\int_V \rho_{obj}(x)\, u(x)\, dx \; + \; \int_V \rho_{bck}(x)\; (1 - u(x))\; dx \right) \; + \; \int_V \rho(x)\, |\nabla u(x)|\, dx \\ \hat{u} &= \arg \min_{u: V \to \{0,1\}} E(u). \end{aligned} \tag{3.20}$$

The final surface estimate $\hat{S}$ can be derived from the relation $int(\hat{S}) = \{\, x \in V \mid \hat{u}(x) = 1 \,\}$. It is essential to observe that solving the minimization problem in (3.20) boils down to minimizing a functional of the general form (1.6). Indeed, the functional in (3.20) reads

$$E(u) = \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x))\ u(x)\,dx\ +\ \int_V \rho_{bck}(x)\,dx\ +\ \int_V \rho(x)\,|\nabla u(x)|\,dx. \tag{3.21}$$

One can note that the second term of the above functional is constant and does not depend on u. Hence, it can be omitted in the minimization. In other words, minimizing (3.21) is equivalent to minimizing

$$E(u) = \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x))\ u(x)\,dx\ +\ \int_V \rho(x)\,|\nabla u(x)|\,dx. \tag{3.22}$$

It is obvious that the functional in (3.22) is of the general form (1.6). The formulation in (1.6) can be obtained by setting $f = \rho_{obj} - \rho_{bck}$ and defining the local metric $||.||_x = \rho(x)\,|.|$.[2] Based on the above observation, we can apply a convex relaxation strategy by adhering to the steps outlined in Section 1.4. In particular, the domain of binary labeling functions $\{\,u\,|\,u : V \to \{0,1\}\,\}$ can be relaxed to $\{\,u\,|\,u : V \to [0,1]\,\}$ by retaining the following important property.

Theorem 8. *Let* $u^\star : V \to [0,1]$ *be a global minimizer of the functional in* (3.22) *over the relaxed domain* $\{\ u\ |\ u : V \to [0,1]\ \}$*. Then, for almost any threshold* $\mu \in (0,1)$*, the binary function* $\mathbf{1}_{\Sigma_\mu(u^\star)} : V \to \{0,1\}$ *with* $\Sigma_\mu(u) = \{\ x \in V\ |\ u(x) > \mu\ \}$ *is also a global minimizer.*

Proof. The above statement can be inferred directly from Theorem 6 and the fact that the functional in (3.22) is a special case of the one in (1.6). ☐

The above theorem is central for solving the minimization problem in (3.4). Basically, it states that the problem boils down to solving

$$\begin{aligned} E(u) &= \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x))\ u(x)\,dx\ +\ \int_V \rho(x)\,|\nabla u(x)|\,dx \\ \hat{u} &= \arg \min_{u:V\to[0,1]} E(u), \end{aligned} \tag{3.23}$$

which poses a constrained convex optimization problem. Once a solution $\hat{u}$ is obtained, a corresponding surface estimate $\hat{S}$ can be derived by choosing a threshold $\mu \in (0,1)$ and setting $int(\hat{S}) = \{\,x \in V\,|\,\hat{u}(x) > \mu\,\}$. In practice, a mesh representation can be calculated by applying the Marching Cube algorithm [89].

Numerical Optimization

In this section, we present, discuss and compare two numerical algorithms for solving the constrained convex minimization problem in (3.23).

Before introducing the numerics, we conduct some analysis of the problem in (3.23). It turns out that it does not pose a general constrained convex optimization problem but one of a very special form, which is an important fact that can be exploited. One could observe that the constraints $u \in [0,1]$ do not play a decisive role. As we do not have any smoothness requirements on u, the constraints ensure that the minimization problem possesses a feasible solution. Note that if we optimize over all real-valued functions $u : V \to \mathbb{R}$, the values of u will diverge to $+\infty$ for $\rho_{obj} - \rho_{bck} < 0$ and to $-\infty$ for $\rho_{obj} - \rho_{bck} > 0$. Yet, after a certain number

---

[2] This local metric can be derived from the Riemannian metric $\langle .,. \rangle_x = \rho(x)^2 \langle .,. \rangle$, where $\langle .,. \rangle$ denotes the Euclidean inner product. See [19] for more details.

of iterations, projecting back to the unit interval $[0,1]$ will not give any changes. In other words, we can obtain a globally optimal solution of (3.23) by ignoring the constraints during the optimization process and projecting the result after a sufficient number of iterations onto the domain determined by the constraints. In practice, the projection could be performed after each iteration or after a certain number of iterations. Yet, the definition of a convergence criterion, that ensures a correct solution, requires more care in that case. We can deduce from the above observations that any method for unconstrained optimization lends itself to solving (3.23).

Linearization and Fixed-Point Iteration

We start solving the minimization problem in (3.23) in a continuous manner by setting up the respective Euler-Lagrange equation

$$\lambda \left(\rho_{obj} - \rho_{bck}\right) - \operatorname{div}\left(\rho \, \frac{\nabla u}{|\nabla u|}\right) = 0. \tag{3.24}$$

Thereby, the arguments are omitted for the sake of simplicity.

An alternative to applying a gradient-based iterative scheme, aiming at decreasing the energy in each step, is to solve the PDE in (3.24) directly. Yet, (3.24) exhibits a non-linear diffusion equation which is not trivial to solve. The basic idea behind the linearized fixed-point iteration method is to consider the non-linear diffusivity term $g = \frac{\rho}{|\nabla u|}$ as constant for a fixed u and to update it only after a certain number of iterations.[3] This leads to a linear PDE of the form

$$\lambda \left(\rho_{obj} - \rho_{bck}\right) - \operatorname{div}\left(g \, \nabla u\right) = 0, \tag{3.25}$$

which after discretization yields a sparse linear system of equations. In the following, we present a concrete numerical realization of this idea.

Let us for convenience summarize the constant part in (3.25) as $f = \lambda \left(\rho_{obj} - \rho_{bck}\right)$. Thus, our goal is to solve

$$f - \operatorname{div}\left(g \, \nabla u\right) = 0 \tag{3.26}$$

on a discretized volume grid $\tilde{V}$ defined in (1.8). For a voxel $(l,m,n) \in \{1,\dots,N_1-2\} \times \{1,\dots,N_2-2\} \times \{1,\dots,N_3-2\}$ within the volume interior (boundary voxels are handled analogously) we obtain

$$\begin{aligned} 0 &= f_{l,m,n} - \operatorname{div}\left(g_{l,m,n} \nabla u_{l,m,n}\right) \\ &= f_{l,m,n} - \frac{\partial}{\partial x}\left(g \, u_x\right)_{l,m,n} - \frac{\partial}{\partial y}\left(g \, u_y\right)_{l,m,n} - \frac{\partial}{\partial z}\left(g \, u_z\right)_{l,m,n}, \end{aligned} \tag{3.27}$$

where u_x, u_y and u_z denote partial derivatives with respect to x, y and z, respectively. The differentiation in the above expression can be carried out via central differences at intermediate grid points. For example, the first term reads

$$\frac{\partial}{\partial x}\left(g \, u_x\right)_{l,m,n} = \left(g \, u_x\right)_{l+1/2,m,n} - \left(g \, u_x\right)_{l-1/2,m,n}. \tag{3.28}$$

Now, we can discretize the partial derivatives of u by means of forward or backward differences, respectively. Thus, for the terms in (3.28) we have

$$\begin{aligned} \left(g \, u_x\right)_{l+1/2,m,n} &= g_{l+1/2,m,n}\left(u_{l+1,m,n} - u_{l,m,n}\right) \\ \left(g \, u_x\right)_{l-1/2,m,n} &= g_{l-1/2,m,n}\left(u_{l,m,n} - u_{l-1,m,n}\right). \end{aligned} \tag{3.29}$$

---

[3] The diffusivity term g is sometimes referred to as lagged diffusivity.

Finally, plugging (3.28) and (3.29) into (3.27) and writing out the expressions for the other two variables y and z yields

$$0 = f_i - \sum_{j \in \mathcal{N}(i)} g_{i \sim j} u_j + \sum_{j \in \mathcal{N}(i)} g_{i \sim j} u_i, \tag{3.30}$$

where $\mathcal{N}(i)$ denotes the 6-neighborhood of voxel i and $g_{i \sim j}$ denotes the diffusivity between voxel i and its neighbor j. In particular, we average the respective values by setting

$$g_{i \sim j} = \frac{g_i + g_j}{2}. \tag{3.31}$$

Note that in the above formulation we revert to a more general indexing, where a single index signifies a voxel. Moreover, it should be noted that evaluating the diffusivities at particular volume grid points requires the addition of a small constant $\epsilon > 0$ that prevents the term to become infinite when $|\nabla u_i| = 0$. More precisely

$$g_i = \frac{\rho_i}{\sqrt{|\nabla u_i|^2 + \epsilon^2}}. \tag{3.32}$$

While the incorporation of such a constant ϵ modifies the original PDE, the modification is negligible for small values. In our experiments, we used $\epsilon = 0.001$.

The second step in the derivation process involves building a system of linear equations and solving it to obtain an update of u. A closer look at the formulation in (3.30) reveals that it can be written in matrix form as $A\vec{u} = b$ by stacking the values of u at all voxels in a single vector $\vec{u} \in \mathbb{R}^N$, where $N = N_1 \cdot N_2 \cdot N_3$ is the number of voxels. The $N \times N$-matrix $A = (a_{ik})_{i,k=1,\dots,N}$ is given by

$$a_{ik} = \begin{cases} -\sum_{j \in \mathcal{N}(i)} g_{i \sim j}, & \text{if } i = k \\ g_{i \sim k}, & \text{if } k \in \mathcal{N}(i) \\ 0, & \text{otherwise} \end{cases} \tag{3.33}$$

and

$$b = (f_1, \dots, f_N)^T. \tag{3.34}$$

It is easy to see that the matrix A is sparse and diagonally dominant (although not strictly diagonally dominant). Hence, an efficient iterative algorithm exploiting this structure, like Successive Overrelaxation (SOR), could be applied to solve the linear system. This leads to the following evolution scheme for $k = 0, 1, 2, \dots$

$$u_i^{(k+1)} = (1 - \omega)\, u_i^{(k)} + \omega \frac{\sum\limits_{j \in \mathcal{N}(i), j<i} g_{i \sim j} u_j^{(k+1)} + \sum\limits_{j \in \mathcal{N}(i), j>i} g_{i \sim j} u_j^{(k)} - f_i}{\sum\limits_{j \in \mathcal{N}(i)} g_{i \sim j}}, \tag{3.35}$$

where $\omega \in \mathbb{R}$ is an over-relaxation parameter that has to be chosen in the interval $(0, 2)$ for the method to converge. Yet, the value giving optimal performance is usually in $(1, 2)$. For the particular energy model at hand we obtained the fastest convergence for $\omega = 1.85$. We should keep in mind that the diffusivities $g_{i \sim j}$ have to be updated after a certain number of iterations. Introducing an index l to explicitly emphasize this produces the generalized evolution process

$$u_i^{(l,k+1)} = (1 - \omega)\, u_i^{(l,k)} + \omega \frac{\sum\limits_{j \in \mathcal{N}(i), j<i} g_{i \sim j}^l u_j^{(l,k+1)} + \sum\limits_{j \in \mathcal{N}(i), j>i} g_{i \sim j}^l u_j^{(l,k)} - f_i}{\sum\limits_{j \in \mathcal{N}(i)} g_{i \sim j}^l}. \tag{3.36}$$

After the linear solver yields a sufficiently good approximation (we iterate for $k = 1, ..., 10$), we update the diffusivities $g^l_{i\sim j}$ and solve the next linear system. Moreover, the constraints $u \in [0, 1]$ are enforced after each iteration by simply clipping the values lying outside the unit interval.
In some cases, different derivation procedures lead to similar resulting update schemes. In fact, the proposed linearized fixed-point iteration technique can be interpreted as a specific quasi-Newton method. This is an important fact which will facilitate further analysis.

Proposition 2. *The linearized fixed-point iteration scheme in* (3.36) *can be interpreted as a specific quasi-Newton method.*

Proof. For convenience we will consider the case when the diffusivity terms are updated after each iteration and the implicit Gauss-Seidel step (SOR) is replaced by an explicit Jacobi one. A generalization to the procedure in (3.36) is straightforward.
First, we observe that the gradient of the functional in (3.23) reads

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

where f summarizes the constant volume subdivision term, i. e. $f = \lambda\,(\rho_{obj} - \rho_{bck})$. By reverting to a discrete setting, we get $\nabla E(\vec{u}) \in \mathbb{R}^N$. A quasi-Newton step is given by

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

where $B^t \in \mathbb{R}^{N\times N}$ and $\omega \in \mathbb{R}$ is a time-step parameter. Based on the previously derived discretization, we set

$$(\nabla E(\vec{u}^t))_i = f_i - \sum_{j\in\mathcal{N}(i)} g^t_{i\sim j} u^t_j + \sum_{j\in\mathcal{N}(i)} g^t_{i\sim j} u^t_i$$

for $i = 1, \dots, N$. The index t in the diffusivities $g^t_{i\sim j}$ indicates that they are defined with respect to u^t, i. e. $g^t_i = \frac{\rho_i}{|\nabla u^t_i|}$.
Finally, we specify $B^t = (b^t_{ik})_{i,k=1,\dots,N}$ as

$$b^t_{ik} = \begin{cases} \sum_{j\in\mathcal{N}(i)} g^t_{i\sim j}, & \text{if} \quad i = k \\ 0, & \text{otherwise} \end{cases} .$$

Summarizing all calculations leads to the evolution scheme

$$u^{t+1}_i = (1-\omega)\, u^t_i + \omega \, \frac{\displaystyle\sum_{j\in\mathcal{N}(i)} g^t_{i\sim j} u^t_j - f_i}{\displaystyle\sum_{j\in\mathcal{N}(i)} g^t_{i\sim j}}.$$

□

Remarks.
1) All matrices B^t and $(B^t)^{-1}$ arising in the optimization are diagonal matrices with positive diagonal elements. Hence, they are symmetric and positive-definite, which guarantees that each update direction is a descent direction.
2) A closer look at the evolution scheme reveals that the matrices $(B^t)^{-1}$ are designed in a way that some components are encouraged while others are damped. For example in areas, where

u is approximately constant, the local gradient ∇u will be small in magnitude as well as the corresponding diagonal entry of $(B^t)^{-1}$, which will result in a damping effect. Analogously, at locations of high variation of u, the diffusivity terms $g = \rho/|\nabla u|$ may become small enough to intensify the respective component of the functional gradient. This results in favorable properties of the proposed fixed-point iteration scheme which is both fast and numerically stable.
3) The over-relaxation parameter ω plays, in essence, the role of a time-step. Hence, in case of unstable behavior (which is unlikely to occur), the value of ω should be decreased.

Primal-Dual Method

Now, we present an alternative numerical technique for solving the minimization problem in (3.23) – the primal-dual method. Similar to the linearized fixed-point iteration approach, it relies on simplification via linearization. However, this time not the Euler-Lagrange equation but the energy functional itself is being linearized. We already encountered a primal-dual formulation in Chapter 2. Yet, here we are confronted with a more general energy model.
The basic idea of the primal-dual method is to convert the minimization problem in (3.23) into a linear saddle-point problem by introducing an auxiliary dual variable $\xi : V \to \mathbb{R}^3$. Indeed, the optimization problem in (3.23) can be equivalently formulated as

$$\min_{u:V\to[0,1]} \max_{\xi\in K} \; \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x)) \; u(x)\, dx \; + \; \int_V \langle \nabla u(x), \xi(x) \rangle \, dx, \tag{3.37}$$

where $K = \{\, \xi : V \to \mathbb{R}^3 \,|\, |\xi(x)| \leq \rho(x) \; \forall x \in V \,\}$. The dual formulation used above is justified by the fact that the inner product of two vectors is maximized if they have the same direction and maximal magnitude. Note that the energy functional in (3.37) is linear in both u and ξ. In this way, we get a minimization problem in u and a maximization one in ξ. Thus, the optimization in (3.37) can be performed by gradient descent in u and gradient ascent in ξ in alternating manner. Starting with some initial $(u^{(0)}, \xi^{(0)})$ and setting $\bar{u}^{(0)} = u^{(0)}$, we obtain the iterative procedure

$$\begin{aligned} \xi^{(k+1)} &= \Pi_K(\xi^{(k)} \; + \; \sigma \, \nabla \bar{u}^{(k)}) \\ u^{(k+1)} &= \Pi_{C_{rel}}(u^{(k)} \; + \; \tau \, (\operatorname{div}(\xi^{(k+1)}) - f)) \\ \bar{u}^{(k+1)} &= 2u^{(k+1)} \; - \; u^{(k)}, \end{aligned} \tag{3.38}$$

where $C_{rel} = \{\; u \mid u : V \to [0,1] \;\}$ and $\sigma, \tau > 0$ denote time-step parameters. Moreover, $\Pi_X(.)$ is the projection operator for the set X. The projection onto C_{rel} is realized by simple clipping while the projection for K involves modifying the length of the respective vector. In our implementation, we set $\sigma, \tau = 0.1$.

Convergence Criterion

As previously discussed, a convergence condition for both numerical schemes should carefully be chosen in order to guarantee a correct solution. Suppose, we are given a convergence tolerance $\epsilon_{conv} > 0$, $u^{(cur)}$ is the current estimate and $u^{(prev)}$ – the one at the previous iteration step. Below, we present three different stopping criteria which we found to work well in practice.

- The iterative procedure is terminated as soon as $|u^{(cur)} - u^{(prev)}| < \epsilon_{conv}$.
- The iterative procedure is terminated as soon as $|E(u^{(cur)}) - E(u^{(prev)})| < \epsilon_{conv}$.

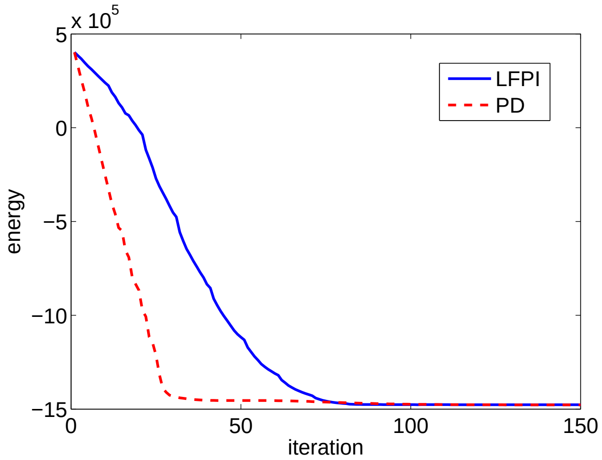


Fig. 3.5: Linearized fixed-point iteration (LFPI) vs. primal-dual (PD) method. Both numerical techniques, evaluated on the "dinoRing" data set (see Section 3.4), offer a fast and stable convergence behavior. Yet, the primal-dual method is substantially faster (about a factor of 3).

- If $\xi^{(cur)}$ denotes the dual variable at the current iteration step, the iterative procedure is terminated as soon as the gap between primal and dual energy falls below ϵ_{conv}, i. e. $|E_p(u^{(cur)}) - E_d(\xi^{(cur)})| < \epsilon_{conv}$, where
$$E_p(u) = \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x))\ u(x)\,dx\ +\ \int_V \rho(x)\,|\nabla u(x)|\,dx \tag{3.39}$$
and
$$E_d(\xi) = \int_V \min\{\ \lambda\,(\rho_{obj}(x) - \rho_{bck}(x))\ -\ \mathrm{div}(\xi(x)),\ 0\ \}\ dx. \tag{3.40}$$
Thereby, the primal energy $E_p(u)$ measures the maximum of E with respect to ξ for a given fixed u. Analogously, $E_d(\xi)$ is the minimal energy with respect to u for a fixed ξ. The identity of both values indicates the presence of a saddle point.

In practice, for stability reasons, it is advisable to check the convergence criterion over a certain number of iterations (more than two) before terminating the evolution process.

Comparative Evaluation

We compare the linearized fixed-point iteration (LFPI) method and the primal-dual (PD) method in terms of runtime, potential for parallel computing and memory requirements.
When talking about runtimes, first of all it should be noted that both numerical schemes possess linear convergence rate. Yet, this fact doesn't reveal the typical physically measured computational times. To this end, for a particular data set we plot the energetic evolution, depicted in Fig. 3.5, and estimate the computational costs for one iteration. It is evident that both numerical techniques offer a fast and stable convergence behavior. Yet, the primal-dual method is substantially faster (about a factor of 3) in terms of number of iterations required. The acceleration factor changes if we measure the runtime in seconds. Surprisingly, using a CPU implementation or a GPU one makes a difference. The computational costs for one iteration on the CPU of the primal-dual scheme are considerably higher than these for the linearized fixed-point iteration approach. Indeed, the first requires two separate loops for updating primal and the dual variable, respectively, while the latter necessitates only one

single loop over the entire volume. As a consequence, the overall computational time of the primal-dual method, executed on the CPU, is only insignificantly lower than that of the linearized fixed-point iteration method. In contrast, a GPU implementation leads to different observations. This time, the computational costs for one primal-dual iteration are lower than these for a linearized fixed-point iteration. This is due to the fact that a primal-dual step envisages a direct update of the respective values while a linearized fixed-point step involves summing over all direct neighbors. Finally, with a GPU implementation we obtain an overall acceleration factor of about 5 in favor of the primal-dual method, measured in seconds.

A closer look at the evolution equations of both numerical schemes – (3.36) and (3.38) – reveals that the two of them are perfectly suited for parallelization as the computations for individual voxels can be performed independently. Hence, a GPU implementation is in both cases straightforward. Yet, some care should be taken when parallelizing the linearized fixed-point iteration method since it involves applying the SOR approach for solving the arising system of linear equations. To this end, the established red-black strategy could be used. Thereby, the values for a particular voxel and its neighbors are updated in an alternating manner. We measured the computational time on a NVIDIA Tesla C2070 GPU for one iteration of both techniques – $0.015s$ for the primal-dual method and $0.027s$ for the linearized fixed-point iteration method.

Finally, we make some notes on the memory requirements of both numerical schemes. The primal-dual method requires maintaining five variables per voxel – the primal variable u, the three-dimensional dual variable ξ and the auxiliary variable $\bar{u}$ (ignoring the handling of a convergence criterion). The linearized fixed-point iteration method involves updating four variables per voxel – the relaxed implicit function u and three diffusivity values along the three coordinate axes. Note that the diffusivities do not need to be stored as they could be calculated on the fly. Yet, this would considerably increase the computational costs. We can conclude that in a straightforward implementation both techniques exhibit comparable memory requirements with the linearized fixed-point iteration approach being slightly more economical.

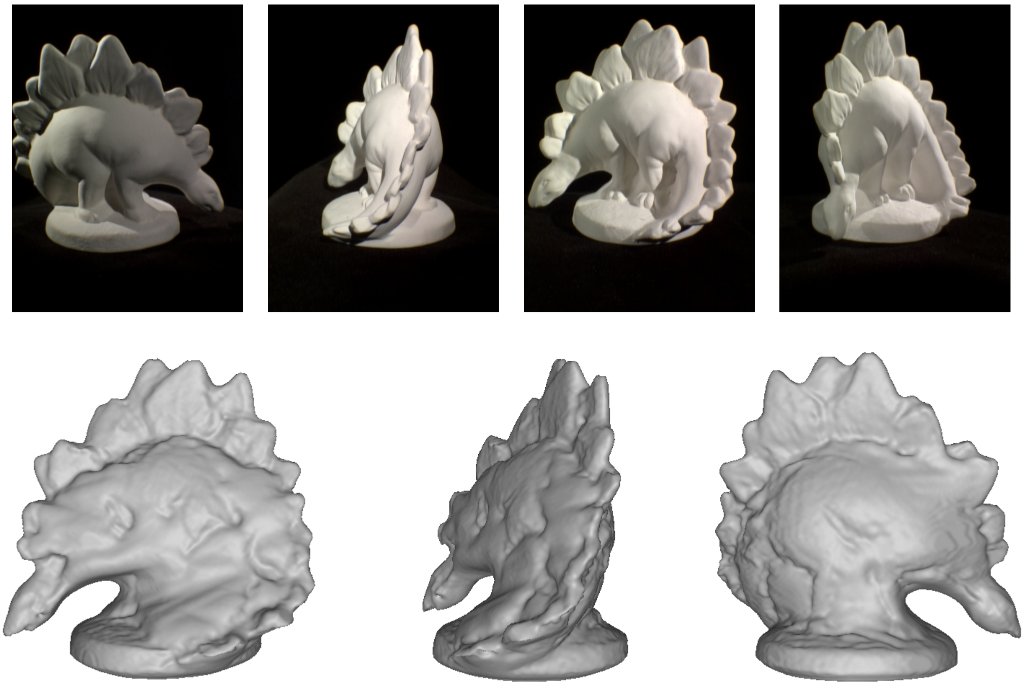
3.4 Experiments

In this section, we complement the comparison of the proposed energy models in Section 3.2 with an experimental validation. Additionally, we present an exhaustive experimental evaluation of the most promising model – energy model III.

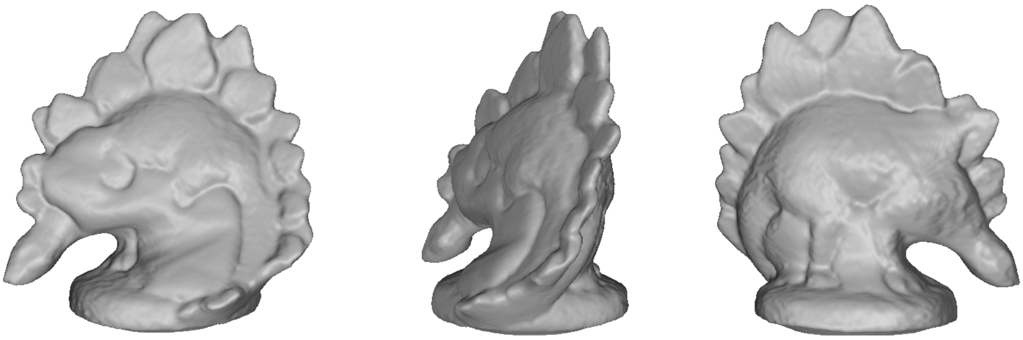
Experimental Comparison of the Three Energy Models

First, we provide a comparison between the presented energy models I, II and III (see Section 3.2). They are tested on the “dinoRing” and “templeRing” data sets which are part of the already mentioned Middlebury multiview stereo evaluation project [112]. The data sets contain 48/47 calibrated images of resolution 640×480 of a plaster dinosaur and a reproduction of a temple in Sicily. Both objects exhibit very different properties. While the dinosaur figurine is relatively smooth and weakly textured, the temple duplicate is well-textured, but of complex geometry in terms of small-scale details and sharp corners.

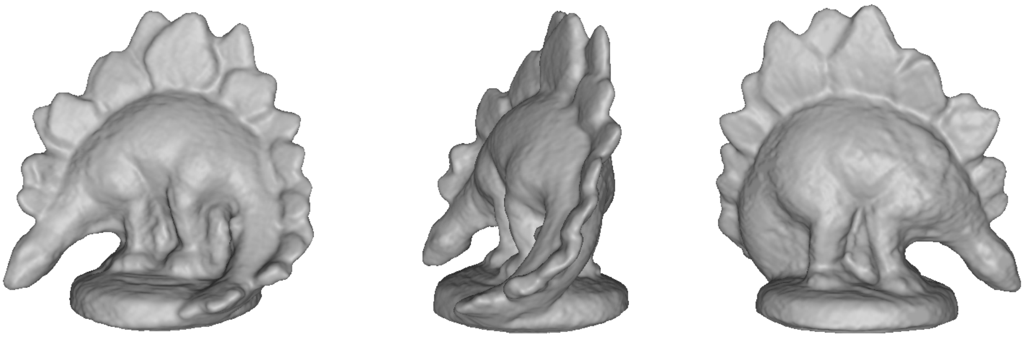
Fig. 3.6 shows reconstructions obtained with the three energy models on the first data set at volume resolution 256^3. The first two models clearly fail to recover the concavities (e. g. at the legs) due to the use of silhouette-based regional terms that act in contradiction to the stereo-based on-surface term (see Fig. 3.4). A decrease of the weighting parameter λ will not lead to the desired effect since this will also cut protruding parts (e. g. the spikes). In



reconstruction with energy model I



reconstruction with energy model II



reconstruction with energy model III

Fig. 3.6: Comparison of energy models I, II and III on the "dinoRing" data set. 4 out of 48 input images of resolution 640×480 and multiple views of the reconstructions obtained with the three energy models. Note that the first two models completely fail to recover deep concavities due to the limitations discussed in Section 3.2. In contrast, energy model III is able to retrieve accurately deep indentations as well as thin protrusions.

reconstruction with energy model I

reconstruction with energy model II

reconstruction with energy model III

Fig. 3.7: Comparison of energy models I, II and III on the "templeRing" data set. 4 out of 47 input images of resolution 640×480 and multiple views of the reconstructions obtained with the three energy models. Although the first model captures the deep concavity at the back, it produces a very noisy reconstruction. The second model successfully suppresses noise, but fails at locations of ambiguous texture. In contrast, the third model achieves the highest accuracy by recovering all large-scale details.

data set	**# images**	**completeness**	**accuracy**	**runtime**
dinoSparseRing	16	98.3 %	0.53 mm	1 h 2 min
dinoRing	48	99.4 %	0.43 mm	3 h 27 min
templeSparseRing	16	91.8 %	1.04 mm	1 h 4 min
templeRing	47	97.8 %	0.72 mm	4 h 12 min

Tab. 3.1: Quantitative evaluation of the reconstructions in Fig. 3.9 and respective runtimes.

this experiment, λ was chosen as the smallest value that retains all relevant surface details, but it is still insufficient to capture the concavities even with the improved photoconsistency estimation. Energy model III follows a purely stereo-based formulation which circumvents the mentioned shortcomings and produces a visibly more accurate reconstruction.

Analogously, Fig. 3.7 offers a comparison on the second data set. Here, the first energy model captures the deep indentation at the back, but the reconstruction is quite noisy and imprecise due to the noisy photoconsistency map (see Fig. 3.4). Although the second model produces a generally more accurate reconstruction, it completely fails in areas of weak or ambiguous texture (e. g. the wall at the back; see Fig. 3.4). In contrast, the third model achieves the highest accuracy by generating a smooth shape preserving all large-scale details.

The computational times of the three methods, which were measured for a single-core implementation on a 2.66 GHz Intel Core2 architecture, range from 15-20 minutes for the first model to more than 4 hours for the third one. Not surprisingly, the increased accuracy of the third model comes at the expense of increased computational efforts. Note that these runtimes can be reduced by a more conservative choice for the parameters α_{max} and γ_{max} and/or a GPU implementation. While a GPU implementation of the optimization is straightforward, a GPU realization of the described voting scheme is infeasible by using a parallelization on a voxel basis due to the high potential computational costs for a single thread.[4]

Analysis of Energy Model III

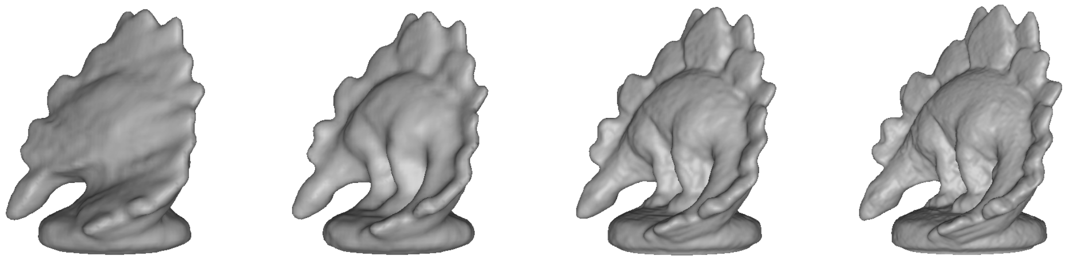
In this section, we give a detailed experimental evaluation of the most viable energy model – model III.

As mentioned in Section 3.2, a banded multiresolution scheme was applied in order to accelerate the computation of the data terms. Reconstructions at intermediate levels for the "dinoRing" data set are shown in Fig. 3.8. Additionally, the evolution of an initial surface towards the final result is depicted for the finest volume resolution of 256^3. Note that the final reconstruction does not depend on initialization due to the convexity of the arising optimization problem. In the visualized evolution process, a generic initialization was used – a box centered in the middle and with the same alignment as the delimitating bounding box. Moreover, we deployed the linearized fixed-point iteration method. Yet, the primal-dual method leads to a fairly similar minimization process with virtually the same result, but at a faster pace. A closer look reveals the difference to local optimization techniques like level sets [114]. While the surface always evolves coherently for level set methods, there are no such constraints for the applied convex optimization as structures can appear and fade freely.

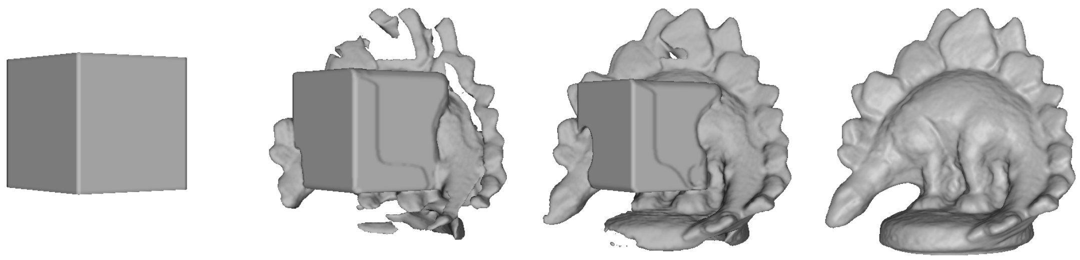
The aim of the next experiment, depicted in Fig. 3.9, is twofold – to present a quantitative evaluation of energy model III and to examine its dependence on the number of input images. As expected, the accuracy of the reconstructions improves when increasing the number of input

---

[4] Note also that the allocated register memory for each thread is quite limited and provides storage for only a couple of local variables.

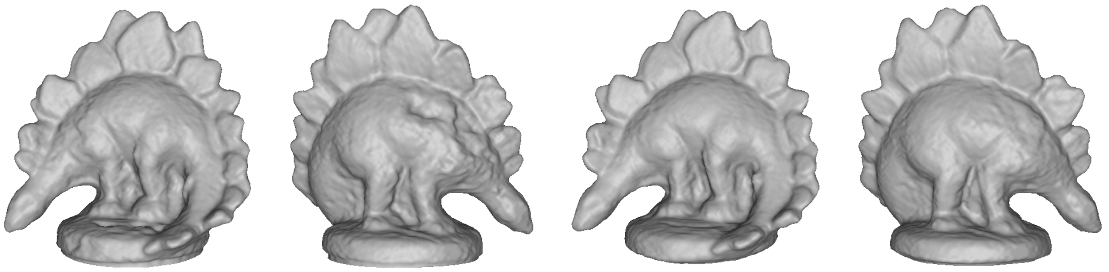


multiresolution scheme

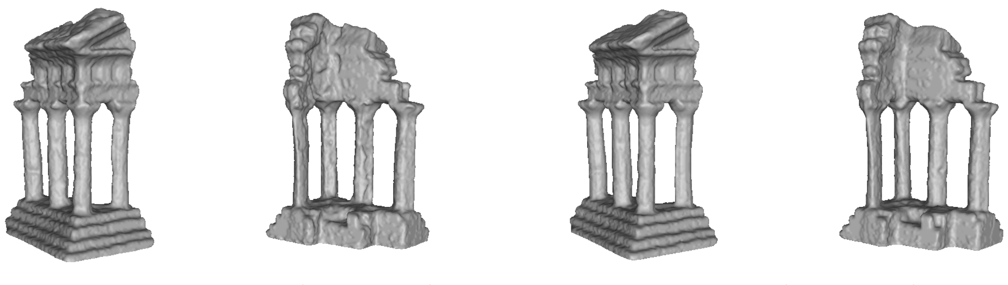


evolution process

Fig. 3.8: Surface evolution towards the final result. *First row:* Successively refined reconstructions with increasing resolutions of the volume: 64^3 (initialization and final result), 128^3 and 256^3. *Second row:* Surface evolution at the finest resolution obtained by thresholding the evolving function u at 0.5 (see Section 3.3). Note that in contrast to level set methods, the evolution process here is not coherent.



dinoSparseRing (16 images) **dinoRing (48 images)**



templeSparseRing (16 images) **templeRing (47 images)**

Fig. 3.9: Middlebury data sets. Reconstructions obtained with energy model III on the established Middlebury data sets – dinoSparseRing (16 images), dinoRing (48 images), templeSparseRing (16 images) and templeRing (47 images). Note how the accuracy of the reconstructions improves when increasing the number of input images, especially in areas of low visibility or poor texture.

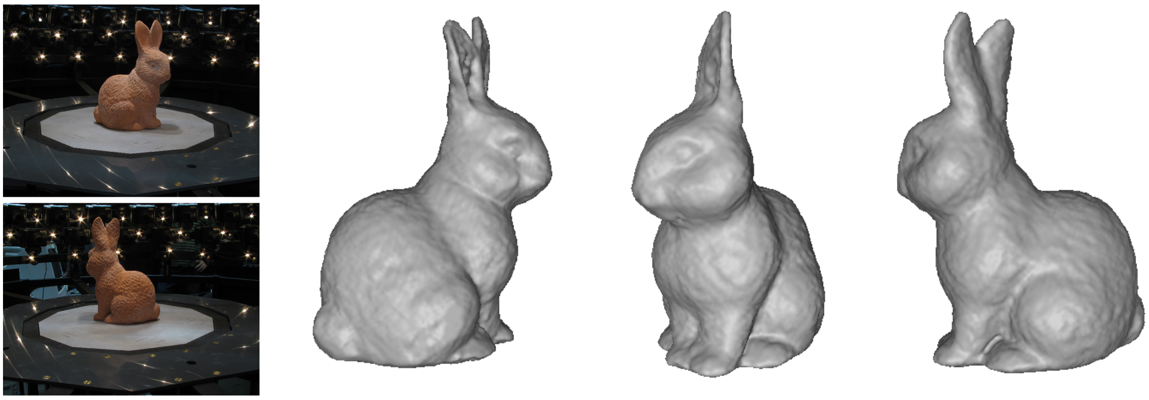


Fig. 3.10: Bunny sequence. 2 out of 33 input images of resolution 1024×768 and multiple views of the reconstructed 3D surface. The data set is particularly challenging due to the presence of homogeneous texture.

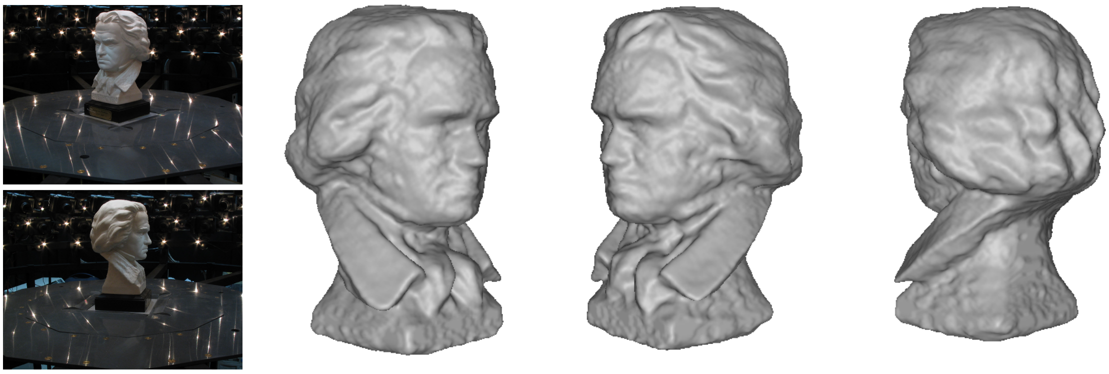


Fig. 3.11: Beethoven sequence. 2 out of 36 input images of resolution 1024×768 and multiple views of the reconstructed 3D surface. The data set is particularly challenging due to the absence of salient texture.

images. The most notable changes are in areas of low visibility (i. e. parts of the object visible in a reduced number of views due to occlusions) and poor texture. See for example the corner of the temple and the back of the dinosaur. Thus, increasing the number of input images seems to be quite helpful to resolve the encountered ambiguities. These observations are confirmed by respective quantitative evaluations shown in Table 3.1. Laser-scanned models of both objects are used as ground-truth in order to evaluate the quality of the reconstructions. The accuracy metric shown is the distance d (in millimeters) that brings 90% of the reconstructed surface within d from some point on the ground truth. The completeness score measures the percentage of points in the ground truth model that are within $1.25mm$ of the reconstructed model. The used volume resolution is 256^3 for "dino(Sparse)Ring" and $256 \times 384 \times 192$ for "temple(Sparse)Ring", respectively. Corresponding runtimes, measured with a single-core implementation on a 2.66 GHz Intel Core2 architecture, are also listed. As already mentioned, the obtained computational times could be reduced with a more conservative choice for the parameters α_{max} and γ_{max} and/or a (partial) GPU implementation.

Fig. 3.10 and 3.11 demonstrate the performance of the proposed approach on two challenging image sequences. The "bunny" sequence (see Fig. 3.10) is challenging due to the presence of homogeneous texture while the "Beethoven" sequence (see Fig. 3.11) is challenging due to the absence of salient texture. Despite these difficulties, which introduce ambiguities in

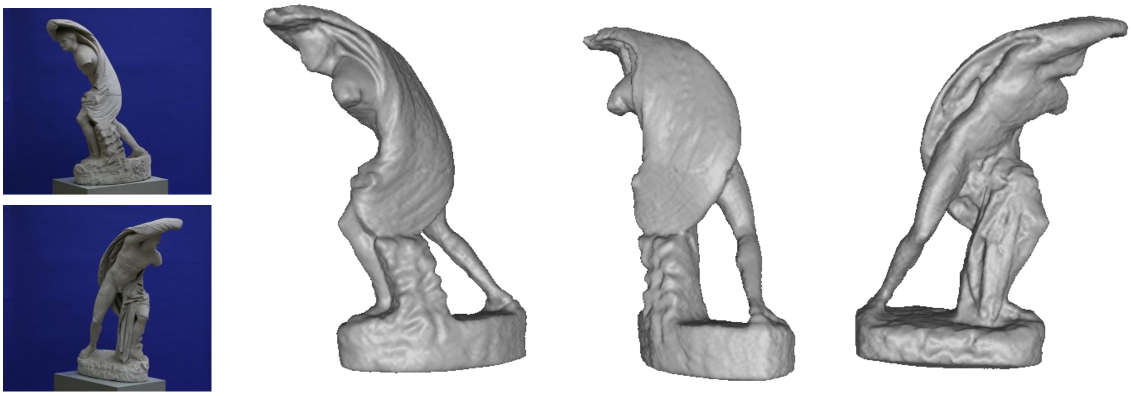


Fig. 3.12: Statue #1. 2 out of 38 input images of resolution 1800 × 1612 and multiple views of the reconstructed 3D surface.

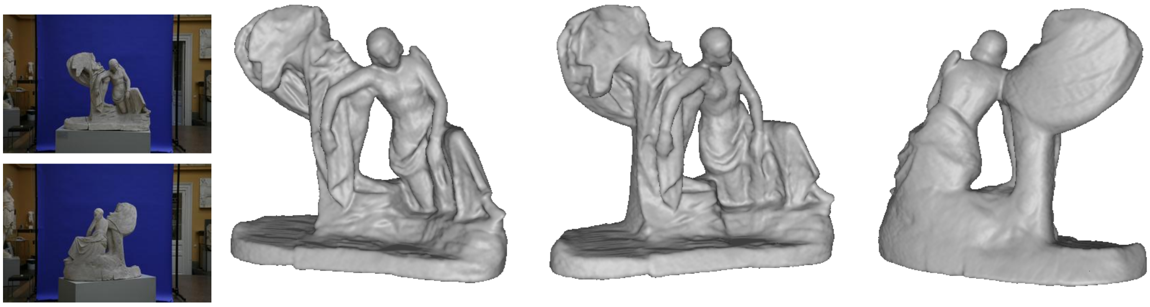


Fig. 3.13: Statue #2. 2 out of 36 input images of resolution 3072 × 2048 and multiple views of the reconstructed 3D surface.

Fig. 3.14: Statue #3. 2 out of 38 input images of resolution 2048 × 3072 and multiple views of the reconstructed 3D surface.

the matching process, the proposed approach produces 3D models of a genuinely acceptable quality. Even though some artifacts could be observed at the inner face of the ears of the bunny or at the cheek of the Beethoven figurine, the reconstructions exhibit all large-scale details and a high degree of smoothness.

In Fig. 3.12, 3.13 and 3.14, we show 3D reconstructions of three statues produced within the framework of a project with the Academic Art Museum in Bonn, Germany. All three

data set	# images	image resolution	volume resolution	runtime
bunny	33	1024×768	$216 \times 288 \times 324$	1 h 38 min
beethoven	36	1024×768	$240 \times 288 \times 360$	2 h 56 min
statue #1	38	1800×1612	$336 \times 456 \times 248$	6 h 26 min
statue #2	36	3072×2048	$300 \times 340 \times 320$	3 h 20 min
statue #3	38	2048×3072	$288 \times 480 \times 288$	6 h 21 min

Tab. 3.2: Experimental settings and runtimes for the reconstructions in Fig. 3.10, 3.11, 3.12, 3.13 and 3.14.

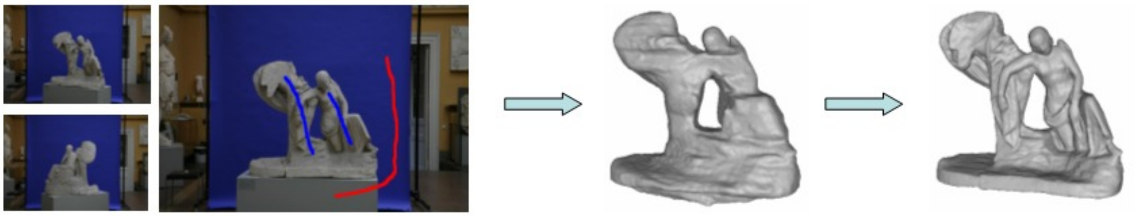


Fig. 3.15: 3D reconstruction pipeline. A complete reconstruction pipeline is obtained by applying the approach, described in Chapter 2, to compute a rough estimate and subsequently refining it with the proposed method. This allows to produce a high-quality 3D model of an object from a collection of calibrated images and a few scribbles marking foreground and background in one of them.

models are duplicates of ancient statues of historical value reproducing a particular event from the Greek mythology. Digitizing these works of art allows to rebuild their original configuration and to unveil the messages left by the sculptor [44]. All reconstructed 3D models are accurate and highly detailed and completely fulfill the quality requirements of the application. These experiments demonstrate the practical value of the proposed approach and image-based modeling in general.

Table 3.2 lists the experimental settings and runtimes for the reconstructions in Fig. 3.10, 3.11, 3.12, 3.13 and 3.14. We used sequences of 30-40 images of resolution ranging from below 1 Mpx up to 6 Mpx. Furthermore, the utilized volume resolutions vary between 20 and 40 million voxels. Although the computational times are relatively high, they could be reduced with some simple heuristics or by means of parallel computing as previously discussed. Moreover, these computational efforts are justified by the utilization of global optimization, which entails a considerable degree of robustness. There is one more point that should be clarified – the specification of the weighting parameter λ (see (3.4)). As we rely on a multiresolution scheme, the parameter should be set in accordance with the current spatial resolution. Increasing the resolution requires decreasing λ in order to retain the same amount of regularization. This issue is discussed in Chapter 5 in more detail. In our implementation, we simply set $\lambda = 1$ at the coarsest level and $\lambda = 0.5$ at any subsequent level.

Finally, we emphasize that a complete 3D reconstruction pipeline is obtained by applying the approach, described in Chapter 2, to compute a rough estimate and subsequently refining it with the proposed method (see Fig. 3.15). This allows to produce a high-quality 3D model of an object from a collection of calibrated images and a few scribbles marking foreground and background in one of them. Note the crucial difference of this image-based modeling pipeline to active laser-scanning techniques. It reveals the main advantage of such technologies – their flexibility and broad range of applicability.

3.5 Discussion

In this chapter, we showed that multiview stereo can be cast as a continuous convex optimization problem. In particular, we considered three different energy models amenable to a convex formulation. While they share the same variational template, they differ in the particular definition of data terms. The first two energy models use silhouette information to derive regional subdivision and classify points in space as being inside or outside the observed object. In contrast, the third model relies exclusively on stereoscopic information to define all involved data terms. In a comparative study, we demonstrated the performance of the three approaches. As the third one exhibits the most complex structure, it offers the highest accuracy, but at the expense of substantial computational efforts. In both qualitative and quantitative experiments, we further explored the capabilities of this energy model. Moreover, we focused on the arising convex optimization problem and discussed two numerical schemes for solving it. While both of them are fast and stable, and lend themselves to the particular problem at hand, they manifest different properties in terms of runtime, potential for parallel computing and memory requirements.

4

Integration of Multiview Stereo and Silhouettes

Science may set limits to knowledge,
but should not set limits to imagination.

Bertrand Russell (1872-1970)

In the previous two chapters, we considered convex formulations for 3D reconstruction from object outlines and stereo information. In this chapter, we unify silhouettes and multiview stereo in a single framework, i. e. we tackle the problem of finding the most photoconsistent surface that exactly fulfills silhouette constraints.

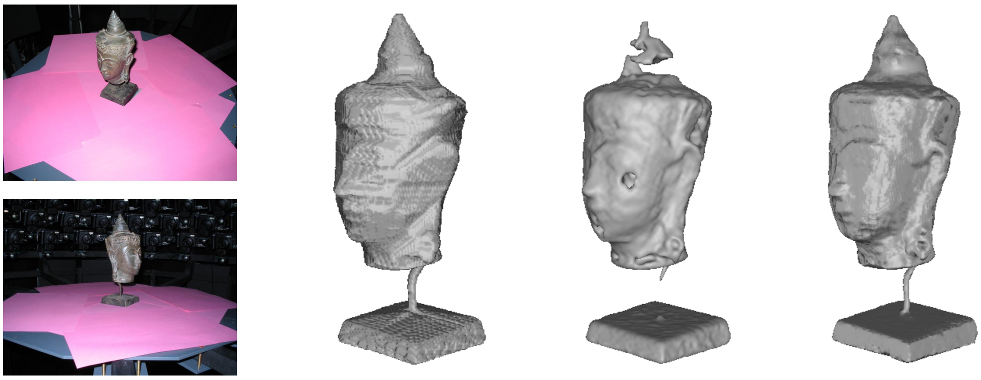
4.1 Introduction

Motivation

So far, we have become acquainted with the silhouette and the stereo cue as powerful tools for multiview 3D reconstruction. Yet, both of them have their strengths and weaknesses. Silhouette-based methods offer considerable robustness, but suffer from limited accuracy, mainly due to the inability to capture concavities. On the other hand, the accuracy of multiview stereo approaches does not depend on the particular geometry of the imaged object, but strongly relies on the Lambertian assumption. As a consequence, there are practical cases, where both paradigms fail to produce a 3D model of acceptable quality, like the one depicted in Fig. 4.1. The image sequence captures a metallic head statue which poses a challenge for silhouette-based methods, due to the presence of indentations, as well as stereo-based ones, due to severe specular reflections and the presence of thin geometric structures like the pedestal. Expectedly, none of the two modalities for image-based modeling is able to deliver an accurate reconstruction. This creates the demand for techniques that combine the advantages of silhouette- and stereo-based methods and promise better prospects of success. Ideally, silhouettes should prevail at fine-scale protrusions and specularities, and stereo information – at surface indentations (see the rightmost reconstruction in Fig. 4.1). To this end, the problem can be formulated as finding a 3D shape that is both photo- and silhouette-consistent.

Previous Work

The first strategy to fuse complementary features like multiview stereo and silhouettes was to use a visual hull (computed from outlines) as initialization for a stereo-based approach by restricting the solution to lie within the specified area [92, 134]. In other words, the following



two out of 33 input images **silhouette-based reconstruction** **stereo-based reconstruction** **silhouettes** & stereo

Fig. 4.1: Silhouette and stereo integration. While silhouette-based methods cannot reconstruct indentations of the surface, since these do not appear in the silhouettes, stereo-based reconstruction methods tend to remove thin structures and are heavily affected by specular reflections such as those of metal objects. In contrast, approaches integrating silhouette and stereo information combine the advantages of both modalities and allow to restore concave areas (around the ears) as well as fine geometric details (such as the pedestal).

cost functional should be minimized

$$\begin{aligned} &E(S) = \int_S \rho(s)\,ds \\ &\text{s. t.}\;\; S \subset VH(Sil_1,\dots,Sil_n), \end{aligned} \tag{4.1}$$

where $VH(Sil_1,\dots,Sil_n) \subset V$ denotes the visual hull computed from the given set of silhouettes and $\rho : V \to [0,1]$ is a photoconsistency map. Note that the constraint in (4.1) can equivalently be formulated as $\pi_i(S) \subset Sil_i \;\forall\, i = 1,\dots,n$. Following the discussion in Chapter 3, one can observe that the above energy model does not preclude the trivial solution – the empty surface. This difficulty was circumvented by using a greedy procedure to find the most photoconsistent shape [92] or by heuristically restricting the search space [134]. Yet, both of these strategies are suboptimal and could produce inaccuracies in the geometry. Apart from these considerations, a grave limitation of the above methodology is that it does not guarantee exact silhouette consistency of the estimated 3D surface.

An alternative technique is to unify both information sources in a single variational formulation. A straightforward approach envisages the minimization of a functional of the form

$$E(S) = \int_S \rho(s)\,ds \;+\; w \sum_{i=1}^{n} \left(\, |\pi_i(S) \backslash (\pi_i(S) \cap Sil_i)| \;+\; |Sil_i \backslash (\pi_i(S) \cap Sil_i)| \,\right), \tag{4.2}$$

where $w \in \mathbb{R}$ is a weighting parameter and $|C|$ signifies the cardinality of a set C. This optimization problem was addressed via various local procedures [26, 33, 40]. Finding a global solution is a genuinely challenging task due to presence of a myriad of local minima. For that reason, the proposed local optimization schemes may lead to a numerically unstable behavior and introduce a bias near the visual hull boundary. Moreover, the final result is highly dependent on initialization. The parameter w governs the degree of silhouette consistency of the obtained surface and balances between the empty solution and a purely silhouette-based estimate.

Some researchers advocated utilizing predetermined surface points to impose silhouette consistency [56, 124, 115]. This can be formulated as minimizing the following cost functional

$$\begin{aligned} E(S) &= \int_S \rho(s)\, ds \\ \text{s. t. } \; & p_j \in int(S) \quad j = 1, \ldots, M, \end{aligned} \tag{4.3}$$

where $p_1, \ldots, p_M \in V$ is a set of points or regions in 3D space and $int(S) \subset V$ denotes the surface interior. Various heuristics was proposed to generate reliable 3D points, usually involving photometric or contour-based criteria. The main drawback of this approach is that it requires premature decisions about voxel occupancy which could lead to spurious artifacts in the reconstructions. Moreover, in general, exact silhouette consistency is hard to achieve with this method.

A mathematically elegant alternative to fuse silhouette and stereo information is stated by the stereoscopic segmentation model [142, 43, 28]. Thereby, appropriate evolution terms enforcing photoconsistency and silhouette consistency criteria are derived by means of the derivative of the reprojection error of the estimated shape. Specifically, the method involves minimizing

$$E(S) = \sum_{i=1}^{n} \int_{\Omega_i} |I_i(z) - T(\pi_{i,S}^{-1}(z))|^2 \, dz, \tag{4.4}$$

where $T : S(\Theta) \subset V \to \mathbb{R}^3$ denotes a texture map calculated from the input images. $\pi_{i,S}^{-1}$ signifies the back-projection mapping of camera i with respect to surface S, i. e. to each pixel the intersection point of the respective viewing ray with the surface is assigned. The energy model in (4.4) is appealing due to the fact that it captures different aspects of the reconstruction process, in particular photometric criteria, silhouette consistency and explicit visibility estimation. However, due to its complexity, it is difficult to overcome local optimization which is prone to undesired local minima. [1]

To address this difficulty and come apart from local evolution schemes, [116] proposed a graph cut framework for silhouette and stereo fusion. The key idea is to incorporate exact silhouette constraints in the construction of the graph while encouraging the cut to follow photoconsistent locations. Unfortunately, the practical applicability of this method is limited due to its high memory requirements. This poses a severe restriction on the volume resolution at which reconstructions can be computed. Additional drawbacks result from the discrete nature of the formulation (see Section 1.4).

Contribution

We propose a convex mathematically transparent framework for silhouette and stereo fusion. The idea is to cast multiview 3D reconstruction as a constrained variational problem, where the minimized functional encodes photometric criteria and *exact* silhouette consistency is imposed by means of convex constraints that restrict the domain of feasible functions. A silhouette-consistent reconstruction is computed by convex relaxation, finding a global minimizer of the relaxed problem and subsequent projection onto the original non-convex set. Due to the inherent convex formulation, it can be shown that the obtained solution of the original problem lies within an energetic bound of the optimal one. In fact, it is the first approach

---

[1] An example of a undesired local minimum is the case of a torus, where the initialization consists of a disk enclosing it. Since local deformations (indentation) of the disk only increase the overall surface area without decreasing th reprojection error, the local minimization procedure will not give rise to a torus but rather remain stuck with the disk - see [66] for an experimental validation.

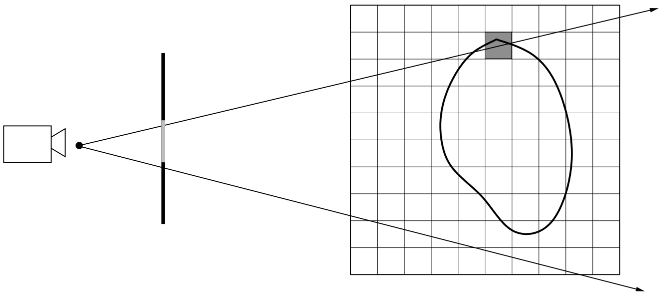


Fig. 4.2: Schematic views of silhouette consistency. A 2D visualization of the object surface and its projection onto the image shows that for a silhouette-consistent shape at least one voxel along each visual ray through a silhouette pixel is occupied, whereas all voxels along rays through non-silhouette pixels are empty. The bright area on the image plane indicates the outlines of the observed object and the shaded voxel is an occupied one along the given viewing ray.

for silhouette and stereo integration that entails any globality guarantees. Compared to established fusion techniques, the proposed formulation does not depend on initialization and leads to a more tractable numerical scheme by removing the bias near the visual hull boundary. All practical benefits are demonstrated on experiments with several challenging real-world data sets including metallic objects, low-texture objects and complex objects with fine-scale details. Moreover, we examine the suitability of two different numerical schemes for solving the arising optimization problem.
The main results in this chapter are published in [68, 24].

4.2 Convex Integration of Silhouettes and Stereo

We revisit the classical weighted minimal surface model for multiview stereovision introduced in Chapter 3

$$E(S) = \int_S \rho(s)\, ds. \tag{4.5}$$

As previously discussed, global minimization of (4.5) is meaningless without any additional constraints as this leads to the empty surface as a solution. Moreover, we saw that given silhouettes could be deployed to further specify the observed object and avoid the trivial result. Yet, the integration of photoconsistency and silhouettes is not an easy task due to the conceptual difference of both cues.

Introducing Silhouette Constraints

The basic idea of the proposed formulation is to impose exact silhouette alignment of the computed shape during the energy minimization process. This naturally prevents the collapse of the surface, since the empty set is clearly not a silhouette-consistent solution.
More specifically, we solve the following constrained optimization problem

$$\begin{aligned} &\min_S \int_S \rho(s)\, ds, \\ \text{s. t.} \qquad &\pi_i(S) = Sil_i \quad \forall\, i = 1, \dots, n. \end{aligned} \tag{4.6}$$

Unfortunately, the above minimization is highly non-convex: Depending on the values of the photoconsistency function ρ, variations in S may give rise to arbitrary increases or decreases in the cost functional. Furthermore, imposing silhouette consistency in the above representation is not straightforward as a small surface deformation may easily violate the constraints.

Convex Relaxation

Surprisingly, these difficulties can be removed by reverting to an implicit representation of the surface S. In order to cast (4.6) as a convex optimization problem, the surface S is again represented by the characteristic function $u : V \rightarrow \{0, 1\}$ of its interior $int(S)$. Hence, changes in the topology of S are handled automatically without reparametrization.

With this implicit surface representation, the optimization problem (4.6) in S is equivalent to the following optimization problem in the binary labeling u

$$\min_{u \in D'} \int_V \rho(x)\, |\nabla u(x)|\, dx, \tag{4.7}$$

where

$$D' = \left\{ u : V \rightarrow \{0,1\} \,\middle|\, \begin{array}{l} \int\limits_{R_{ij}} u(r)\, dr \geq \delta \ \text{ if } j \in Sil_i \ \forall i, j \\ \int\limits_{R_{ij}} u(r)\, dr = 0 \ \text{ if } j \notin Sil_i \ \forall i, j \end{array} \right\}. \tag{4.8}$$

Thereby, R_{ij} denotes the visual ray through pixel j of image i. In this implicit formulation, the silhouette consistency constraint in (4.6) gives rise to equality and inequality constraints. These constraints are based on the following reasoning: For all visual rays from the camera center passing through a pixel inside a silhouette *at least one* of the voxels along this ray should be occupied, whereas for a visual ray passing through a pixel outside a silhouette *all* voxels along the respective ray should be empty (see Fig. 4.2). In the spatially continuous formulation adopted in this paper (where there exists no notion of discrete *voxels*), the constant δ in (4.8) denotes the thickness of material below which the considered object becomes translucent. It should be noted that this material-dependent parameter δ gives rise to a more realistic physical model than imposing

$$\int\limits_{R_{ij}} u(r)\, dr > 0 \text{ if } j \in Sil_i \ \forall\, i, j.$$

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

There are a few important remarks regarding the formulation in (4.7) that should be made. In case of imperfect silhouettes, the above constraints can be applied only to those pixels for which we have certain confidence about being inside or outside a silhouette. Note that this property is in contrast to most of the existing approaches which require apparent contours of the imaged object to be provided. In a discrete setting, the volume resolution should be determined according to the given image resolution. When the respective volume sampling is too coarse, some discretization artifacts could appear in the reconstruction. Unfortunately, increasing the voxel resolution entails a considerable computational and memory burden and is not always possible in practice. In such cases subsampling the input silhouettes could be helpful. Note that the constraints in (4.7) could give an empty set. This can happen when all voxels along a visual ray passing through a silhouette pixel in one of the images project to background in another image. Although this is unlikely to happen in a real scenario, it can occur in case of noisy silhouette input. Such difficulties can be circumvented by restricting the computations to the visual hull of the object and holding the values of u for all other voxels fixed.

Now, let us concentrate on the optimization of the energy model in (4.7). Interestingly, the functional is convex. Yet, due to the constraint that u is a binary-valued function, the overall minimization problem (4.7) is non-convex (because the space of binary functions is non-convex). This can be resolved by relaxing the binary constraint and allowing the function u to take on values in the interval $[0,1]$. The relaxed problem therefore becomes that of minimizing a convex functional over a convex set

$$\min_{u \in D} \int_V \rho(x) \, |\nabla u(x)| \, dx, \tag{4.9}$$

where

$$D = \left\{ u : V \rightarrow [0,1] \;\middle|\; \begin{array}{l} \int\limits_{R_{ij}} u(r)\, dr \geq 1 \text{ if } j \in Sil_i \; \forall i,j \\ \int\limits_{R_{ij}} u(r)\, dr = 0 \text{ if } j \notin Sil_i \; \forall i,j \end{array} \right\} \tag{4.10}$$

is the set of continuous valued functions u which are silhouette-consistent with respect to all images i and all rays j. Again, the corresponding constraints follow the formulation in (4.8). Surprisingly, in this implicit representation of the shape S by a relaxed labeling function $u : V \longrightarrow [0,1]$, the set D of silhouette-consistent configurations is convex.

Proposition 3. *The set D of all silhouette-consistent functions defined in* (4.10) *forms a convex set.*

Proof. As D is defined by a set of inequality constraints, the claim could be proven by generalizing Corollary 1 to a continuous setting. Yet, in the following, we expose a direct proof. Let $u_1, u_2 \in D$ be two elements of D. Then any convex combination $u = \alpha u_1 + (1-\alpha) u_2$ with $\alpha \in [0,1]$ is also an element in D. In particular, $u(x) \in [0,1]$ for all x. Moreover,

$$\int\limits_{R_{ij}} u \, dr = \alpha \int\limits_{R_{ij}} u_1 \, dr + (1-\alpha) \int\limits_{R_{ij}} u_2 \, dr \geq 1 \;\text{ if } j \in Sil_i,$$

and similarly

$$\int\limits_{R_{ij}} u \, dr = \alpha \int\limits_{R_{ij}} u_1 \, dr + (1-\alpha) \int\limits_{R_{ij}} u_2 \, dr = 0 \;\text{ if } j \notin Sil_i.$$

Thus $u \in D$. □

The above statement implies that (4.9) poses a constrained convex optimization problem which can be solved by means of any of the methods presented in Section 1.3. In Section 4.4, we discuss the numerical minimization of the model. In particular, we adapt two techniques, already encountered in Chapter 2 and 3 – linearized fixed-point iteration and primal-dual method. While we compared both procedures on a linear model combining regional terms and on-surface photoconsistency (fitting into the general form (1.5)), here we complete the evaluation by testing them on the weighted minimal surface model in (4.6). Surprisingly, although the functional in (4.6) can be considered as a special case of (1.5) (for $\lambda = 0$), the two numerical schemes demonstrate a genuinely different behavior.

Binary Solution via Thresholding

Since we are interested in minimizers of the non-convex binary labeling problem (4.7), a straightforward methodology is to threshold the solution of the convex problem appropriately. Although this does not guarantee finding the global minimum of (4.7), the proposed strategy entails a series of advantages compared to classical local optimization techniques. Intuitively, for smooth functionals extending the set of feasible functions, computing the global minimum over this domain and subsequently projecting to the nearest point within the original set is expected to give a more accurate estimate than a simple gradient descent procedure. In particular, the proposed methodology has the following advantages:

- It allows to incorporate *exact* silhouette constraints without making premature hard decisions about voxel occupancy along each viewing ray passing through a silhouette pixel.
- It does not depend on initialization since the relaxed functional is optimized globally.
- It leads to a simple and tractable numerical scheme which does not rely on a locally estimated surface orientation and thus does not introduce a bias near the visual hull boundary.
- The algorithm provides a solution to the binary optimization problem (4.7) which lies within an energetic bound of the optimal solution (sometimes called ϵ-minimizer in the literature). This is stated in the following proposition.

Proposition 4. *Let E denote the functional in* (4.9) *and u^* be a minimizer. Let $D' \subset D$ be the set of binary silhouette-consistent functions defined in* (4.8). *Furthermore, let $u' \in D'$ be the (global) minimum of* (4.7) *and $\tilde{u}$ the solution obtained with the above procedure. Then the computed solution lies within an energetic bound of the optimum.*

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

$$E(\tilde{u}) - E(u') \leqq E(\tilde{u}) - E(u^\star),$$

where the right hand side can easily be computed. □

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

$$\tilde{u}(x) = \begin{cases} 1, & \text{if } u^*(x) \geq \mu \\ 0, & \text{otherwise} \end{cases}, \tag{4.11}$$

where

$$\mu = \min\left\{ \left(\min_{i \in \{1,\dots,n\}, j \in Sil_i} \ \max_{x \in R_{ij}} u^*(x) \right) \ , \ 0.5 \right\}. \tag{4.12}$$

This threshold μ provides the closest (in any L^p-norm) silhouette-consistent binary function to the solution of the relaxed problem. Note that, in the absence of any additional constraints, 0.5 would be the threshold that realizes the projection. If additional constraints are available and the constant 0.5 does not lead to a feasible solution, the largest possible value (and closest to 0.5) possesses the desired property. [2]

---

[2] It should be noted that the above thresholding could alternatively be realized by computing the binary labeling function $\tilde{u}$ of lowest energy instead of the closest one. We tested this procedure and found out that it typically gives a threshold very close to the one specified in (4.12). As a result, both thresholds give rise to visually indistinguishable reconstructions while the proposed alternative is much faster to estimate.

Proposition 5. *The reconstructed surface exactly fulfills all silhouette constraints, i.e.* $\tilde{u} \in D'$.

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

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

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

In practice, thresholding by μ can be replaced by subtracting μ from the values of u^* and extracting the 0-level. This methodology has the advantage that it produces smoother surfaces than the thresholding procedure without affecting the energy which is defined on a voxel basis.

4.3 Imposing Silhouette Consistency

The proposed approach boils down to solving the constrained convex optimization problem (4.9). In Section 1.3, we considered different numerical techniques for accomplishing such tasks. Some of them, in particular the gradient descent method and the primal-dual method, involve permanently projecting the current estimate onto the convex domain. Thus, the efficient realization of this projection is of paramount importance. In this paragraph, we address the numerical implementation of the projection operator for the specified set.

A Set of Linear Constraints

One can observe that the set D in (4.10) is formed by three types of linear equality and inequality constraints:

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

$$\int_{R_{ij}} u(r)\,dr \;=\; 0 \text{ if } j \notin Sil_i. \tag{4.13}$$

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

$$\int_{R_{ij}} u(r)\,dr \geq 1 \text{ if } j \in Sil_i. \tag{4.14}$$

A Simple Iterative Projection Scheme

Individually, each type of constraint can be easily imposed as follows. The first constraint is enforced by simply clipping the values of u to the interval $[0,1]$ for all x. The second type of constraint in (4.13) is imposed by simply setting all values of u along the respective ray Sil_i to zero. This concerns all voxels outside the visual hull of the imaged object. The third

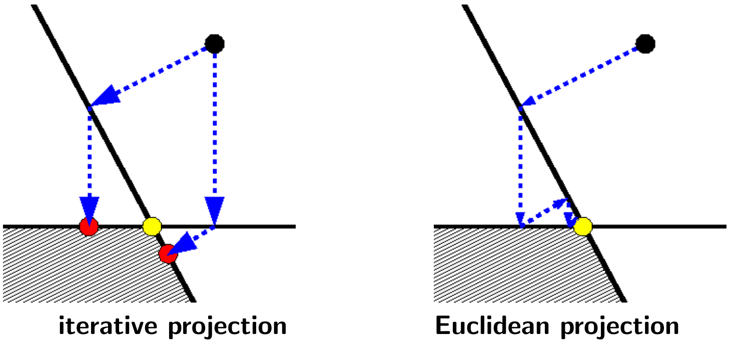


Fig. 4.3: Schematic view of projections of an initial configuration (black dot) onto two non-orthogonal linear constraints. While the sequential projection onto each individual hyperplane (left) leads to configurations (red dots) within the set of feasible solutions (shaded area), these configurations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

type of constraint in (4.14) requires more effort. Assume that it is violated for some pixel j in silhouette Sil_i, i. e. there exists some residual $\epsilon > 0$ such that

$$\int_{R_{ij}} u\, dr = 1 - \epsilon. \tag{4.15}$$

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

$$u_{new}(x) = u(x) + \frac{\epsilon}{\int_{R_{ij}} dr} \qquad \forall\, x \in R_{ij}. \tag{4.16}$$

Note that the above procedure realizes projection onto the hyperplane associated with the respective constraint (with respect to any L^p-norm). Iterating the individual projections can be done quite efficiently. It leads to a configuration which is guaranteed to fulfill all constraints.

However, enforcing these constraints in a different order will generally produce a different result. Since the constraints defined above are generally not orthogonal, sequentially projecting onto each respective hyperplane will generally not give rise to the Euclidean projection onto the convex domain. It will lead to a feasible solution but typically not the closest one among all. This can be seen in the schematic drawing of Figure 4.3, left side: By sequentially projecting from an initial configuration (black dot) onto two non-orthogonal hyperplanes, one ends up with a solution (red dot) in the feasible set that is not the closest one to the input configuration (yellow dot). In our scenario, this will happen if the same voxel lies on two rays that both violate the constraint. Depending on which of the two constraints is imposed first, we may obtain a different solution.

In practice, we find that when the evolution step is small enough, this issue is not crucial and the simple sequential projection often produces acceptable results. In order to avoid computations of ray-volume intersections any time the constraint is checked, one can compute the set of relevant voxels to each viewing ray in a preprocessing step and store them in lists. However, the size of this data structure could grow significantly when the resolution of input images is high. For this reason, in our implementation, we stored only the first and last voxel along each ray. Another important issue when using constraints is the frequency of enforcing them. In our implementation, we achieved a stable behavior when applying the first constraint after each optimization iteration and the silhouette constraints after each 10 iterations. Nevertheless, one may ask if there is a more elegant solution to impose the constraints since enforcing each constraint sequentially as above not only depends on the order of projections, but it is also likely to shrink the current solution more than necessary.

Euclidean Projection onto the Convex Set

Ideally, one would like to compute the Euclidean projection onto the convex set of feasible solutions such that all constraints are taken into account simultaneously. Since there is a huge number of them, a central challenge is to find an efficient algorithm to do this. Fortunately, there exist such algorithms in the literature. In this work, we make use of a method that was first published by Dykstra and Boyle [32, 14]. By alternating projections onto each convex set in a recursive manner, this approach allows to compute the Euclidean projection onto the intersection of finite number of closed convex sets.

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

$$\begin{aligned}
\vec{u}_0^k &= \vec{u}_p^{k-1}, \\
\vec{u}_i^k &= \Pi_i(\vec{u}_{i-1}^k - \vec{v}_i^{k-1}), \quad i = 1, 2, \dots, p, \\
\vec{v}_i^k &= \vec{u}_i^k - (\vec{u}_{i-1}^k - \vec{v}_i^{k-1}), \quad i = 1, 2, \dots, p,
\end{aligned} \tag{4.17}$$

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

Similar to the simple procedure described in Section 4.3, this recursive algorithm also applies the individual projections Π_i sequentially. Yet, it systematically separates the residuals associated with various projection directions so as to make sure that the computed estimates $\vec{u}_p^k$ gradually approach the Euclidean projection onto the intersection of half-spaces.

Proposition 6. *The algorithm defined in* (4.17) *converges to the Euclidean projection onto the intersection of convex sets.*

Proof. For the proof we refer to [32, 14]. □

Figure 4.3, right side, shows a visualization of this recursive projection method for the case of two non-orthogonal constraints.

It can be noted that the dimensionality of all involved variables in (4.17) is equal to the number of voxels and is usually in the order of multiple million. Hence, some care should be taken when implementing the above procedure to reduce the memory requirements. First, one can observe that the updates in $\vec{u}_i^k$ only involve the previously projected values $\vec{u}_{i-1}^k$ and thus can be evaluated on the fly. Moreover, the auxiliary vectors $\vec{v}_i^k$ always have the direction of the normals to the corresponding hyperplanes. Hence, only their lengths have to be stored instead of the full configurations.

---

[3] Recall that they can be realized by the simple procedure in (4.16).

4.4 Implementation

This section gives more details on the implementation of the proposed approach. In particular, the estimation of photoconsistency, involved in (4.9), as well as the numerical optimization of (4.9) are discussed.

Photoconsistency Estimation

Obviously, the method is independent of the particular computation of the photoconsistency map $\rho : V \to [0, 1]$. In the presented implementation, the voting scheme, described in Chapter 3, was used because of its high precision. See Chapter 3 for more details.

Numerical Optimization

In this paragraph, we focus on the numerical optimization of the constrained convex optimization problem in (4.9). To this end, we adopt two procedures, already encountered in Chapter 3 – linearized fixed-point iteration and primal-dual method.

Linearization and Fixed-Point Iteration

The basic idea of the linearized fixed-point iteration method, introduced in Chapter 3, is to solve the respective Euler-Lagrange equation directly by linearizing the non-linear part and derive a linear system of equations. Regarding the functional in (4.9), the Euler-Lagrange equation reads

$$0 = \operatorname{div}\left(\rho\, \frac{\nabla u}{|\nabla u|}\right) = \rho \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \langle \nabla \rho, \frac{\nabla u}{|\nabla u|}\rangle. \tag{4.18}$$

It can be observed that the only source of non-linearity in (4.18) is the diffusivity $g = \frac{\rho}{|\nabla u|}$. Starting with an initialization u^0, we can compute g and keep it constant. For constant g (4.18) is linear and discretization yields a sparse linear system of equations of the form $A\vec{u} = \vec{0}$ which we solve with the SOR method. More specifically, we iteratively compute an update of u at voxel i by

$$u_i^{(l,k+1)} = (1-\omega)u_i^{(l,k)} + \omega \frac{\sum\limits_{j\in\mathcal{N}(i), j<i} g^l_{i\sim j} u_j^{(l,k+1)} + \sum\limits_{j\in\mathcal{N}(i), j>i} g^l_{i\sim j} u_j^{(l,k)}}{\sum\limits_{j\in\mathcal{N}(i)} g^l_{i\sim j}}, \tag{4.19}$$

where $\mathcal{N}(i)$ denotes the 6-neighborhood of i. Finally, $g_{i\sim j}$ denotes the diffusivity between voxel i and its neighbor j. It is defined as

$$g^l_{i\sim j} = \frac{g^l_i + g^l_j}{2}, \qquad g^l_i = \frac{\rho_i}{\sqrt{|\nabla u^l_i|^2 + \epsilon^2}}, \tag{4.20}$$

where $\epsilon = 0.001$ is a small constant that prevents the diffusivity to become infinite when $|\nabla u^l_i|^2 = 0$ and $|\nabla u^l_i|^2$ is approximated by standard central differences. The over-relaxation parameter ω has to be chosen in the interval $(0, 2)$ for the method to converge. The optimal value depends on the linear system to be solved. Empirically, for the specific problem at hand, we obtained a stable behavior for $\omega = 1.3$. After the linear solver yields a sufficiently good approximation (we iterate for $k = 1, ..., 10$), one can update the diffusivities and solve the next linear system. Empirically, we observed that the particular choice of a stopping criterion

is not a crucial issue in practice. We obtained virtually the same results by measuring the energy decay, the evolution step and the gap of the dual energy (see Theorem 3).
However, some more care should be taken in this case since here we are confronted with a constrained optimization problem. In Chapter 3, we saw that the linearized fixed-point iteration method can be interpreted as a specific quasi-Newton method of the form

$$\vec{u}^{t+1} = \vec{u}^t - \omega(B^t)^{-1}\nabla E(\vec{u}^t), \tag{4.21}$$

where $E(.)$ denotes the functional in (4.9), $\vec{u} \in \mathbb{R}^N$ summarizes the values of u on the N-dimensional discretized volume grid $\tilde{V}$ and $B^t \in \mathbb{R}^{N\times N}$ is defined as $B^t = (b^t_{ik})_{i,k=1,\dots,N}$ with

$$b^t_{ik} = \begin{cases} \sum_{j\in\mathcal{N}(i)} g^t_{i\sim j} & \text{if} \quad i = k \\ 0 & \text{otherwise} \end{cases} . \tag{4.22}$$

Moreover, the gradient $\nabla E(\vec{u}) \in \mathbb{R}^N$ is given by

$$(\nabla E(\vec{u}))_i = -\text{div}\left(\rho\,\frac{\nabla u_i}{|\nabla u_i|}\right) \tag{4.23}$$

for all voxels $i = 1,\dots,N$. Reverting to Section 1.3, we recall that a generalization of a quasi-Newton method to the constrained case could be done by solving a quadratic problem to estimate the evolution direction in each step. Yet, the involved computational costs would be prohibitive. For that reason, we propose a different strategy. One can observe that the evolution scheme in (4.21) boils down to gradient descent if we set $B^t = I$, where $I \in \mathbb{R}^{N\times N}$ denotes the identity matrix. The gradient descent method can easily be generalized to the constrained case by simple back-projection onto the feasible set (see Section 1.3). Hence, we can construct the following strategy for solving (4.9): We retain the iterative procedure in (4.19) and impose the constraints after each iteration by projecting the current estimate to the feasible set. Upon convergence, we switch to a gradient descent procedure so as to ensure correctness of the algorithm. In fact, this two-step approach can be interpreted as a single quasi-Newton scheme with respect to an appropriate series of transformation matrices $B^0, B^1, B^2, \dots, B^{final} = I$. Moreover, we found out in our experiments that the switch to the gradient descent method is not necessary in practice since it usually does not lead to any appreciable changes. This is not surprising, regarding the stable behavior of the linearized fixed-point iteration method (see Fig. 4.4).

Primal-Dual Method

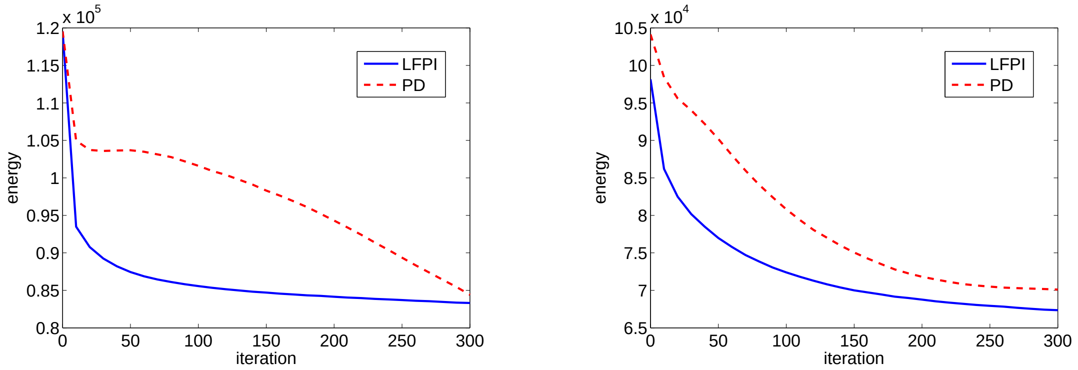
The primal-dual method derives a saddle-point problem by introducing an auxiliary variable $\xi : V \to \mathbb{R}^3$. More specifically, the optimization problem in (4.9) can be equivalently formulated as

$$\min_{u\in D}\ \max_{\xi\in K} \int_V \langle \nabla u(x), \xi(x)\rangle\, dx, \tag{4.24}$$

where $K = \{\ \xi : V \to \mathbb{R}^3 \mid |\xi(x)| \le \rho(x) \quad \forall\, x \in V\ \}$. Now, the optimization in (4.24) can be performed by gradient descent in u and gradient ascent in ξ in alternating manner. Thus, starting with some initial $(u^{(0)}, \xi^{(0)}) \in D \times K$ and $\bar{u}^{(0)} = u^{(0)}$, we get the following iterative procedure

$$\begin{aligned} \xi^{(k+1)} &= \Pi_K(\xi^{(k)} + \sigma\,\nabla\bar{u}^{(k)}) \\ u^{(k+1)} &= \Pi_D(u^{(k)} - \tau\,\text{div}(\xi^{(k+1)})) \\ \bar{u}^{(k+1)} &= 2u^{(k+1)} - u^{(k)}, \end{aligned} \tag{4.25}$$

where $\sigma, \tau > 0$ denote time-step parameters and $\Pi_C(.)$ is the projection operator for the set C. In our implementation, we observed stable behavior for $\sigma, \tau = 0.1$.



head sequence **warrior sequence**

Fig. 4.4: Linearized fixed-point iteration (LFPI) vs. primal-dual (PD) method. Note that the linearized fixed-point iteration method manifests quite fast and stable energy decay. In contrast, the primal-dual method exhibits unpredictable (see the jump in the left plot) and considerably slower behavior.

Comparative Evaluation

As in Chapter 3, we present a comparative evaluation of both numerical methods by studying their suitability for solving the convex optimization problem at hand given in (4.9). To this end, we pick two image sequences exhibiting different geometric and photometric properties - the "head" sequence in Fig. 4.5 and the "warrior" sequence in Fig. 4.8. While in the first case both photoconsistency and silhouette constraints play an important role due to the presence of concavities as well as thin structures like the pedestal, in the second case the silhouette information prevails as the obtained reconstruction is close to the visual hull model.

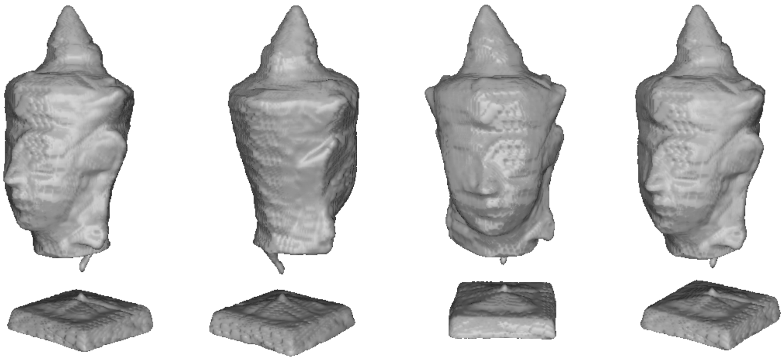
In Fig. 4.4, the energetic evolution for the two proposed numerical schemes – the linearized fixed-point iteration method (LFPI) and the primal-dual method (PD) – is depicted. As can be observed, the linearized fixed-point iteration method manifests quite fast and stable energy decay. In fact, the demonstrated results were reached after 100-120 iterations without any further visual changes. In contrast, the primal-dual method offers unpredictable and considerably slower behavior. This appears unexpected, regarding the observations in Chapter 3. We believe that it is due to some stability weaknesses of the primal-dual method, even with the appropriate time-step settings. Note that plotting the energetic evolution with respect to the physical time will not give any substantial differences here since the computational time for the two numerical procedures is dominated by the constraint enforcement. As mentioned in Chapter 3, a GPU implementation of both schemes is straightforward by parallelizing the computations over the discretized volume grid.[4] See Table 4.1 for GPU runtimes obtained with the linearized fixed-point iteration method.

4.5 Experimental Results

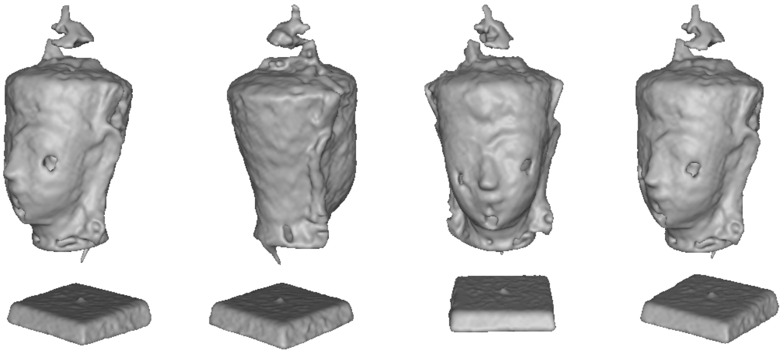
In this chapter, we proposed a convex optimization method to combine stereo and silhouette information for multiview reconstruction. In the following, we provide a series of experiments to assess the properties of the method. Although the presented approach achieves state-of-the-art accuracy, we believe that its main advantage is its robustness, i. e. the ability to operate in many practical situations, where traditional methods fail. To emphasize this, we

---

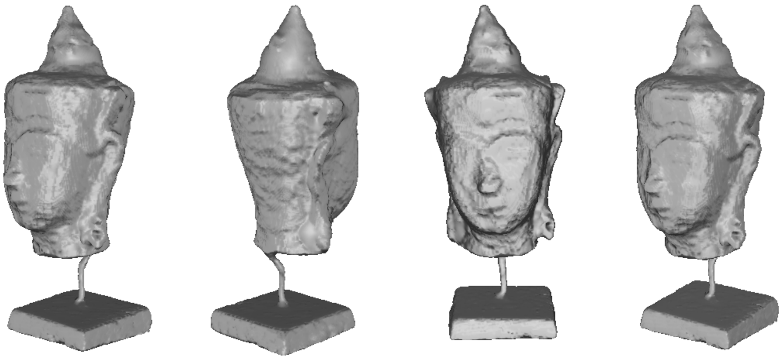
[4] Recall that SOR can be parallelized with a red-black strategy.



multiview stereo with silhouette-based regional terms



spatial propagation of stereo information



proposed integration of stereo and silhouettes

Fig. 4.5: Experimental comparison with alternative multiview stereo methods on the reconstruction of a metallic head (see Chapter 3 for more details on the methodology). While both the approach utilizing silhouette-based regional terms (upper row) and the more elaborate model relying on volumetric photo-consistency (middle row) fail to recover the pedestal of the statue and produce oversmoothing effects or erroneous carving, the proposed method (bottom row) accurately recovers all relevant details.

focus on four types of objects which exhibit important challenges for image-based modeling techniques, namely

- We consider objects that are not Lambertian such as metallic and shiny ones. For such objects the key assumption underlying the stereo approach that points on the surface have the same color when seen from different views is violated such that the stereo information alone is likely to provide suboptimal reconstructions.

- We consider objects with complex and fine-scale details. In traditional weighted minimal-

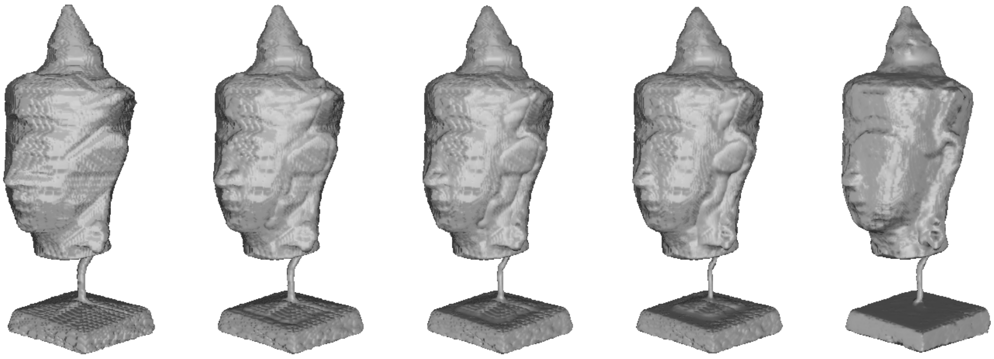


Fig. 4.6: Minimization process. Surface evolution starting from the visual hull, obtained by projecting the current estimate onto the original domain. Note that the presented method is able to generate accurate shapes starting from this initialization since it does not take the local surface orientation into account.

surface approaches, elongated and fine-scale details are typically suppressed as this drastically reduces the surface area.

- We consider objects that exhibit little prominent texture. Such objects are known to be difficult for stereo approaches since the matching of similar structures no longer provides a reliable discrimination of good and bad matches.
- We consider image sequences acquired with a hand-held camera lacking perfect color and camera calibration. Analogously, the matching process is quite challenging under such conditions.

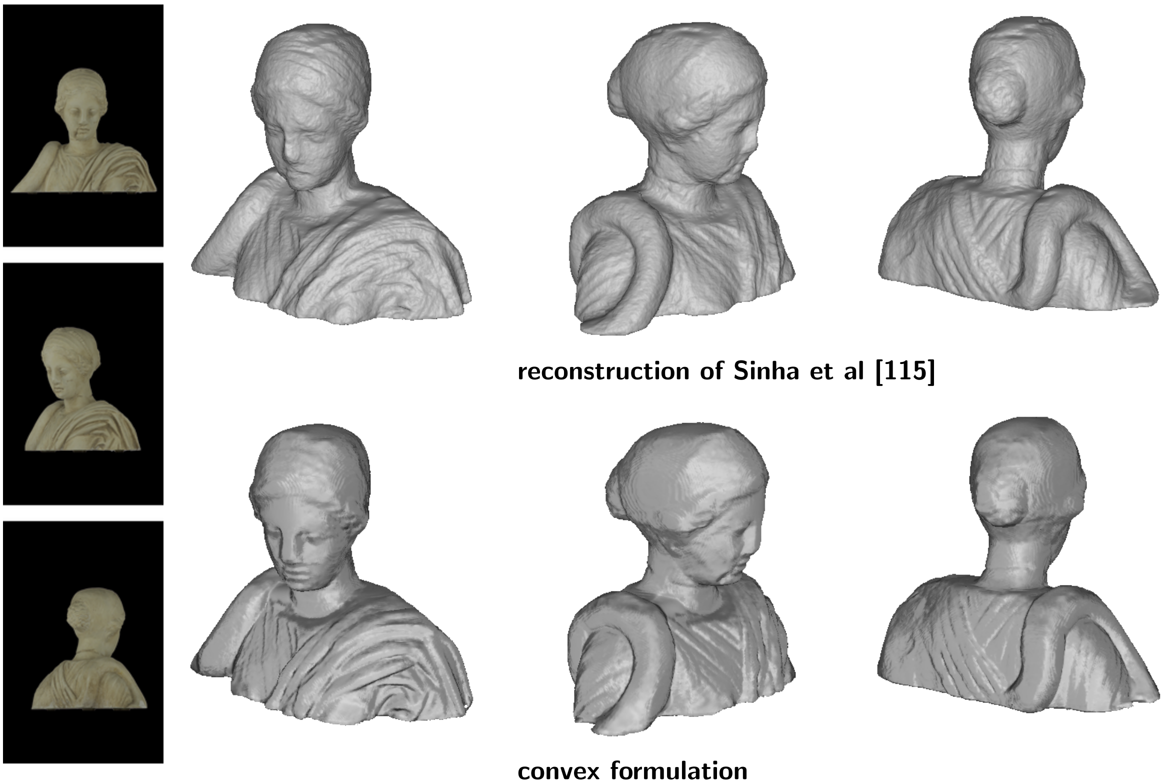
In addition, we compare the proposed method to other reconstruction methods already introduced in Chapter 3, and also to alternative techniques imposing simultaneously photometric and silhouette consistency.

Comparison to Alternative Multiview Stereo Methods

We validate the proposed approach on a scene of a head statue with complex reflection properties containing thin structures (the pedestal); see Fig. 4.5. Scenes of this type are a known challenge for variational stereo-based methods due to the violation of the Lambertian assumption and the presence of a regularizer which introduces a bias towards shapes with small area. To confirm this, we applied two of the energy models introduced in Chapter 3 – energy model II and III. The first method produces clear oversmoothing effects at concavities and small-scale structures. The second approach retrieves shape indentations but also leads to erroneous carving at thin parts and specularities. In contrast, the introduced technique produces accurate reconstructions of thin structures (the pedestal) as well as concave areas by incorporating silhouette constraints in the optimization process. Note that all three models are based on a classical minimal surface formulation but use different methodologies to avoid the empty surface as a solution. Note also that all three methods use silhouette information to define regional terms, for initialization or to constrain the domain of admissible functions.

Absence of Bias Towards the Visual Hull

The proposed method is based on minimizing a convex energy under convex constraints. Although the computed solution does not depend on the initialization, in practice we can



reconstruction of Sinha et al [115]

convex formulation

Fig. 4.7: Hygeia sequence. 3 out of 36 input images of resolution 2008×3040 and multiple views of the reconstructed surface compared to the reconstruction of Sinha et al. [115]. The result obtained with the proposed approach exhibits a higher grade of smoothness while recovering surface details more accurately (for example the face and the creases of the cloth).

accelerate computation by initializing the solution with the visual hull. In Fig. 4.6, we show intermediate steps in the evolution process of the proposed approach. Usually, local minimization techniques use the surface orientation to identify locations to deform the current shape in order to minimize the resulting reprojection error. However, this could lead to instabilities and introduce a bias near the visual hull boundary by involving surface points beyond the contour generator. In contrast, the introduced method recovers shape indentations effortlessly while retaining silhouette alignment during the optimization process.

Comparison to the Approach of Sinha et al. [115]

Fig. 4.7 shows a comparison between the proposed convex relaxation method and the approach of [115] on an image sequence of a statue of a Greek goddess. The experiments show that both approaches give rise to a high-quality silhouette-consistent reconstruction. The proposed method, however, offers visible improvements in the area of the face and the creases of the cloth. Overall, it provides a smoother reconstruction that nevertheless preserves the relevant structures. On the other hand, the recovered surface seems to be slightly overflattened at some locations (e.g. on the nose or the back), which is due to the discrepance between image and volume sampling, discussed previously. Note that the input photographs are of relatively high resolution while the computational costs and memory requirements of a volumetric representation with a comparable sampling rate would be prohibitive. It should be noted that the formulation in [115] differs from the proposed one not only in the way of fusing photometric and silhouette information but also in other aspects.

Fig. 4.8: Warrior sequence. 2 out of 24 input images of resolution 1600 × 1600 and multiple views of the reconstructed surface. Note that thin structures (for example the handle of the hammer) as well as concavities (for example at the chest) are reconstructed accurately.

iterative projection **Euclidean projection**

Fig. 4.9: Iterative vs. Euclidean projection. Imposing silhouette consistency by means of Euclidean projection onto the feasible set of surfaces is particularly important to correctly handle occluded concave parts (see marked area). While the iterative projection, described in Section 4.3, often produces artifacts in such cases, the more accurate Euclidean projection overcomes the numerical difficulties and gives better reconstructions.

Reconstruction of Complex and Fine-scale Objects

While the above objects were topologically rather simple, Fig. 4.8 shows experimental results obtained on a considerably more complex geometry of a warrior figurine that has many small-scale structures such as the hammer and the sword. Again, we observe that the proposed reconstruction exhibits a relatively high degree of smoothness while preserving all fine-scale geometric details. Note that the image sequence consists of only 24 images from camera positions arranged on a hemisphere around the object, which implies large baselines and thus limited accuracy of the estimated photoconsistency.

Fig. 4.9 shows a comparison between the reconstructions obtained by applying the simple iterative projection method and the more accurate Euclidean projection in the course of optimization (see Section 4.3). While the simple iterative approach produces artifacts in the

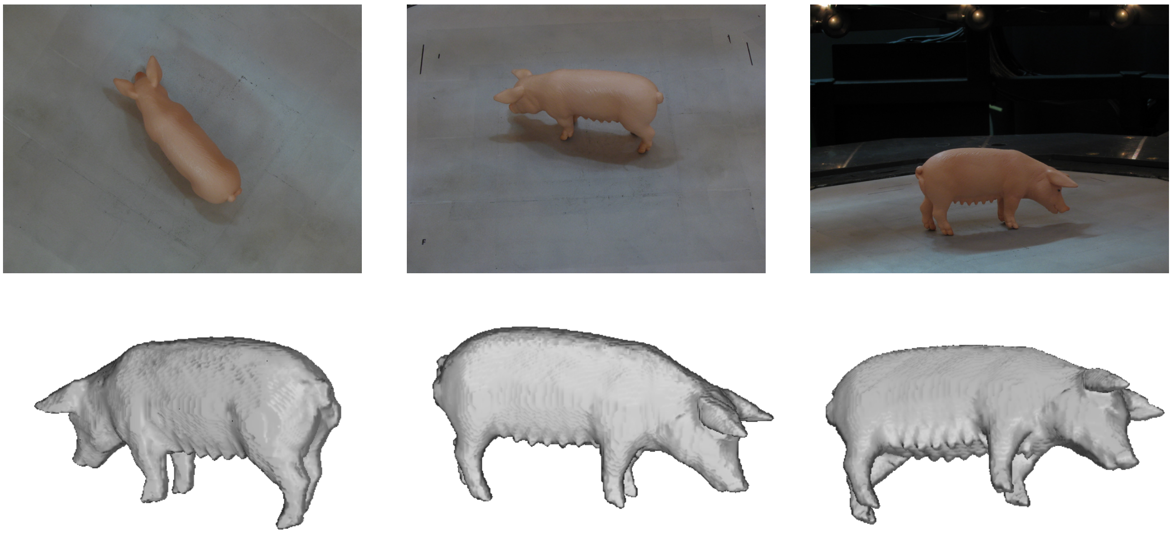


Fig. 4.10: Sow sequence. *First row:* 3 out of 27 input images of resolution 1024×768. *Second row:* Multiple views of the reconstructed surface. Despite a lack of prominent texture, the proposed method allows an accurate reconstruction even of fine-scale details of the object.

area below the fur at the back (marked in red), the Euclidean one overcomes these numerical difficulties and gives a more accurate reconstruction. Generally, the simple iterative procedure performs poorly in the presence of occluded concave areas. However, in most practical scenarios such cases do not occur. In fact, in our experiments the “warrior” sequence was the only one, where we could observe considerable differences between the results obtained by both techniques.

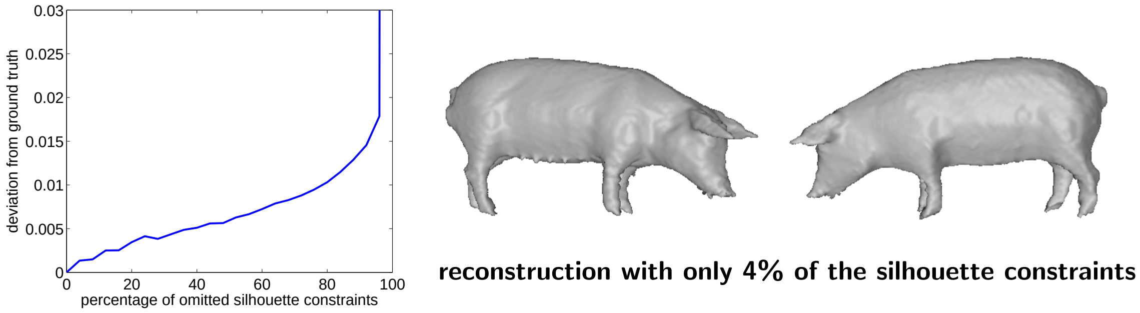
Reconstruction of Low-textured Objects

The next experiment, shown in Figure 4.10, illustrates that the proposed method allows to compute good quality reconstructions even for objects that exhibit rather little prominent texture. For such experiments, where the stereo-matching provides almost no relevant information regarding good or bad matches, the proposed method essentially computes a Euclidean minimal surface that exactly fulfills all silhouette constraints. The results show that even in the absence of reliable stereo information, highly detailed reconstructions could be obtained.

Robustness to Missing Silhouette Information

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

In order to explore the behavior of the proposed method in case of incomplete silhouette information, we assume that only a certain percentage $p \in [0, 1]$ of all constraints of type



reconstruction with only 4% of the silhouette constraints

Fig. 4.11: Robustness to missing silhouette information. *Left:* Completeness of the reconstruction obtained with the proposed approach applied on the image sequence in Fig. 4.10 for reduced number of silhouette constraints. The omitted constraints of type (4.14) were chosen randomly. As a ground truth served the reconstruction obtained with complete information (see Fig. 4.10) and the deviation from it was estimated by evaluating (2.25). *Right:* Reconstruction obtained with only 4% of the constraints of type (4.14). The corresponding deviation value is about 0.018. Although the accuracy of fine-scale details degrades by reducing the number of silhouette constraints, the overall large-scale accuracy remains high even when ignoring 96% of them.

(4.14) is taken into account for reconstruction. Note that omitting constraints of type (4.13) will have less impact on the reconstruction due to the nature of the silhouette fusion process. In addition, we assume for objectivity that the subset of ignored constraints has been selected randomly, i. e. a certain proportion of it is randomly switched off.
Let $u_{full} : V \to \{0,1\}$ denote the solution computed with complete silhouette information and $u_p : V \to \{0,1\}$ the solution obtained using only a fraction $p \in [0,1]$ of it. To quantify the decay in performance with fewer and fewer constraints, we define the reconstruction error with respect to the solution with complete information as the relative deviation ϵ_p:

$$\epsilon_p = \frac{\displaystyle\int_V |u_{full}(x) - u_p(x)|\, dx}{\displaystyle\int_V u_{full}(x)\, dx + \int_V u_p(x)\, dx}. \tag{4.26}$$

In particular, we have $\epsilon_p \in [0,1]$, with $\epsilon = 0$ if and only if both reconstructions are identical, and $\epsilon = 1$ if u_p is the empty set.
Figure 4.11 shows the average reconstruction error computed as a function of the percentage of missing silhouette constraints. It shows that the reconstruction accuracy gradually degrades with decreasing value of p. Yet, for values $p \leq 0.96$ the reconstruction error is below 0.02. The figure depicts also the reconstruction with only 4% of the silhouette constraints (i. e. $p = 0.96$). A direct comparison to Fig. 4.10 reveals that many of the fine-scale details are oversmoothed while all relevant large-scale parts of the object are preserved. We can conclude that due to the utilized silhouette fusion scheme the proposed approach enjoys considerable resilience to missing silhouette information. This assures its high practical applicability.

Reconstruction from a Hand-held Camera

The calibration of the input images is a crucial ingredient of the reconstruction pipeline. While the above reconstructions were all conducted with imagery obtained either from the internet or in controlled environments with precalibrated cameras, one may ask if the proposed method is sufficiently robust to also deal with outdoor sequences generated with a hand-held camera. In many real-world applications one cannot precalibrate the camera – possibly because the

Fig. 4.12: Reconstruction from a hand-held camera. 2 out of 28 input images of resolution 2048×3072 and multiple views of the reconstructed surface. Despite severe illumination variations in the input images, the creases of the cloth are recovered in high detail.

images were acquired by another person or because it is not possible to include calibration patterns in the scene.

In the following experiment, we demonstrate the robustness of the proposed approach to operate with input images lacking perfect color and camera calibration, even though the imaged object is well-textured and does not feature complex reflectance properties. We acquired 28 photographs of an ancient statue with a hand-held camera (see Fig. 4.12). Fig. 4.12 depicts multiple views of the recovered surface. Generally, the reconstruction exhibits a high grade of accuracy despite the severe intensity changes in the images and the imprecisions of the calibration. Note, in particular, the detailed recovery of the creases of the cloth. Note also the correct handling of the non-trivial topology of the reconstructed surface.

Energy Bound and Runtimes

As detailed in Section 4.2, the proposed approach does not give a global minimum of the underlying energy functional (4.6). Yet, the computed solution lies within an energy bound, dependent on the particular input data, around the global minimum (see Proposition 4). Here, we present the perimeter of this upper bound for all demonstrated experiments.

Apart from robustness, another crucial issue for a silhouette and stereo integration approach is the computational time needed. To this end, we used a GPU implementation of the linearized fixed-point iteration scheme, where the SOR optimization in a red-black strategy runs on the GPU. For comparison, we implemented the simple iterative projection method (see Section 4.3) as well as the more accurate Euclidean projection method (see Section 4.3). Note, however, that due to their sequential nature a GPU implementation of the constraint enforcement step is not meaningful. As a result, the back-projection dominate the overall computational time of the optimization.

Table 4.1 summarizes the energy bounds and runtimes for all presented experiments. In order to deliver insight into how tight the bounds are, we estimated the energy gap as the ratio of the energy at the computed solution (i. e. after thresholding) and the solution of the relaxed problem (i. e. before thresholding). The variations seem to correlate with the accuracy of the photoconsistency map and the geometry of the imaged object (e.g. the presence of concavities). The computational times were measured on a PC with 2.83 GHz and 8 GB of main memory, equipped with a NVIDIA Tesla C2070 graphics card. Note that the optimization runs partially on the GPU (SOR scheme) and partially on the CPU (constraint

	# views	image resolution	energy before thresholding	energy after thresholding	energy gap	runtime (iter. proj.)	runtime (Eucl. proj.)
head	33	1024×768	86315.2	107146	1.24	14 s	29 s
hygeia	36	2008×3040	127877	165292	1.29	91 s	322 s
warrior	24	1600×1600	68160.8	109829	1.61	19 s	97 s
sow	27	1024×768	59661.2	70429.6	1.18	6 s	16 s
statue	28	2048×3072	112417	146437	1.3	47 s	181 s

Tab. 4.1: Energy bounds and runtimes for all experiments. The optimization runs partially on the GPU (SOR scheme) and partially on the CPU (constraint enforcement). For all experiments the volumetric resolution was between 15 and 20 million voxels.

enforcement). Note also that the runtime of the proposed approach strongly depends on the resolution of the input images since it gives the number of constraints to be considered. The runtime also depends on the particular choice of projection method. However, the dependence on the volume resolution is less relevant since the evolution scheme runs on the GPU. For all experiments the volumetric resolution was within the range 15-20 million voxels. The reported times do not include the photoconsistency estimation as the latter is very sensitive to the choice of methodology.

4.6 Discussion

In this chapter, we proposed a framework for integrating silhouette and stereo information in 3D reconstruction from multiple images. The key idea is to cast multiview stereovision as a convex variational problem and to impose exact silhouette constraints by restricting the domain of feasible functions. Relaxation of the resulting formulation leads to the minimization of a convex functional over the convex set of silhouette-consistent functions, which can be performed in a globally optimal manner using classical techniques. A solution of the original problem is obtained by projecting the computed minimizer onto the corresponding restricted domain. We presented an algorithm which allows to efficiently compute the projection of the current solution onto the convex set of silhouette-consistent configurations. We also proved that the final reconstruction is within an energetic bound of the optimum.

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

5

Anisotropic Minimal Surfaces and Minimal Ratios

A theory is a policy rather than a creed.

Joseph John Thomson (1856-1940)

In the previous chapters, we considered different convex energy models designed with focus on particular paradigms for image-based modeling. In this chapter, we extend the class of functionals amenable to convex optimization and alongside the range of applicability of the respective approaches.

5.1 Introduction

Motivation

As pointed out in Chapter 3, normal information has become an established tool in multiview stereo since it captures small-scale surface details which are difficult to recover with other methodologies. Yet, estimating shape orientation by means of stereo matching is quite vulnerable to noise due to the imprecision of the matching process. Thus, we are facing the non-trivial question of how to exploit the high-frequency information provided by a normal field, possibly corrupted by noise, and enhance the reconstructed 3D model without reducing its overall accuracy. In particular, we are looking for a unified framework which allows to integrate different cues like silhouettes, photoconsistency and surface normal information in a robust and transparent manner. The ultimate goal is to combine their advantages and extend the range of applicability of established techniques. Such an approach should be able to reconstruct fine-scale shape details on thin protruding structures – a test scenario for which most of the existing methods fail (see Fig. 5.1). Once again, convex optimization proves useful and powerful and promotes reaching that goal.

Furthermore, one could observe that most of the energy models, considered so far, exhibit a linear structure of a data term and a smoothness term, balanced by means of a weighting parameter λ, which fit into the general form (1.5). Setting this parameter in an adequate manner is not a trivial task. In fact, the regularizer introduces a shrinking bias in the reconstruction process as it prefers shapes of small area. As a consequence, a too small value for λ leads to oversmoothing effects, whereas a too large value produces generally a noisy 3D model. Interestingly, there exist also other energy models exempt from such weighting parameters. An example poses the minimal ratio model, where one minimizes the ratio between a data term and a smoothness one instead of a linear combination. The minimal ratio model is particularly appealing due to some useful properties. Firstly, it does not suffer from a shrinking bias, inherent to minimal surface models. Secondly, the minimal ratio model is scale invariant, i. e. for a given object the quality of reconstruction does not depend on the spatial scale at which it is perceived. Thirdly, a globally optimal solution can be obtained by

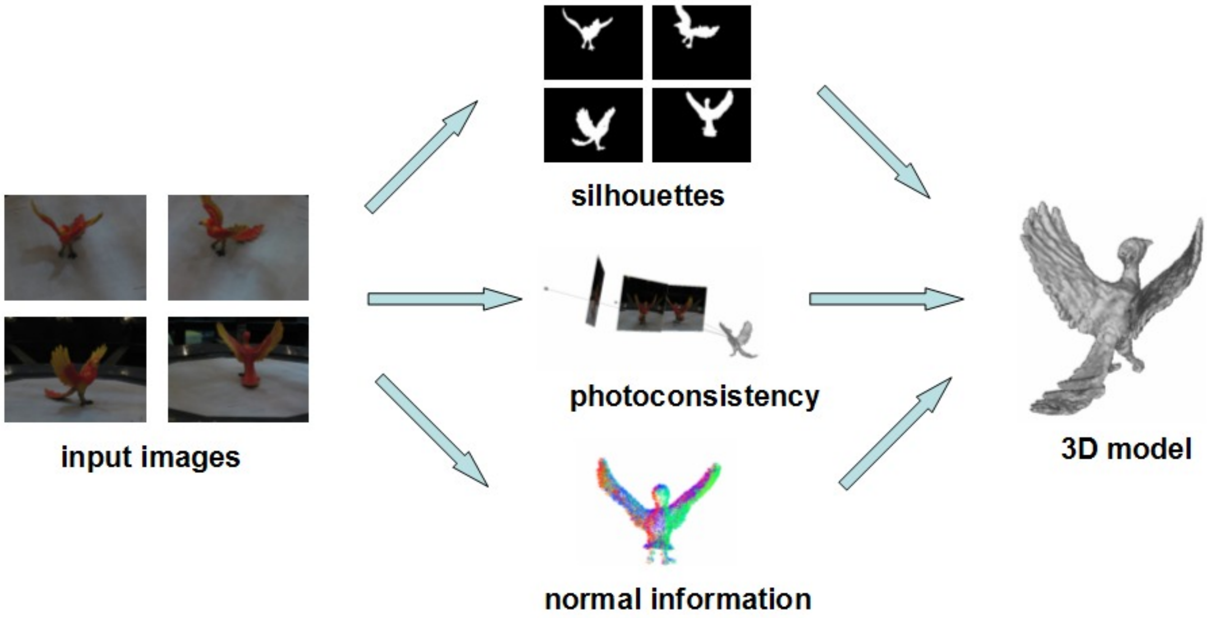


Fig. 5.1: Accurate 3D reconstructions for challenging objects can be obtained by fusing different cues like silhouettes, photoconsistency and normal information. A unified framework allows to recover thin protrusions (e. g. the wings of the bird) as well as fine-scale details on them (e. g. the feathering).

solving a sequence of convex optimization problems.

Previous Work on Anisotropic Minimal Surfaces

In Chapter 3, we became acquainted with different methods for multiview stereo based on surface normal integration, e. g. [12, 41, 49]. Essentially, they minimize a cost functional of the form

$$E(S) = -\lambda \int_S \langle N_S(s), F(s) \rangle \, ds + \int_S \rho(s) \, ds, \tag{5.1}$$

where N_S denotes the local surface normal of S and $F : V \to S^2$, where $S^2 \subset \mathbb{R}^3$ is the unit sphere, represents an estimate of the unit outward orientation of the observed shape. Moreover, $\rho : V \to (0, 1]$ is a photoconsistency map, reflecting the agreement of corresponding image projections, and $\lambda \in \mathbb{R}$ is a weighting parameter. Note that the last term in (5.1) serves as a smoothness term needed to suppress noise in the normal field F. Typically, the normal estimates obtained by means of stereo matching are quite noisy due to the imprecision of the matching process. This is in contrast to photometric stereo approaches, where the local shape orientation is derived from estimated reflectance properties of the object and measured color observations [139, 34]. Due to the high reliability of the shading information, photometric stereo methods usually do not impose any surface priors like smoothness and minimize a simple functional like

$$E(S) = -\int_S \langle N_S(s), F(s) \rangle \, ds, \tag{5.2}$$

where $F : V \to S^2$ denotes the estimated normal field. Yet, the high accuracy of the computed normals comes at the expense of restricted applicability of photometric stereo as this paradigm requires controlled illumination conditions.

An alternative strategy to integrate provided surface normal information is by means of an anisotropic minimal surface model with respect to an appropriate Finsler metric introducing directional selectivity. Such approaches were developed within the context of discrete optimization [64, 76]. Formally, they involve minimizing an energy functional of the form

$$E(S) = \int_S L_{S(s)}(N_S(s))\, ds, \tag{5.3}$$

where $L_x(.)$, $x \in V$, signifies a general Finsler metric. It is important to notice that minimizing the functional in (5.3) alone is not meaningful as it poses a minimal surface problem and as such gives the empty set as a solution. Yet, the model in (5.3) lends itself to a regularizer in a variational formulation.

Subsequently, based on previous work on anisotropic diffusion processes [136], anisotropic minimal surface models were generalized to the domain of continuous optimization [143, 97]. It was shown that the model in (5.3) gives rise to a convex optimization problem after relaxation and thus can be solved globally also in a continuous setting.

Previous Work on Minimal Ratios

Minimal ratio models were developed as an alternative to classical minimal surface models and linear models combining data terms and regularizer. They can be classified in two categories – discrete and continuous ones.

Discrete minimal ratio models were explored within the graph cut framework [57, 77]. In [57], the following functional is regarded

$$E(S) = -\frac{\displaystyle\int_S \langle N_S(s), F(s)\rangle\, ds}{\displaystyle\int_S \rho(s)\, ds}, \tag{5.4}$$

where $F : V \to S^2$ denotes a normal field with the desired surface orientation and $\rho : V \to (0, 1]$ is a function encoding certain spatial preference. Note that the minimal value of the functional in (5.4) is negative, i. e. minimizing it is equivalent to maximizing its absolute value. The functional in (5.4) is particularly interesting due to some favorable properties, which will be discussed later in the chapter, and its global optimizability. In [77], the class of ratio models, that can be optimized globally in a discrete setting, was extended by

$$E(S) = \frac{\displaystyle\int_{int(S)} \rho_{obj}(x)\, dx}{\displaystyle\int_S \rho(s)\, ds} \tag{5.5}$$

and

$$E(S) = \frac{\displaystyle\int_S \rho(s)\, ds}{\displaystyle\int_{int(S)} dx}, \tag{5.6}$$

where $int(S) \subset V$ denotes the interior of S and $\rho_{obj} : V \to \mathbb{R}$ is a given volumetric potential function.

Minimal ratio models were considered also in a spatially continuous setting within the framework of total variation minimization [120]. Thereby, the following functional was proposed

$$E(u) = \frac{\lambda \int_V \rho_{obj}(x)u(x)\,dx + \int_V |\nabla u|\,dx}{\int_V \rho'_{obj}(x)u(x)\,dx + K}, \tag{5.7}$$

where $u : V \to [0,1]$ denotes a relaxed labeling function, $\rho'_{obj} : V \to \mathbb{R}_{\geq 0}$ is a non-negative volumetric potential function and $\lambda \in \mathbb{R}$, $K > 0$ are constants. It turns out that a global minimum of (5.7) can be obtained by solving a single convex optimization problem.

Contribution

The main contribution of this chapter is to propose an energy minimization framework for multiple view reconstruction that allows to combine multiview photoconsistency, silhouette and normal information. We show that the reconstruction can be efficiently determined as an anisotropic minimal surface which favors not only locations of good photoconsistency but also orientations that are consistent with the specified normal field. By adaptively reducing the smoothing along the predetermined directions, the inherent shrinking bias of minimal surface models is alleviated. We show that relaxation techniques can still be applied to this general framework to derive a convex optimization problem and find a globally optimal solution.

Additionally, we explore minimal surface models in more detail and provide an exact mathematical characterization of both the inherent shrinking bias and the scale dependency. As an alternative, we propose a minimal ratio model addressing these limitations. It is based on solving a sequence of convex optimization problems. We prove that the proposed minimal ratio solution does not suffer from the above problems. Firstly, it does not exhibit a shrinking bias and allows for better reconstructions of concavities and protrusions. In particular, we prove that any disjoint surface with the same energy can be added without affecting the overall energy costs. Secondly, the minimal ratio formulation does not have any tuning parameters and is shown to be scale invariant. These properties are confirmed in experimental evaluations.

The main results in this chapter are published in [69, 74].

5.2 Anisotropic Minimal Surface Model

Anisotropic Metrics

This paragraph introduces the main concept of anisotropic metrics preferring certain orientation selectivity by generalizing the already encountered weighted minimal surface model.

Our starting point is a classical multiview stereo approach introduced in Chapter 3 (see (3.1)). According to it, the most photoconsistent shape is obtained by minimizing

$$E(S) = \int_S \rho(s)\,ds, \tag{5.8}$$

where $\rho : V \to (0,1]$ denotes a photoconsistency map. The model encourages the surface to pass through points with high observation agreement. Its minimization identifies shapes with minimal overall area with respect to a local isotropic metric induced by ρ. However, the above formulation does not explicitly model the orientation of the estimated shape. To this end, in case of given surface orientation, a generalization has to be developed. This

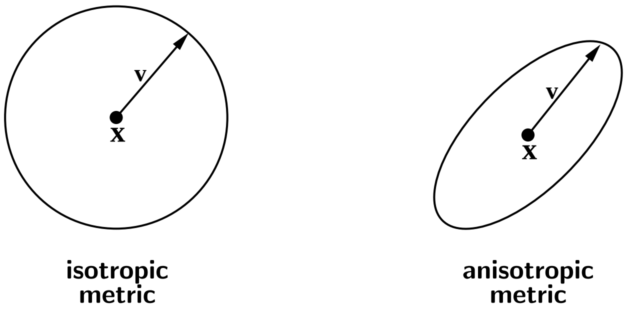


Fig. 5.2: Local distance maps. Examples of local distance maps – $\sqrt{v^T v} = 1$ (isotropic case) and $\sqrt{v^T D v} = 1$ (anisotropic case). While isotropic metrics treat all directions equally, anisotropic metrics possess directional selectivity.

can be achieved by switching to a more general metric and introducing a family of positive semi-definite anisotropic tensors $D(x) \in \mathbb{R}^{3\times 3}, x \in V$ tolerating certain directional selectivity. Now, the minimal surface model (5.8) reads

$$E(S) = \int_S \sqrt{N_S(s)^T D(s) N_S(s)}\, ds. \tag{5.9}$$

Obviously, the tensors $D(x)$ can be designed to energetically favor certain shape orientations while suppressing others. Note that D is defined pointwise, i. e. it is spatially varying. However, in the remainder of this section, we will omit the argument for the sake of simplicity. The energy model (5.9) can sill be interpreted as a minimal surface formulation according to the Riemannian metric induced by D (i. e. $|v|_D = \sqrt{v^T D v}$).[1] The Euclidean metric, which treats all spatial directions equally, appears as a special case with $D = I$, where $I \in \mathbb{R}^{3\times 3}$ denotes the identity matrix. Examples of local distance maps in 2D of the Euclidean and the more general Riemannian metric are visualized in Fig. 5.2. The weighted minimal surface model (5.8) also appears as a special case for $D = \rho^2\, I$.

Now, we are confronted with the question of defining the family of anisotropic tensors D appropriately. Let us assume that a vectorfield $F : V \to \mathbb{R}^3$ is provided representing an estimate of the unit outward orientation of the sought-after shape. In practice, meaningful normal estimates can be computed only for points on the surface of the observed object. For all other points we can set $D = \rho^2\, I$, which corresponds to the conventional isotropic case. Thus, in the sequel, we assume $F : V \to S^2 \subset \mathbb{R}^3$. In Section 5.2 we give more details on how orientation information can be obtained from the input images. Based on this data, we would like to suppress regularization along the corresponding normal and encourage the process along the tangent plane so as to enforce the surface to assume the desired orientation. Moreover, the photoconsistency map ρ, which encodes spatial preference, should be taken into account. This can be achieved by setting

$$D = \rho^2 \left(\tau F F^T + \frac{3-\tau}{2} (I - F F^T) \right), \tag{5.10}$$

where $\tau \in [0, 1]$ is a weighting parameter that controls the distortion of the corresponding metric, i. e. the tolerance of the normal field F. In effect, the formulation in (5.10) realizes a basis transformation and subsequent scaling. The first term treats the component along

---

[1] In a mathematically rigorous sense, a Riemannian metric is defined as a family of scalar products. Here, with a Riemannian metric we refer to the variable metric induced by the family of scalar products $\langle v_1, v_2 \rangle_{D,x} = \sqrt{v_1^T D(x) v_2}$.

F and scales it by τ, whereas the second one affects the tangential components. It is easy to verify that the choice $\tau = 1$ gives the original model (5.8). On the other hand, $\tau = 0$ will completely turn off smoothing along the vectorfield F. In our experiments we found out $\tau = 0.15$ to be a good compromise for a moderately noisy normal field.

One could wonder about the concrete definition of D in (5.10). To justify this formulation, we show some favorable properties resulting from it.

Proposition 7. *For a normalized vector F, $\rho > 0$ and $\tau \in [0,1]$ the matrix D defined in* (5.10) *is symmetric and positive semi-definite with $tr(D) = 3\rho^2$.*

Proof. The symmetry of D is obvious.

In order to show that D is positive semi-definite, we observe for $v \in \mathbb{R}^3$

$$v^T D v = \rho^2 \, \frac{3\tau - 3}{2} \, \left(v^T F\right)^2 + \rho^2 \, \frac{3 - \tau}{2} \, |v|^2 \geq \rho^2 \, \tau \, |v|^2 \geq 0$$

due to $|F| = 1$.

Finally, we obtain

$$tr(D) = \rho^2 \, \left(\frac{3\tau - 3}{2} \, |F|^2 + \frac{3\,(3 - \tau)}{2} \right) = 3\rho^2.$$

□

The condition that D is symmetric and positive semi-definite justifies the correctness of the definition of the proposed anisotropic metric. The condition that $tr(D)$ does not depend on the choice of the parameter τ assures that the overall smoothing remains fixed. This is an important property as it makes the model in (5.9) suitable for a smoothness term in a more complex variational formulation. In this sense, the normal information is integrated in a soft manner.

Moreover, it should be noted that the inverse of the matrix in (5.10) as well as its square root can easily be computed as

$$\begin{aligned} D^{-1} &= \frac{1}{\rho^2} \left(\frac{1}{\tau} F F^T + \frac{2}{3 - \tau} (I - F F^T) \right) \\ D^{1/2} &= \rho \left(\sqrt{\tau} F F^T + \sqrt{\frac{3 - \tau}{2}} (I - F F^T) \right). \end{aligned} \tag{5.11}$$

This will be useful for optimization purposes (see Section 5.2).

Fusing Photoconsistency, Orientation and Silhouettes

In this paragraph, we propose specific energy functionals based on the anisotropic minimal surface model in (5.9), where it serves as a data-aware smoothness term. In particular, we generalize already encountered variational formulations.

Adding Regional Terms

As previously discussed, a undesired property of minimal surface models of the form (5.9) is that the empty set always exhibits a global minimum. One way to avoid this trivial solution is to derive additional information from the images giving a closer specification of the observed object. This can be achieved by introducing regional maps $\rho_{obj} : V \to \mathbb{R}$ and $\rho_{bck} : V \to \mathbb{R}$

defining costs for each point within the volume for being inside or outside the imaged shape, respectively. Now, we obtain the following energy model

$$\begin{aligned} E(S) &= \lambda \left(\int_{int(S)} \rho_{obj}(x)\, dx + \int_{ext(S)} \rho_{bck}(x)\, dx \right) \\ &+ \int_S \sqrt{N_S(s)^T D(s) N_S(s)}\, ds, \end{aligned} \tag{5.12}$$

where $int(S), ext(S) \subset V$ denote the surface interior and exterior, respectively, and $\lambda \in \mathbb{R}$ is a weighting parameter. In Chapter 2 and 3, we became acquainted with different methodologies for defining ρ_{obj} and ρ_{bck} based on object outlines and multiview stereo.

Next, we tackle the optimization of the functional (5.12). To this end, the first steps are a conversion to an implicit representation $u = \mathbf{1}_{int(S)}$, where $\mathbf{1}_{int(S)}$ denotes the characteristic function of $int(S)$, and subsequent relaxation. This yields

$$E(u) = \lambda \int_V (\rho_{obj}(x) - \rho_{bck}(x))\, u(x)\, dx + \int_V \sqrt{\nabla u(x)^T D(x) \nabla u(x)}\, dx, \tag{5.13}$$

where $u \in C_{rel} = \{\ \hat{u} \mid \hat{u} : V \to [0,1]\ \}$. The above formulation is derived from the relations $N_S = -\frac{\nabla u}{|\nabla u|}$ and $ds = |\nabla u| dx$. Note that the "binary" version of (5.13), i. e. optimization over the set of binary functions $u \in C_{bin} = \{\hat{u} | \hat{u} : V \to \{0,1\}\}$, is equivalent to (5.12). Fortunately, the optimization of (5.12) turns out to be as simple as minimizing (5.13) over C_{rel}, which exhibits a constrained convex optimization problem (see Proposition 1). In particular, we obtain

Theorem 9. *Let $u^* : V \to [0,1]$ be a global minimizer of the functional* (5.13). *Then, the characteristic functions of all upper level sets (i.e. thresholded versions)*

$$\Sigma_{\mu,u^*} = \{x \in\ V \mid u^*(x) > \mu\}, \qquad \mu \in (0,1), \tag{5.14}$$

of u^* are also global minimizers of (5.13).

Proof. The statement follows directly from the fact that the functional in (5.13) is a special case of the one in (1.6) (see Theorem 6). □

Hence, we can find a global minimum of (5.12) by minimizing the convex functional in (5.13) over the convex C_{rel} and thresholding the result at some $\mu \in (0,1)$. We can conclude that incorporating the proposed generalized anisotropic model does not entail any substantial difficulties with the optimization and retains globality of the solution.

Incorporating Silhouette Constraints

In certain practical scenarios, obtaining reliable volume subdivision terms may be a challenging task. In such cases, a reasonable alternative could be to retain the original minimal surface model (5.9), but to restrict the domain of feasible shapes in order to exclude the trivial solution. The object silhouettes serve as a useful tool that could provide such constraints

$$\begin{aligned} E(S) &= \int_S \sqrt{N_S(s)^T D(s) N_S(s)}\, ds, \\ \text{s. t.} \qquad & \pi_i(S) = Sil_i \quad \forall\, i = 1, \ldots, n. \end{aligned} \tag{5.15}$$

Note that for $D = \rho^2 I$ the above formulation boils down to the model proposed in Chapter 4.

Unfortunately, global optimization of (5.15) is not a trivial task. Nevertheless, a global minimum can be obtained up to an energetic upper bound. Reverting to an implicit representation and subsequent relaxation yields

$$
\begin{aligned}
E(u) = \int_V \sqrt{\nabla u(x)^T D(x) \nabla u(x)}\, dx&, \\
\text{s. t.} \qquad u \in [0,1]& \\
\int_{R_{ij}} u(r)\, dr \geq 1 \text{ if } j \in Sil_i& \\
\int_{R_{ij}} u(r)\, dr = 0 \text{ if } j \notin Sil_i&,
\end{aligned}
\tag{5.16}
$$

where R_{ij} denotes the visual ray through pixel j of image i.[2] It is evident that (5.16) exhibits a constrained convex optimization problem for which the global minimum can be obtained. Since we are interested in finding minimizers of the original non-convex problem (5.15), we threshold the solution of the convex problem u_{min} appropriately

$$
\tilde{u}(x) = \begin{cases} 1, & \text{if } u_{min}(x) \geq \mu \\ 0, & \text{otherwise} \end{cases}, \tag{5.17}
$$

where

$$
\mu = \min \left\{ \left(\min_{i \in \{1,\ldots,n\}, j \in Sil_i} \; \max_{x \in R_{ij}} u_{min}(x) \right) , 0.5 \right\}. \tag{5.18}
$$

In analogy to (4.12), the threshold is estimated such that the computed binary solution is the closest one that still fulfills exact silhouette consistency. Note that minimizing (5.15) is equivalent to minimizing the "binarized" version of (5.16) (where $u \in [0,1]$ is replaced by $u \in \{0,1\}$).

Now, a question about the benefits from the proposed formulation arises as we do not optimize the original problem (5.15) in a globally optimal manner. Fortunately, all globality guarantees, obtained for weighted isotropic minimal surfaces (see Chapter 4), carry over to this more general case.

Proposition 8. *Let u' be a (global) minimum of the "binary" version of* (5.16)*, $\tilde{u}$ the computed solution and u_{min} a (global) minimum of* (5.16)*. Then, a bound $\gamma(u_{min}, \tilde{u})$ exists such that*

$$
E(\tilde{u}) - E(u') \leq \gamma(u_{min}, \tilde{u}).
$$

Proof. As the introduced anisotropic metric preserves convexity, the proof is analogous to the proof of Proposition 4.

Since the binary functions form a subset of the real-valued functions, we have the relation

$$
E(u_{min}) \leq E(u') \leq E(\tilde{u})
$$

As a consequence, we obtain the inequality

$$
E(\tilde{u}) - E(u') \leq E(\tilde{u}) - E(u_{min}).
$$

□

Again, we can conclude that, also for this variational framework, incorporating the proposed anisotropic model does not affect any globality guarantees obtained for the solution.

---

[2] Actually, modeling silhouette constraints in a continuous setting requires representing the thickness of the material by a parameter $\delta > 0$ (see (4.8)). Yet, for the sake of simplicity, we set $\delta = 1$, which makes sense from a discrete point of view.

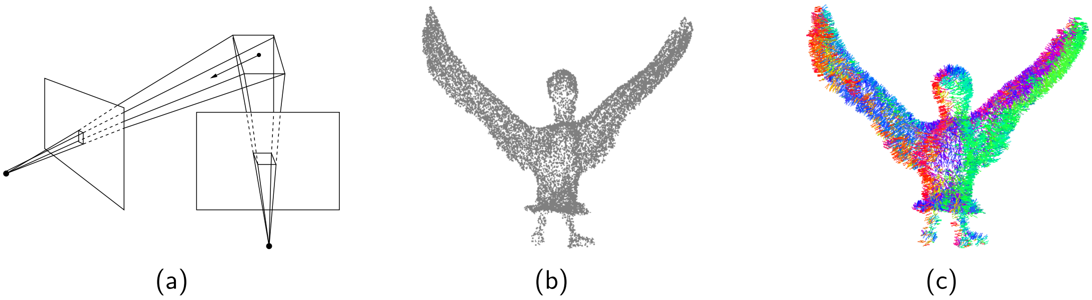


Fig. 5.3: Surface normal estimation. (a) The orientation of a point in space is obtained, based on a local planar patch. The optimal orientation is given by the maximal photometric agreement of the projections of the patch onto the images, where it is visible. (b) A point cloud for the data set in Fig. 5.5, generated with the software at [100]. (c) Corresponding color-coded normal vectors.

Implementation and Numerics

In this paragraph, we present a particular implementation of the proposed energy models and discuss their numerical optimization.

Data Term Computation

Following the formulation in (5.10), we need to define multiple data measures: a photoconsistency map ρ, regional subdivision costs ρ_{obj}, ρ_{bck} and an outward normal field F.

The photoconsistency estimation that we used in our experiments is based on the voting scheme proposed in Chapter 3. Moreover, we used the underlying propagating approach to derive volumetric assignment costs for object interior/exterior.

In order to obtain an estimate of a normal field F representing the surface orientation, we assume a sparse oriented point cloud

$$\begin{aligned} P &= \{\, p_i \mid p_i \in V \,\} \\ O &= \{\, v_i \mid v_i \in S^2 \,\} \,. \end{aligned}$$

As discussed in Chapter 3, such data can be obtained via an optimization procedure over the local photometric consistency (see Fig. 5.3). In our experiments, we used the approach of [41], an implementation of which is publicly available at [100]. According to it, the most truthful surface normal is obtained from the distortion of the projected local tangent patch. A sample oriented point cloud for a real image sequence, obtained with the above procedure, is visualized in Fig. 5.3 (b), (c). Based on this data, we define the vectorfield F as

$$F(x) = \begin{cases} v_i, & \text{if } x = p_i \\ 0, & \text{otherwise} \end{cases} . \tag{5.19}$$

In practice, we replace F with a semi-dense blurred version $\tilde{F}$ in order to account for inaccuracies due to image noise.

Primal-Dual Optimization

As mentioned previously, (5.13) and (5.16) pose classical constrained convex optimization problems. Hence, any iterative local optimization procedure will provide the global minimum. However, the particular choice of minimization method will affect the speed of convergence.

As we already encountered similar variational models in Chapter 2, 3 and 4, one could ask if the optimization procedures proposed there can be adapted to the current formulations. While the introduced anisotropic minimal surface model preserves convexity, the numerical optimization of the arising convex problems is, in this case, slightly more challenging. In the following, we generalize the previously adopted primal-dual scheme.
First, we observe that the energy functionals in (5.13) and (5.16) are both of the form

$$E(u) = \int_V \sqrt{\nabla u^T D \nabla u}\, dx + \int_V f u\, dx, \tag{5.20}$$

where $f : V \to \mathbb{R}$ summarizes the constant part not dependent on u, i. e. $f = \rho_{obj} - \rho_{bck}$ for the first energy model and $f = 0$ for the second one. Crucial in the proposed numerical procedure is the splitting of D as $D = D^{1/2} D^{1/2}$. Following a dual formulation, we introduce an auxiliary variable $\xi : V \to \mathbb{R}^3$ which allows for the following conversion

$$\begin{aligned} E(u) &= \int_V \sqrt{\nabla u^T D \nabla u}\, dx + \int_V f u\, dx \\ &= \int_V |D^{1/2} \nabla u|\, dx + \int_V f u\, dx \\ &= \max_{|\xi| \le 1} \int_V \langle \xi, D^{1/2} \nabla u \rangle\, dx + \int_V f u\, dx. \end{aligned} \tag{5.21}$$

Now, we obtain a new functional

$$E(u, \xi) = \int_V \langle \xi, D^{1/2} \nabla u \rangle\, dx + \int_V f u\, dx \tag{5.22}$$

that should be minimized with respect to u and maximized with respect to ξ under the constraints $u \in C_{gen}$, where $C_{gen} = C_{rel}$ in the first case and $C_{gen} = D$ (see (4.10)) in the latter case, and $|\xi| \le 1$. This states a typical saddle point problem that can be solved by a projected gradient descent/ascent strategy. Denoting by $K = \{\ \xi : V \to \mathbb{R}^3 \mid |\xi| \le 1\ \}$ a vectorfield mapping within the unit ball, a primal-dual optimization scheme for the anisotropic model in (5.20) can be described as follows.
We choose $(u^0, \xi^0) \in C_{gen} \times K$ and let $\bar{u}^0 = u^0$. We choose two time-steps $\tau, \sigma > 0$. Then, we iterate for $k \ge 0$

$$\begin{aligned} \xi^{(k+1)} &= \Pi_K(\xi^{(k)} + \sigma(D^{1/2} \nabla \bar{u}^{(k)})) \\ u^{(k+1)} &= \Pi_{C_{gen}}(u^{(k)} + \tau(\operatorname{div}(D^{1/2} \xi^{(k+1)}) - f)) \\ \bar{u}^{(k+1)} &= 2u^{(k+1)} - u^{(k)}, \end{aligned} \tag{5.23}$$

where Π_K and $\Pi_{C_{gen}}$ denote projections onto the corresponding sets.
Both projections are realized by simple normalization and clipping, respectively, as described in Chapter 2, 3 and 4. Note that the matrix square root $D^{1/2}$ can easily be computed according to the construction (see (5.11)). Recall, however, that $D^{1/2}$ is, in general, spatially varying.

5.3 Minimal Ratio Model

While minimal surface models, like the ones in (5.8) and (5.9), are shown to provide good reconstructions and robustness to noise, they suffer from two important shortcomings.

- Minimal surfaces are known to exhibit a *shrinking bias*. In particular, the global optimum is always the empty set. As already discussed, this trivial solution can be suppressed by restricting the optimization to a given region or by introducing additional data terms or constraints. Nevertheless, the shrinking bias still prevails in the sense that indentations and protrusions are energetically disfavored.

- The traditional linear model of data terms plus regularity requires choosing an appropriate weighting parameter (see Chapter 2 and 3). As discussed in [93], this weight cannot be estimated automatically as it defines the spatial scale at which reconstructions are to be computed. In other words, this linear model is *not scale invariant.* For a given object the quality of reconstruction highly depends on the spatial scale at which it is perceived. Usually, the appropriate parameter, balancing between data fidelity and smoothness, is set by the user. Yet, this is a cumbersome task which may require multiple test runs.

In this section, we propose an alternative energy model – the minimal ratio model – which avoids these limitations. Surprisingly, it is also amenable to convex relaxation.

If $F : V \to S^2$ denotes an outward normal field and $\rho : V \to (0,1]$ – a classical photoconsistency map, the minimal ratio model involves solving[3]

$$E(S) = -\frac{\displaystyle\int_S \langle N_S(s), F(s)\rangle \, ds}{\displaystyle\int_S \rho(s)\, ds} \to \min . \tag{5.24}$$

Since during minimization the above energy functional becomes negative, optimal surfaces aim at maximizing the numerator magnitude while minimizing the denominator. Minimization of (5.24) gives a 3D shape that optimally fulfills photoconsistency and normal field alignment criteria. Yet, in contrast to the traditional linear model, this is achieved without the specification of any parameters. Note that the functional in (5.24) exhibits a fairly different structure than the ones considered so far.

In the following, we discuss some important properties of the minimal ratio model.

Absence of a Shrinking Bias

We start with the shrinking bias inherent to the minimal surface model. It consists in the property that the model poses a strong preference to shapes of small overall surface area. The following proposition is a precise statement that ratio optimization does not exhibit a shrinking bias.

Proposition 9. *Let $S, S' \subset V$ be two arbitrary disjoint surfaces with the same ratio energy: $E(S) = E(S')$. Then, adding the surface S' to the surface S does not affect the overall energy, i. e. $E(S \cup S') = E(S)$.*

Proof. For simplicity, we will denote the numerator and denominator in (5.24) by $N(S) = -\int_S \langle N_S(s), F(s)\rangle \, ds$ and $D(S) = \int_S \rho(s)\, ds$.

From the condition $E(S) = E(S')$ we can derive

$$N(S') = \frac{D(S')N(S)}{D(S)}$$

and hence

$$\begin{aligned} E(S \cup S') &= \frac{N(S) + N(S')}{D(S) + D(S')} = \frac{N(S) + \frac{D(S')N(S)}{D(S)}}{D(S) + D(S')} \\ &= \frac{N(S)}{D(S)} = E(S). \end{aligned}$$

□

---

[3] Alternatively, we could use the proposed anisotropic minimal surface model in the denominator. Yet, we refuse to use it here as we want to explore the properties of both models separately.

The above claim states that one can iteratively expand a given surface by surface elements of the same ratio cost without affecting the total energy. Similarly, removing areas of the same ratio cost does not decrease the energy, which implies that the model has no shrinking bias. This is in contrast to the minimum surface model (see (5.8) and (5.9)), where the energy would simply double (or halve) when adding (or removing) same cost surface elements.

The absence of shrinking bias implies that the ratio optimization does not have any inherent preference for a specific geometry. Hence, its accuracy does not depend on the presence of protrusions or indentations on the reconstructed surface.

Scale Invariance

Next, we show that the minimal ratio model (5.24) is scale invariant.

Proposition 10. *For any arbitrary surface estimate* S *and scaled version* $S' = \gamma S$ *with* $\gamma > 0$, *the ratio energy remains unchanged, i. e.* $E(S) = E(S')$, *provided that the data remains fixed*

$$\begin{aligned} F(x) &= F(\gamma x) \\ \rho(x) &= \rho(\gamma x). \end{aligned}$$

Proof. Via change of variables, we obtain

$$\begin{aligned} E(S') &= -\frac{\displaystyle\int_{S'} \langle N_{S'}(s), F(s) \rangle \, ds}{\displaystyle\int_{S'} \rho(s) \, ds} \\ &= -\frac{\gamma^2 \displaystyle\int_{S} \langle N_{S}(s), F(s) \rangle \, ds}{\gamma^2 \displaystyle\int_{S} \rho(s) \, ds} = E(S). \end{aligned}$$

□

The above property of the model is particularly useful when applying a multiresolution scheme. In that case, the scale invariance guarantees that a correct solution is computed at each resolution level without the cumbersome need to adjust respective parameters. Note that using the same weighting parameter at all levels will not give a constant smoothing performance.

Remarks. While absence of a shrinking bias and scale invariance are useful properties, there are practical cases, where the proposed minimal ratio model will not give a satisfactory 3D model. For example, in the presence of strong noise in the data terms, one may wish to suppress it by increasing the regularization. Yet, the minimal ratio model does not explicitly take the noise level into account and always finds a scale-invariant shape, which may result in a noisy reconstruction. Analogously, if the observed object exhibits small-scale surface details, which have to be preserved, it makes sense to reduce the smoothing effect during the reconstruction process. In this case, ratio minimization could deliver an oversmoothed reconstruction. Nevertheless, the minimal ratio model stands out from all other models considered so far with the aforementioned properties. They make it appealing not only from a theoretical point of view, but also from a practical point of view, as they guarantee a reasonable result in a “non-degenerate” scenario without the need for specifying any parameters.

Continuous Global Optimization

While the above observations indicate that it is worthwhile studying minimal ratio formulations, the major computational challenge is to actually solve the optimization problem (5.24). A closer look at the ratio model doesn't suggest the existence of a straightforward convexification strategy. Surprisingly, a global minimum of (5.24) can still be obtained by solving a sequence of convex optimization problems, as we show in the sequel.

First, we apply the divergence theorem to convert the integral over the surface orientation to a regional integral over the surface interior. Thus, we obtain the following equivalent formulation

$$E(S) = -\frac{\displaystyle\int_{int(S)} \mathrm{div} F \, dx}{\displaystyle\int_S \rho(s) \, ds}, \tag{5.25}$$

where $int(S) \subset V$ denotes again the interior of S. The next step is a standard one – we switch to an implicit representation $u = \mathbf{1}_{int(S)}$, where $\mathbf{1}_{int(S)}$ signifies the characteristic function of $int(S)$, and relax the resulting binary condition. This yields

$$\begin{aligned} E(u) &= -\frac{\displaystyle\int_V \mathrm{div} F \cdot u(x) \, dx}{\displaystyle\int_V \rho(x) |\nabla u(x)| \, dx}, \\ \text{s. t.} &\qquad u \in [0,1]. \end{aligned} \tag{5.26}$$

Now, we come to a central observation. We note that the Dinkelbach's method for fractional optimization [30] can be applied to find the minimal ratio in (5.26). It consists in sequentially minimizing

$$\begin{aligned} G(u,\lambda) &= -\int_V \mathrm{div} F \cdot u(x) \, dx - \lambda \int_V \rho(x) |\nabla u(x)| \, dx, \\ \text{s. t.} &\qquad u \in [0,1] \end{aligned} \tag{5.27}$$

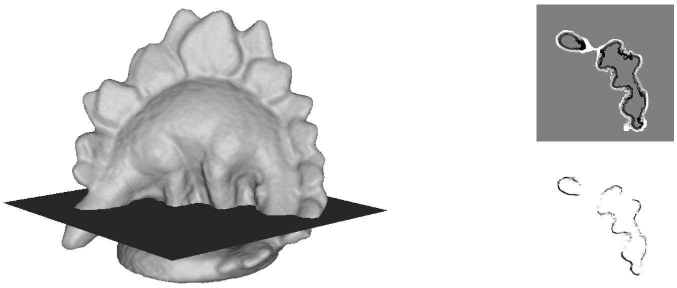
for different values of λ. Now, it is crucial to observe that, for a fixed parameter λ, (5.27) exhibits a constrained convex optimization problem of the form (3.4) that we regarded in Chapter 3. Hence, for each subproblem the global minimum can be obtained. Fortunately, the ability to optimize each subproblem globally implies the global optimizability of the original problem (5.26).

First, we present the minimization procedure before discussing its correctness. In summary, we have the following simple scheme for minimizing (5.26)

```
(0) Initialize: pick $u$ arbitrary and set $\lambda = E(u)$.
(1) Compute minimizer $u^*$ of $G(.,\lambda)$.
(2) Set $u := u^*$ and $\lambda := E(u^*)$.
(3) If $\lambda$ has decreased go to step (1), otherwise stop.
```

Now, we prove the correctness of the algorithm.

Proposition 11. *The above optimization procedure computes a (global) minimum of* (5.26).



cross-section **data terms**

Fig. 5.4: Volumetric data terms. Visualized are cross-sections through both utilized data volumes for the "dinoRing" image sequence (see Fig. 5.7). *Left:* Corresponding slice. *Right:* Inverted regional term $-\rho_{int}$ specifying the surface interior (above) and photoconsistency measure ρ (below) for the given volume slice. Intensity values correspond to estimated costs.

Proof. As before, we denote $N(u) = -\int_V \text{div}F \cdot u(x)\,dx$ and $D(u) = \int_V \rho(x)|\nabla u(x)|\,dx$. Upon convergence, the method gives a solution u_{min} and a ratio λ_{min} such that $\lambda_{min} = E(u_{min})$. Convergence implies that $G(u_{min}, \lambda_{min}) = 0$ and $u_{min} = \arg\min G(u, \lambda_{min})$. From these two statements it follows

$$0 = N(u_{min}) - \lambda_{min} D(u_{min}) \leq N(u) - \lambda_{min} D(u)$$

for all u that fulfill the constraints in (5.26). Hence, we obtain

$$\lambda_{min} \leq \frac{N(u)}{D(u)} = E(u)$$

for all feasible functions u, thus λ_{min} is the optimal ratio. □

So far, we have discussed the minimization of the relaxed problem (5.26). Since we are interested in finding minimizers of the original binary problem (5.25), a question about the validity of the thresholding theorem in this case arises. In fact, the thresholding theorem is applicable to each of the subproblems, which allows for its generalization to the minimal ratio model (5.25).

Theorem 10. *Let* $u^* : V \to [0,1]$ *be a global minimizer of the functional* (5.26)*. Then, the characteristic functions of all upper level sets (i.e. thresholded versions)*

$$\Sigma_{\mu,u^*} = \{x \in V \,|\, u^*(x) > \mu\}, \qquad \mu \in (0,1), \tag{5.28}$$

of u^* *are also global minimizers of* (5.26)*.*

Proof. As mentioned above, the proof is based on the observation that all subproblems involved in the optimization of (5.26) are of the general form (1.6) for which the above statement was proven (see Theorem 6). Hence, the above optimization procedure will still give a globally optimal solution if we replace the minimizer $u^\star$ in step (1) by a thresholded version $\mathbf{1}_{\Sigma_{\mu,u^\star}}$. □

Implementation and Numerics

In this section, we give more details on the particular choice of data terms and the numerical optimization of the proposed minimal ratio model.

Data Terms

Following the formulation in (5.24), we need to define two data measures: a photoconsistency map $\rho : V \to (0,1]$ and an outward normal field $F : V \to S^2$.
Once again, we used the voting scheme proposed in Chapter 3 to obtain a photoconsistency map ρ.
Instead of directly estimating a normal field F representing the shape orientation, we compute a regional term ρ_{int} assigned to the interior of the surface, which can be interpreted as divergence of a corresponding vectorfield (see (5.25)). In order to obtain meaningful orientations, ρ_{int} should be defined only in a vicinity of the surface with positive values inside and negative values outside of it. A respective vectorfield $F : V \to S^2$ with $\text{div}F = \rho_{int}$ can be derived as

$$\begin{aligned} F_1(x_1,x_2,x_3) &= \frac{1}{3}\int_0^{x_1} \rho_{int}(x_1',x_2,x_3)\,dx_1' \\ F_2(x_1,x_2,x_3) &= \frac{1}{3}\int_0^{x_2} \rho_{int}(x_1,x_2',x_3)\,dx_2' \\ F_3(x_1,x_2,x_3) &= \frac{1}{3}\int_0^{x_3} \rho_{int}(x_1,x_2,x_3')\,dx_3', \end{aligned} \tag{5.29}$$

where $F = (F_1\ F_2\ F_3)^T$. Note that this definition is unique up to the addition of a divergence-free vectorfield. In our implementation, we used the propagating scheme proposed in Chapter 3 to compute ρ_{int} within a band of 3 voxels around the hypothetical surface.
A real example of estimated data volumes ρ and $-\rho_{int}$ is depicted in Fig. 5.4.

Numerical Optimization

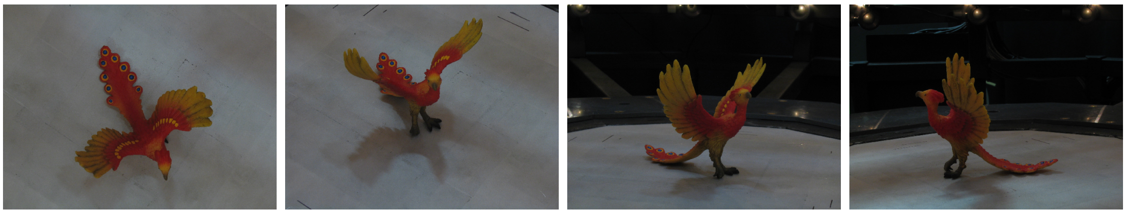
As mentioned previously, the minimization of (5.26) involves solving a sequence of convex optimization problems of the form (5.27). Following the discussion in Chapter 3, the primal-dual method offers an efficient numerical scheme for solving them. In our experiments, we apply this procedure by initializing in each step (consisting in solving one subproblem) with the result from the previous iteration. See Chapter 3 for more details.

5.4 Experiments

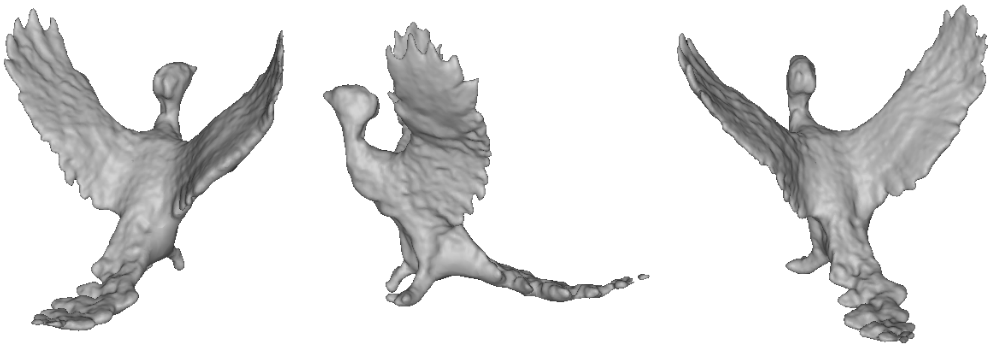
In this section, we explore experimentally the behavior of the anisotropic minimal surface model and the minimal ratio model.

Anisotropic Minimal Surface Model

In Fig. 5.5, the potential of the anisotropic minimal surface model (5.15) is demonstrated by comparing it to the isotropic model proposed in Chapter 4 as well as to a purely normal-based reconstruction approach [60]. The imaged object of a bird figurine is geometrically quite challenging due to the presence of thin structures like the wings as well as fine-scale details on them like the feathering. The normal information is able to capture a big part of the high-frequency content, but its large-scale accuracy suffers at some locations due to weak texture, occlusions or strong projective distortion. As a consequence, the Poisson reconstruction [60] exhibits a quite high degree of detailedness, but also some large-scale artifacts. Expectedly, the isotropic minimal surface model accurately recovers the wings and the legs due to the integration of silhouette constraints, but fails to retrieve the feathering as the photoconsistency map is incapable to encode such high-frequency information. Note that the photoconsistency



input images (4/21)



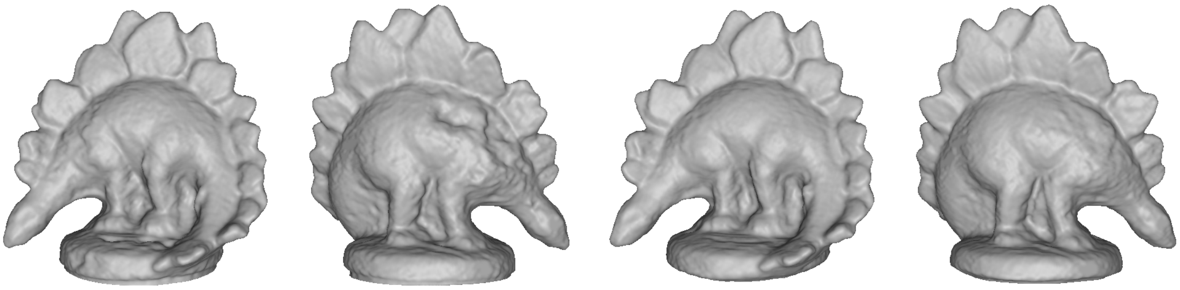
Poisson reconstruction [60]

isotropic minimal surface model

anisotropic minimal surface model

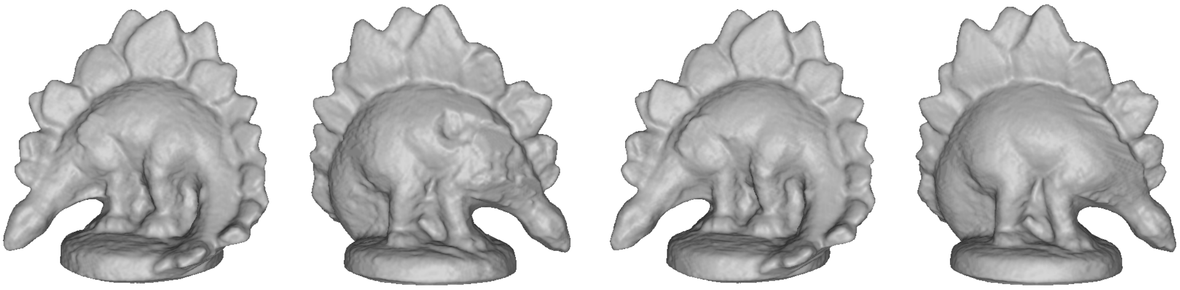
Fig. 5.5: Bird sequence. *First row:* 3 out of 21 input images of resolution 1024×768. *Second row:* Normal-based Poisson reconstruction obtained from the normal field in Fig. 5.3. *Third row:* Isotropic minimal surface reconstruction obtained from silhouettes and photoconsistency information (see Chapter 4). *Fourth row:* Anisotropic minimal surface reconstruction obtained from silhouettes, photoconsistency and normal information by minimizing (5.15).

is practically not meaningful for surface pieces of thickness up to 2 voxels. In contrast, the proposed anisotropic minimal surface model (5.15) delivers a genuinely accurate reconstruction



dinoSparseRing (16 images) **dinoRing (48 images)**

isotropic model



anisotropic model

Fig. 5.6: Dino sequences. See Fig. 3.6 for example input images. *First row:* Reconstruction with the isotropic minimal surface model (see Chapter 4). *Second row:* Reconstruction with the proposed anisotropic model (5.12) which additionally takes surface normal information into account. Note that the incorporation of normal information generally improves the reconstructions, especially at locations of weak texture, even though the improvements are minor.

	isotropic model	anisotropic model
dinoSparseRing	0.53 mm / 98.3 %	**0.48 mm / 98.6 %**
dinoRing	0.43 mm / 99.4 %	**0.42 mm / 99.5 %**

Tab. 5.1: Quantitative evaluation of the reconstructions in Fig. 5.6.

result by integrating silhouette, photoconsistency and normal information. In particular, all relevant surface details as well as a big portion of the fine-scale structure are recovered. A multi-modal formulation seems to be the best recipe to address objects of the given complexity. We proceed with a quantitative evaluation of the anisotropic model in (5.12) on the already encountered “dinoSparseRing” and “dinoRing” image sequences. The reconstructions with energy model III proposed in Chapter 3 and the anisotropic model in (5.12) are depicted in Fig. 5.6 and quantitative numbers are given in Table 5.1. The experiment demonstrates that the incorporation of normal information generally improves the reconstructions, especially at locations of weak texture, even though the improvements are minor. This is not surprising since the regional terms play a decisive role in this model, i. e. the normal information is used only to enhance the reconstructions.

Minimal Ratio Model

Now, we focus on the minimal ratio model given in (5.24).

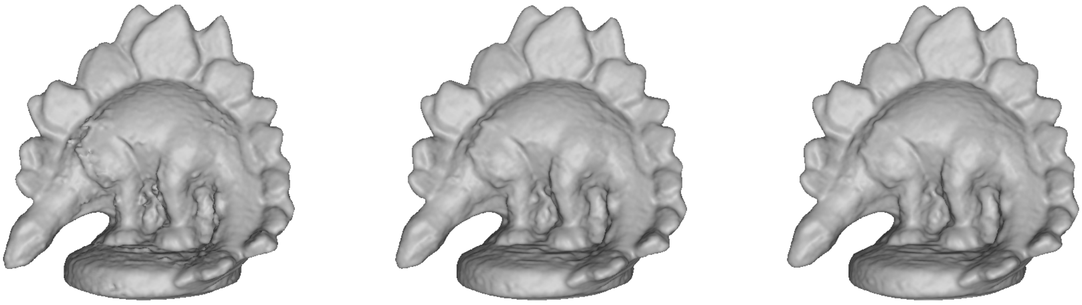


Fig. 5.7: Ratio minimization for the “dinoRing” image sequence. Reconstructions at different iteration steps and corresponding ratios. The leftmost 3D model poses the initialization, the middle one – an intermediate reconstruction and the rightmost – the final one. Note that the amount of regularization progressively increases during the optimization.

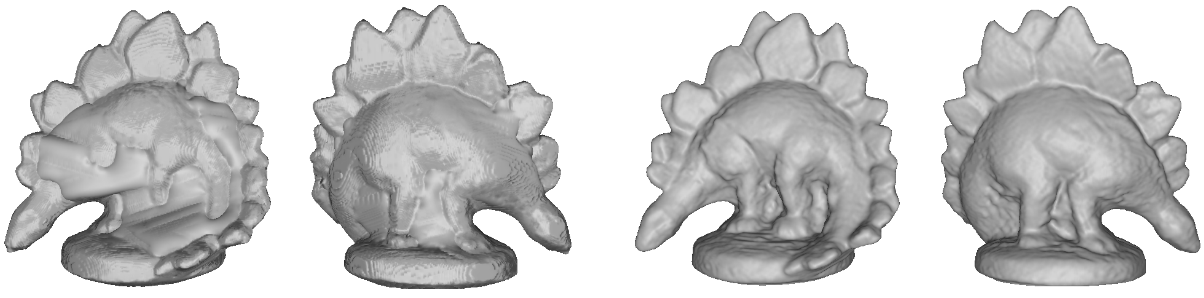


Fig. 5.8: Minimal surface model vs. minimal ratio model (“dinoRing” image sequence). Note that the minimal surface model produces clear oversmoothing effects by filling in deep concavities (e.g. at the legs). In contrast, the minimal ratio model, which is free of shrinking bias, accurately recovers the complete geometry.

Fig. 5.7 shows reconstructions at different iteration steps during the process of ratio minimization for the “dinoRing” image sequence. Note that the amount of regularization progressively increases during the optimization while the ratio number decreases. The rightmost 3D model poses the final result at ratio -2.49476. It was reached after 6 iteration steps by starting from an initial ratio of -0.0001. Recall that the final reconstruction exhibits an optimal 3D shape in terms of scale invariance.

As previously mentioned, the absence of a shrinking bias is an important property of the minimal ratio model. In Fig. 5.8, we demonstrate this property experimentally on the “dinoRing” sequence. The data set seems to be a very challenging test scenario for the minimal surface model which produces clear oversmoothing effects by filling in deep concavities (e. g. at the legs). In contrast, the minimal ratio model, which is free of shrinking bias, accurately recovers the complete geometry. Note that none of the compared models uses a weighting parameter allowing to control the amount of desired smoothing. On that condition, the minimal surface model strongly depends on the accuracy of the estimated photoconsistency measure and on the geometry of the recovered shape. In contrast, the success of the minimal ratio model is not affected by the particular structure of the reconstructed object.

Computational Time

Obviously, the proposed energy models are suitable for a GPU implementation as they could be considered as generalizations of already encountered formulations. On a PC with 2.83 GHz, equipped with a NVIDIA Tesla C2070 graphics card, we measured computational times of 14s for the reconstruction of the bird figurine (see Fig. 5.5) with the anisotropic minimal surface model and 3.1s for the reconstruction of the dinosaur figurine (see Fig. 5.7) with the minimal ratio model. Note that the computational time of the numerical scheme in (5.23) is dominated by the realization of the respective projections. Note also that each iteration of the ratio minimization is initialized with the result from the previous iteration, which significantly reduces the overall computational costs.

5.5 Discussion

In this chapter, we extended the class of energy models amenable to convex optimization by including anisotropic metrics and minimal ratios.

We proposed a specific anisotropic metric allowing to integrate provided surface normal information in a soft manner. Based on it, we generalized previously proposed variational formulations to include this additional source of information. In particular, we built an anisotropic minimal surface model completed by exact silhouette constraints to obtain a 3D shape that optimally fulfills silhouette, photoconsistency and normal alignment criteria. Additionally, we extended a previously proposed energy model fusing data terms and a regularizer by incorporating a more general smoothing scheme aware of directional selectivity. We proved that both models can be optimized globally by means of convex relaxation and thresholding techniques. In experiments on real-world data, we demonstrated that stereo-based reconstruction results can be enhanced both qualitatively and quantitatively by incorporating normal information.

Moreover, we proposed a continuous framework for ratio optimization based on relaxation and sequential convex optimization providing a globally optimal solution. In a theoretical investigation, we explored some useful properties of the minimal ratio model, in particular scale invariance and absence of a shrinking bias. The absence of a shrinking bias was confirmed experimentally on a real-world data set by performing a comparison to the classical minimal surface model.

6

Vesicle Membrane Reconstruction from Fluorescence Imaging

Real is what can be measured.

Max Planck (1858-1947)

So far, we have explored convex models for multiview 3D reconstruction. In this chapter, we focus on another application related to it – 3D segmentation. We will see that convex optimization offers a powerful tool to design robust and accurate methods also within this domain.

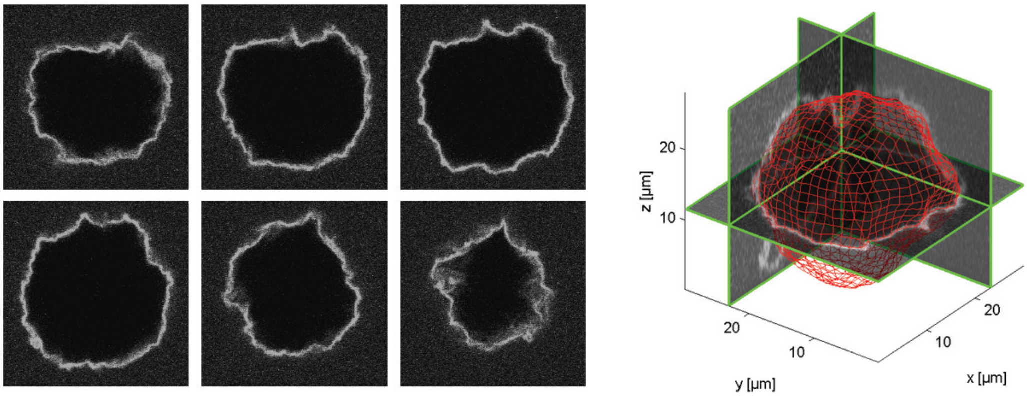
6.1 Motivation

Convex optimization proved useful for solving the image-based modeling problem. We considered different convex formulations fusing various cues like silhouettes, multiview stereo and surface normal information. We demonstrated the practical value of all of them in terms of accuracy and robustness. A closer look at these variational approaches reveals that they follow the same modeling strategy – appropriate volumetric data terms are defined first and subsequently a 3D surface is obtained by solving a 3D segmentation problem on the underlying volume with respect to the given criteria. Obviously, this optimization machinery can be applied to solve directly a 3D segmentation problem, where measured volume data is provided by a respective application. An example poses the reconstruction of a vesicle membrane deformation process under osmotic pressure (see Fig. 6.1). Thereby, the membrane surface can be measured by exploiting fluorescence properties of the material in a suitable experimental environment. Yet, typical measurements are noisy due to the imprecision of the utilized technology. For example, interior and exterior regions could exhibit very similar intensity characteristics while the membrane gauging is usually genuinely blurry (see Fig. 6.1). This makes the 3D segmentation problem quite challenging and requires a robust and powerful optimization scheme. Once again, we revert to convex modeling which proves viable to accomplish the given task.

6.2 Previous Work

The segmentation problem is usually addressed in the context of processing one single image instead of a 3D volume. Image segmentation is one of the fundamental problems in computer vision and has been extensively studied for decades. It is beyond the scope of the current work to discuss the myriad of existing works. Here, we focus only on the most closely related approaches.

Among the existing segmentation techniques variational methods prevail due to their accuracy, robustness and mathematical elegance. Since the pioneering work of Kass et al. [59]



input volume slices

superimposed segmentation

Fig. 6.1: Volumetric 3D segmentation. *Left:* Multiple input volume slices, *right:* superimposed segmentation result. The volume data captures the deformation of a vesicle membrane under osmotic pressure, imaged by means of fluorescence microscopy. Note that the data set is quite challenging due to the similarity of the intensity characteristics of interior and exterior regions as well as the noisy membrane measurements.

considerable efforts have been made on exploring different energy models relying on various image cues. Three paradigms have proven to be particularly useful: regional statistics, edge terms and orientation information.

One of the seminal works demonstrating the potential of regional intensity statistics is [95]. Therein, the segmentation problem is posed in terms of a piecewise smooth or piecewise constant image approximation with minimal boundary length. More specifically, in the more general case the following energy functional is being minimized

$$\begin{aligned} E(C,f,g) = \alpha \left(\int_{int(C)} (f(x) - I(x))^2 \, dx \; + \int_{\Omega \setminus int(C)} (g(x) - I(x))^2 \right) + \\ \mu \left(\int_{int(C)} |\nabla f|^2 \, dx \; + \int_{\Omega \setminus int(C)} |\nabla g|^2 \, dx \right) + \nu \int_C ds, \end{aligned} \tag{6.1}$$

where $I : \Omega \subset \mathbb{Z}^2 \to \mathbb{R}$ denotes the given intensity image, the functions $f, g : \Omega \to \mathbb{R}$ model the foreground and background intensity characteristics, respectively, and $\alpha, \mu, \nu \in \mathbb{R}$ are constants. The above functional is minimized with respect to the functions f and g as well as a one-dimensional parametric contour $C : \theta \subset \mathbb{R} \to \Omega$ separating the image domain into two regions $int(C)$ and $\Omega \setminus int(C)$. This methodology was generalized in [148] and [63], where regional intensity statistics is estimated to achieve an optimal region assignment

$$E(C, \alpha_1, \alpha_2) = - \int_{int(C)} \log P(I(x) \mid \alpha_1) \, dx \; - \int_{\Omega \setminus int(C)} \log P(I(x) \mid \alpha_2) \, dx \; + \nu \int_C ds. \tag{6.2}$$

Thereby, α_1 and α_2 summarize the statistical parameters for the two regions. Crucial in this formulation is the modeling of the corresponding probability maps. They could encode not only characteristic mean intensity values, but also respective variances. The segmentation is obtained by alternatingly minimizing the underlying energy functional with respect to the contour and the parameters of the probability models. In [110] and [129], the region parameters are estimated interactively, which allows for globally optimal segmentations according to the specified user input.

Although the reliance on regional image criteria entails considerable robustness, it often results in imprecise or oversmoothed segmentations. In order to account for this drawback and increase the accuracy of the segmentation, researchers suggested the use of edge terms attracting the curve (in 2D) or the surface (in 3D) towards the locations of sudden brightness changes that commonly arise at object boundaries. Geodesic active contours [19, 62] pose the segmentation problem as a weighted minimal surface model, where the weights reflect a local Riemannian metric based on the magnitude of the image gradient and therefore account for the presence of edges. More specifically, the following energy functional is considered

$$E(C) = \int_C g(s)\, ds, \tag{6.3}$$

where the weighting function $g : \Omega \to \mathbb{R}$ is defined as

$$g(x) = \frac{1}{1 + |\nabla I(x)|^p} \tag{6.4}$$

with $p = 1$ or 2. Note that the model in (6.3) is an analog to the stereo-based weighted minimal surface model (3.1).

While the incorporation of edge terms can result in substantial improvements of object boundary alignment, it is still unable to handle thin or small-scale geometric details. To this end, Vasilevskiy and Siddiqi [130] demonstrated the potential of using orientation information to segment narrow elongated structures like blood vessels. The curve (or surface) outward orientation is provided in the form of a dense vectorfield $v : \Omega \to S^2$ defined on the image domain. A typical choice for the normal field is the normalized or unnormalized image gradient. The segmentation is finally obtained by minimizing

$$E(C) = -\int_C \langle N_c(s), v(s)\rangle\, ds, \tag{6.5}$$

where $N_C \in S^2$ denotes the outward contour normal. Note that the model in (6.5) is an analog of the 3D reconstruction model in (5.2).

6.3 Contribution

In this chapter, we propose a unified framework integrating regions, edges and orientation with the goal of combining the advantages of all utilized paradigms to achieve robust and accurate 3D segmentations regardless of the available geometric structure. Although the presented model is built upon established concepts, it differs from alternative methods in their concrete realization. In particular, we derive an anisotropic formulation, tailored to the 3D segmentation problem at hand, and a specific edge attraction term. We show that a globally optimal solution of the constructed model can be obtained by means of convex optimization. The particular structure of the proposed framework as well as the benefits from the convex optimization in this context are demonstrated experimentally by comparisons to alternative state-of-the-art segmentation approaches.

The main results in this chapter are published in [70].

6.4 Anisotropic Energy Model Integrating Regions, Edges and Orientation

This section introduces the underlying energy functional accounting for various complementary information sources encoded in the data. As previously mentioned, the proposed model imposes regional intensity subdivision, edge alignment and orientation integration.

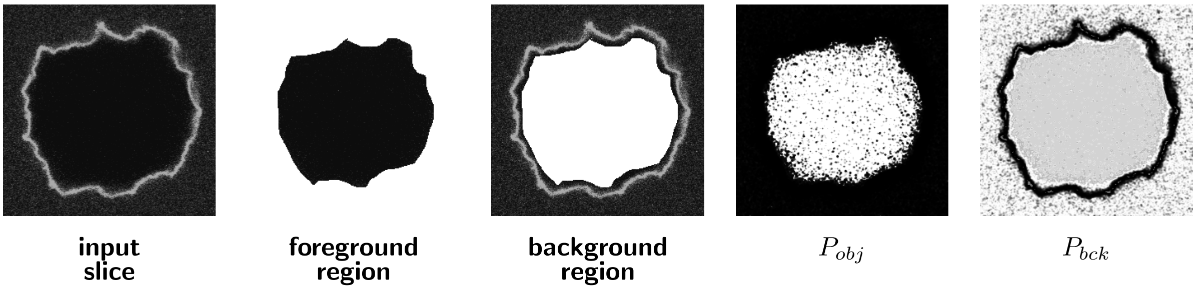


Fig. 6.2: Regional subdivision via interactive user input. The user is required to roughly segment one of the slices of the volume data. Based on the provided segmentation, foreground/background histograms are built which are, in turn, used to assign intensity likelihoods to each voxel conditioned on being inside or outside the membrane surface. An example is illustrated for one of the slices in Fig. 6.1. Note that due to the noisy input data and the similarity in the intensity characteristics of both regions these likelihoods alone do not allow a reliable separation of object and background.

We start with some notations. Let $I : V \subset \mathbb{R}^3 \longrightarrow [0,1]$ denote the measured volumetric data (see Fig. 6.1 for an example). The goal is to partition the volume into two subsets by identifying the vesicle membrane surface $S \subset V$. We will indicate the interior region according to a certain surface estimate S as $int(S)$ and the exterior one as $ext(S)$, respectively.

The user is required to specify a rough identification of both regions by marking them in one of the input volume slices (Fig. 6.2). This user input is used to build initial foreground/background histograms which are adaptively refined during the evolution process. To this end, a precise segmentation is not needed. Alternatively, strokes or scribbles specifying both regions could be provided. After smoothing and normalization, derived histograms are used to define probabilities $P_{obj}, P_{bck} : V \longrightarrow [0,1]$ for observing a certain intensity, given that the respective voxel is inside or outside the membrane surface. However, due to the noisy input data and the similarity in the intensity characteristics of both regions these intensity likelihoods do not permit a reliable separation of object and background. Based on the probabilistic volume subdivision, a surface estimate is desired which separates voxels with high values for P_{obj} and low values for P_{bck} from voxels with opposite characteristics. Simultaneously, favorable shape properties like smoothness should be imposed. These observations naturally lead to the following energy model

$$E(S) \;=\; -\mu \left(\int_{int(S)} \log P_{obj}(x)\, dx + \int_{ext(S)} \log P_{bck}(x)\, dx \right) \;+\; \int_S ds, \tag{6.6}$$

where $\mu \in \mathbb{R}_{\geq 0}$ is a parameter controlling the weighting of data term and regularization. For a good surface estimate P_{obj} is expected to be close to 1 within the interior region $int(S)$ and P_{bck} close to 1 within the exterior region $ext(S)$. Hence, the overall energy costs will be low. Analogously, for inaccurate estimates the corresponding probability values will generally vary, which after applying the logarithm would result in high data fidelity costs. In effect, minimization of the functional in (6.6) specifies a shape with the desired properties. Note that both foreground/background histograms are updated permanently by reinitializing them with the voxel values within the current regions $int(S)$ and $ext(S)$. This adaptive procedure allows to adjust the histograms to the input data rather than relying on the specified user input during the entire optimization process.

Unfortunately, regional intensity statistics is not sufficient to achieve the required precision of the extracted surface due to the noise in the estimated probability maps (see Fig. 6.2). Yet, a closer inspection of the volume data reveals that the boundary of the interior region can

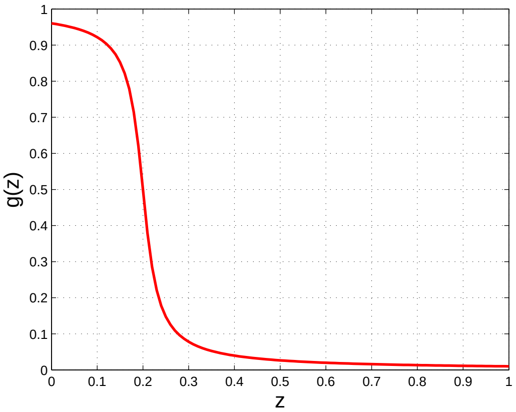


Fig. 6.3: Edge attraction function. Plotted is the proposed function $g(z) = \frac{1}{2} - \frac{\mathrm{atan}(\frac{z-\beta_1}{\beta_2})}{\pi}$ with parameters $\beta_1 = 0.2$ and $\beta_2 = 0.025$.

be identified as areas, where the intensity changes from dark to bright. This requirement can be imposed by encouraging the shape normal at each point to align with the local intensity gradient. Taking this into account, we come up with the following energy model

$$\begin{aligned} E(S) &= -\mu_1 \left(\int_{int(S)} \log P_{obj}(x)\, dx + \int_{ext(S)} \log P_{bck}(x)\, dx \right) \\ &\quad -\mu_2 \int_S \langle N_S(s), \nabla I(s) \rangle\, ds \;+\; \int_S ds, \end{aligned} \tag{6.7}$$

where N_S denotes the (normalized) outward surface normal map. Again, $\mu_1, \mu_2 \in \mathbb{R}_{\geq 0}$ are constant weighting parameters. The second term in (6.7), referred in the literature as flux [130], favors the alignment of the surface orientation with the vectorfield defined by ∇I. In the ideal case, when shape normals exactly coincide with the intensity gradient, the corresponding costs will be negative and maximal in magnitude. On the other hand, surface normals pointing in the opposite direction will cause positive energy costs scaled by the respective deviation. It is important to note that abrupt intensity changes in the input data, i. e. areas of high gradient, can be observed not only at the boundary between vesicle membrane and interior region, but also at the boundary between membrane and exterior region. However, the estimated region statistics is additionally used to remove this ambiguity. In fact, regional information and flux are exploited as complementary cues. The balancing between them via the weighting parameters μ_1 and μ_2 should be established based on the confidence in both information sources in the particular situation.

The energy model in (6.7) allows for further extensions by exploiting the fact that the vesicle membrane can be identified in the input volume data by locations of relatively high intensity due to its fluorescence properties. A simple but very effective way to take this observation into account is to replace the Euclidean metric used by the regularization term in (6.7) by a more general Riemannian measure based on the local data content. This yields the following functional

$$\begin{aligned} E(S) &= -\mu_1 \left(\int_{int(S)} \log P_{obj}(x)\, dx \;+\; \int_{ext(S)} \log P_{bck}(x)\, dx \right) \\ &\quad -\mu_2 \int_S \langle N_S(s), \nabla I(s) \rangle\, ds \;+\; \int_S g(I(s))\, ds, \end{aligned} \tag{6.8}$$

where $g : [0,1] \to \mathbb{R}_{\geq 0}$ is a non-negative, monotonically decreasing function. The new weighted minimal surface term alleviates the smoothing and encourages the shape to pass through locations of low costs according to g. To this end, the function should be monotonically decreasing in order to assign low costs to bright locations which vote for the presence of membrane edges and to penalize dark regions specifying interior and exterior. Yet, this is not the only requirement that should be followed when designing the function g. When conducting different measurements, the particular intensity distribution of the volume data, including the brightness characteristics of the vesicle membrane, may vary. Hence, from a practical point of view, it is preferable to provide the user with the flexibility to specify the particular complexion of the weighting function g. We propose the following definition

$$g(z) = \frac{1}{2} - \frac{\operatorname{atan}(\frac{z-\beta_1}{\beta_2})}{\pi}, \tag{6.9}$$

where $\beta_1 \in [0,1], \beta_2 \in (0,1]$ are constant parameters. The parameter β_1 determines the location of the inflection point of the function while β_2 controls the smoothness of the falling edge (see Fig. 6.9). In effect, the value of β_1 indicates the lower limit of the range in which a pixel is categorized as "bright". Large values will restrict the range and classify only high intensities as vesicle membrane. The parameter β_2 can be interpreted as a kind of tolerance of the estimated intensity splitting. Large values will give a smooth function and will smear the boundary while small values will sharpen the splitting.

Although the energy functional in (6.8) incorporates a broad variety of data features, it still does not address one relevant point. The applied scanning technique produces distorted volume data in a sense that the spacing in different directions is not equal. In fact, the resolution and, thus, the sampling in the z-direction is significantly coarser than along the other two spatial dimensions. As a consequence, the isotropic weighted minimal surface term $\int_S g(I(s))\,ds$, which imposes the shape regularity, treats the slicing direction with the same priority as the other two, which leads to a different amount of smoothing due to the unequal spacing. Hence, different scaling is needed along this particular direction, which is achieved by the following anisotropic generalization

$$\begin{aligned} E(S) &= -\mu_1 \left(\int_{int(S)} \log P_{obj}(x)\,dx + \int_{ext(S)} \log P_{bck}(x)\,dx \right) - \mu_2 \int_S \langle N_S(s), \nabla I(s) \rangle\,ds \\ &+ \int_S g(I(s)) \sqrt{N_S^T(s) D(s) N_S(s)}\,ds \end{aligned} \tag{6.10}$$

where $\left\{ D(x) \in \mathbb{R}^{3\times 3} \mid x \in V \right\}$ denotes a family of positive semidefinite symmetric tensors tolerating certain directional selectivity. Compare the above formulation to the discussion of anisotropic minimal surfaces in Section 5.2. In addition to encouraging the surface to pass through preselected locations according to the edge map g, the anisotropic energy model favors certain shape orientations while suppressing others. This can be achieved with the appropriate design of the tensors $D(x)$. Note that D is defined pointwise. However, in the remainder of this section, we will omit the argument for the sake of simplicity. Our intention is to define D in a way that allows for different scaling in the z-direction. This aspect can be modeled by setting

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \tag{6.11}$$

where $\gamma \in [0,1]$ is a constant parameter. Ideally, γ should be set to the squared ratio between the sampling in the z-direction and the x/y-direction, respectively. Note again that for $\gamma = 1$ we obtain the original isotropic model (6.8).

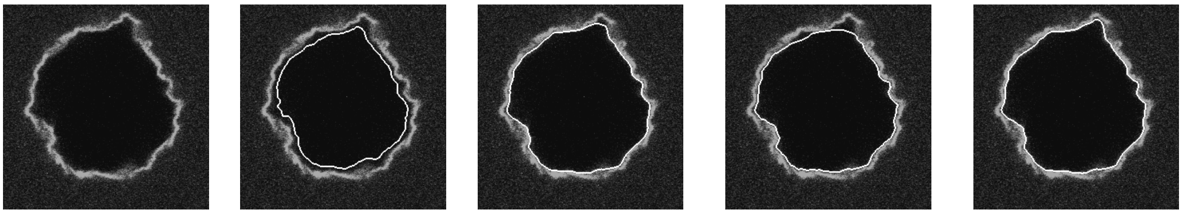


Fig. 6.4: Contribution of different data features. Results of various energy models obtained by combining different cues applied to the 3D data set in Fig. 6.1. The segmentation is visualized for slice #38 and superimposed on the input data. Note that only the complete model (6.10) – shown on the right – yields tight and accurate segmentation results.

Now, the final model defined in (6.10) contains all constructs necessary to achieve our primary goal. The relevance of the various features used in the model depends on the particular input data. In cases, where one of the cues prevails, the corresponding energy term will play a dominant role, which should be reflected by an adequate parameter setting. However, in a typical practical situation, all utilized paradigms will have a certain contribution to the final segmentation result. An example demonstrating the contribution of the different data features is illustrated in Fig. 6.4. Superimposed segmentation results for one of the slices with different models, obtained by omitting one of the cues, are visualized. The regional terms alone, which are based on the specified user input, are not sufficient to produce accurate segmentations, even in combination with a flux term. This is due to the blurred membrane measurements resulting in high intensity characteristics in the vicinity of the surface (see Fig. 6.2). Note that a more accurate user input will not drastically improve the segmentations since more bright pixels will enter the foreground histogram and diminish the discriminative power of the regional terms. Substantial increase of the accuracy of segmentation can be observed by adding the edge attraction term to the model. Yet, the absence of a flux term is still noticeable, especially at small-scale contour structures. Finally, the effect of the anisotropic generalization is investigated by comparing it to the original isotropic model (i. e. for $\gamma = 1$). As expected, the isotropic regularizer produces significant oversmoothing at areas of abrupt slice-to-slice changes. This can be suppressed by reducing the overall regularization, but only at the cost of degrading the contour alignment along the slice-plane. In summary, regional terms and edge term seem to be decisive for the overall quality of the segmentation results, but the flux term and the anisotropic scaling also contribute significantly to the precision of the estimated boundary.

6.5 Variational Minimization

This section deals with the minimization of the energy functional proposed in (6.10).
First, we observe that the flux term in (6.10) can be rewritten as a regional term over the surface interior via the *divergence theorem*. Now, we end up with the following energy functional equivalent to (6.10)

$$\begin{aligned} E(S) &= -\int_{int(S)} \left(\mu_1 \log P_{obj}(x) + \mu_2\, \Delta I(x)\right)\, dx \,-\mu_1 \int_{ext(S)} \log P_{bck}(x)\, dx \\ &+ \int_S g(I(s)) \sqrt{N_S^T(s) D(s) N_S(s)}\, ds. \end{aligned} \tag{6.12}$$

Note that $\mathrm{div}(\nabla I) = \Delta I$. Basically, the above formulation states that a flux term can be incorporated by adequately adjusting the interior region statistics. This observation will considerably facilitate further optimization.

Solution I: Minimization by Level Sets

The level set method [29, 98] is an established technique for minimizing energy functionals for shape estimation like (6.12). The key idea is to represent the surface implicitly by a function $\phi : V \to \mathbb{R}$ whose values reflect the signed distance from it, where negative values indicate the interior region and non-negative values - the exterior, respectively. The formulation proposed in this section is based on the framework of [22]. Note, however, that the model in (6.12) surpasses the one in [22] in its completeness. In particular, the anisotropic regularization scheme poses the main challenge when developing a valid generalization.
The construction of an implicit surface representation implies that both interior/exterior regions can easily be accessed by means of the Heaviside function

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6.13}$$

In the level set framework, a surface deformation equation is converted to an evolution equation with respect to the implicit function ϕ. Hence, topological changes are handled automatically without additional efforts. By means of the Heaviside function, we can rewrite the functional in (6.12) as a functional over the set of implicit functions, which yields

$$\begin{aligned} E(\phi) &= -\int_V (\mu_1 \log P_{obj}(x) + \mu_2\, \Delta I(x))\, (1 - (H \circ \phi)(x))\; dx \; - \mu_1 \int_V \log P_{bck}(x) (H \circ \phi)(x)\, dx \\ &+ \int_V g(I(x)) \sqrt{\nabla (H \circ \phi)(x)^T D(x) \nabla (H \circ \phi)(x)}\, dx. \end{aligned} \tag{6.14}$$

The terms $1 - (H \circ \phi)(x)$ and $(H \circ \phi)(x)$ in the above formulation act as filtering functions by specifying interior and exterior voxels, respectively. Note that the composite function $H \circ \phi$ is a binary labeling function identifying both regions. More precisely $H \circ \phi \equiv \mathbf{1}_{ext(S)}$, where $\mathbf{1}_{ext(S)}$ denotes the indicator function of the exterior region. Note also the relations $N_S(x) = \frac{\nabla(H \circ \phi)(x)}{|\nabla(H \circ \phi)(x)|}$ and $ds = |\nabla(H \circ \phi)(x)|\, dx$ which give rise to the last term in (6.14).
The first step towards minimizing the functional in (6.14) is to compute its Euler-Lagrange equation. We obtain (the argument x is omitted for simplicity)

$$\begin{aligned} \frac{\partial E}{\partial \phi} &= \delta(\phi) \left(\mu_1 \log P_{obj} + \mu_2\, \Delta I - \mu_1 \log P_{bck}\right) + \delta'(\phi) g(I) \sqrt{\nabla \phi^T D \nabla \phi} \\ &- \mathrm{div} \left(\delta(\phi) g(I) \frac{D \nabla \phi}{\sqrt{\nabla \phi^T D \nabla \phi}} \right) \\ &= \delta(\phi) \left(\mu_1 \log P_{obj} + \mu_2\, \Delta I - \mu_1 \log P_{bck}\right) + \delta'(\phi) g(I) \sqrt{\nabla \phi^T D \nabla \phi} \\ &- \delta'(\phi) g(I) \frac{\nabla \phi^T D \nabla \phi}{\sqrt{\nabla \phi^T D \nabla \phi}} - \delta(\phi) \mathrm{div} \left(g(I) \frac{D \nabla \phi}{\sqrt{\nabla \phi^T D \nabla \phi}} \right) \\ &= \delta(\phi) \left(\mu_1 \log P_{obj} + \mu_2\, \Delta I - \mu_1 \log P_{bck}\right) - \delta(\phi) \mathrm{div} \left(g(I) \frac{D \nabla \phi}{\sqrt{\nabla \phi^T D \nabla \phi}} \right). \end{aligned} \tag{6.15}$$

Thereby, $\delta(.)$, which denotes the Dirac delta function, appears as the derivative of the Heaviside function, i. e. $\delta(z) = H'(z)$. It is not surprising that all terms in the above expression are scaled by $\delta(\phi)$ since only the 0-level of ϕ representing the surface is of interest. In practice, $\delta(\phi)$ is replaced by a smoothed version $\delta_\epsilon(\phi)$ for numerical reasons.
The functional in (6.14) can be minimized via gradient descent by solving the following partial differential equation

$$\phi_t = -\frac{\partial E}{\partial \phi}. \tag{6.16}$$

Numerically, the above equation can be solved by applying classical discretization schemes like the forward Euler method. It should be noted that after each update, ϕ will no longer reflect the exact distance to the surface. For that reason, the function should permanently be reinitialized in order to circumvent numerical difficulties.
Although the level set method leads to a tractable numerical scheme and avoids the need for explicit handling of topological changes, it suffers from one important limitation. Due to the binary nature of the representation, whereas the localization of the surface is implicitly encoded by the sign of a distance map ϕ, this technique always involves optimization over a set of binary labeling configurations. As a consequence, due to the minimization over a non-convex domain any iterative procedure can get stuck in a undesired local minimum that may be far from the expected solution. In effect, the final result strongly depends on the initialization.

Solution II: Minimization by Convex Relaxation

In this paragraph, we show that convex optimization poses a competitive alternative to the level set method for minimizing the underlying cost functional (6.10).
First, we observe that the implicit function ϕ in (6.14) always appears in concatenation with the binary Heaviside function H. Hence, an equivalent formulation can be obtained by substituting $u = H \circ \phi$. This leads to the following minimization problem

$$\begin{aligned} E(u) &= -\int_V (\mu_1 \log P_{obj}(x) + \mu_2\, \Delta I(x))\,(1-u(x))\,dx \;-\mu_1 \int_V \log P_{bck}(x) u(x)\,dx \\ &+ \int_V g(I(x))\sqrt{\nabla u(x)^T D(x) \nabla u(x)}\,dx \to \min, \\ &\quad \textbf{s. t.} \quad u : V \to \{0,1\}\,. \end{aligned} \tag{6.17}$$

Now, one can notice that the above functional is convex, but the minimization is carried out over the set of all binary labeling functions which is not convex. As a result, we have a constrained non-convex optimization problem and obtaining an exact solution is not straightforward. Once again, we relax the domain of feasible functions and allow also intermediate values in the interval $[0,1]$ as an effective remedy to this problem. Thus, we obtain a more tractable constrained convex optimization problem

$$\begin{aligned} E(u) &= -\int_V (\mu_1 \log P_{obj}(x) + \mu_2\, \Delta I(x))\,(1-u(x))\,dx \;-\mu_1 \int_V \log P_{bck}(x) u(x)\,dx \\ &+ \int_V g(I(x))\sqrt{\nabla u(x)^T D(x) \nabla u(x)}\,dx \to \min, \\ &\quad \textbf{s. t.} \quad u : V \to [0,1]\,. \end{aligned} \tag{6.18}$$

Now, we have, as desired, a convex functional optimized over a convex domain. However, recall that we are interested in solving the original "binary" problem (6.17). Again, we make use of the thresholding theorem.

Theorem 11. *If* $u^* : V \to [0,1]$ *is a minimizer of the functional in* (6.18)*, then for almost any threshold* $\nu \in (0,1)$ *the binary function* $\mathbf{1}_{\Sigma_{\nu,u^*}}(x) : V \to \{0,1\}$ *with* $\Sigma_{\nu,u} = \{x \mid u(x) > \nu\}$ *is also a (global) minimizer of* (6.18)*.*

Proof. The statement follows directly from Theorem 6 and the fact that the functional in (6.18) is a special case of the one in (1.6) (when omitting the constant part). □

In other words, solving the minimization problem (6.17) boils down to solving (6.18). In the following, we focus on the numerical minimization of (6.18).

In Section 5.2, we proposed a general anisotropic minimal surface model and discussed different variational formulations building on it. As the current functional at hand (6.18) contains an anisotropic regularizer, the numerical scheme in (5.23) can be adapted for its minimization. However, in the sequel we propose an alternative procedure – a generalization of the linearized fixed-point iteration method based on Successive Overrelaxation (SOR) introduced in Chapter 3.

The first step is to set up the Euler-Lagrange equation which states a necessary condition for a minimum of (6.18) (the argument x is omitted for simplicity)

$$0 \;=\; \mu_1 \left(\log P_{obj} - \log P_{bck}\right) + \mu_2 \Delta I - \operatorname{div}\left(g(I)\frac{D\nabla u}{\sqrt{\nabla u^T D \nabla u}}\right). \tag{6.19}$$

The key idea is to solve the above equation directly instead of deriving an evolution PDE like (6.16). Yet, (6.19) exhibits a non-linear differential equation which is not trivial to be solved directly. The source of non-linearity is given by the diffusivity $\rho = \frac{g(I)}{\sqrt{\nabla u^T D \nabla u}}$. Thus, a straightforward way to address this difficulty is to apply a linearization scheme. Starting with an initialization $u^0 = 0.5$, we can compute ρ and keep it constant. For constant ρ, (6.19) is linear and spatial discretization yields a sparse linear system of equations which can be solved efficiently with iterative approaches like SOR. Similar to (3.36), the k-th update of u at voxel i is computed as

$$u_i^{(l,k+1)} = (1-\omega)u_i^{(l,k)} + \omega \frac{\sum\limits_{j \in \mathcal{N}(i), j<i} \rho^l_{i\sim j} u_j^{(l,k+1)} + \sum\limits_{j \in \mathcal{N}(i), j>i} \rho^l_{i\sim j} u_j^{(l,k)} - d_i}{\sum\limits_{j \in \mathcal{N}(i)} \rho^l_{i\sim j}}, \tag{6.20}$$

where $\omega \in (0,2)$ is a fixed relaxation parameter. Thereby, the index l counts the updates in the diffusivity terms, $\mathcal{N}(i)$ denotes the 6-neighborhood of i and $d_i = \mu_1 \left(\log P_{obj,i} - \log P_{bck,i}\right) + \mu_2 \Delta I_i$ summarizes the constant part of (6.19) that does not depend on u, i. e. the righthand side of the linear system. Finally, $\rho^l_{i\sim j}$ denotes the diffusivity between voxel i and its neighbor j while distinguishing between different spatial directions according to the anisotropic tensor D. It is defined as

$$\rho^l_{i\sim j} = \begin{cases} \gamma\,\dfrac{\rho^l_i + \rho^l_j}{2}, & \text{if voxels } i \text{ and } j \text{ are neighbors in the } z\text{-direction} \\ \dfrac{\rho^l_i + \rho^l_j}{2}, & \text{otherwise.} \end{cases} \tag{6.21}$$

Thus, diffusivities between neighboring voxels are estimated by averaging the corresponding values which are, in turn, computed as

$$\rho_i^l = \frac{g(I)_i}{\sqrt{\nabla u_i^{l\,T} D \nabla u_i^l + \epsilon^2}}, \tag{6.22}$$

where $\epsilon = 0.001$ is a small constant that prevents the diffusivity to become infinite when $\nabla u_i^{l\,T} D \nabla u_i^l = 0$ and ∇u_i^l is approximated by standard central differences. The overrelaxation parameter ω has to be chosen in the interval $(0, 2)$ for the method to converge. In our experiments, we set $\omega = 1.85$. After the linear solver yields a sufficiently good approximation (we iterated for $k = 1, ..., 10$), one can update the diffusivities ρ_i^l and solve the next linear system. Iterations could be stopped as soon as the energy decay becomes negligible.

6.6 Experiments

This section presents an experimental validation of the proposed approach. Since the segmentation method was developed with focus on a particular application, we give some details on the data acquisition process before demonstrating its performance on real-world data sets.

Experimental Setup

Materials

To mimic mechanical properties of cell membranes, 1,2-dioleoyl-sn-glycero-3- phosphocholine (DOPC) was chosen as main lipid component. The addition of 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-(cap Biotinyl) (capBioDPPE) gave rise to a drastically increased rigidity of the coupling of streptavidin to the lipid membrane. Lipid components were dissolved in chloroform and mixed to 1 mg/ml total lipid concentration in a molar ratio of DOPC/capBioDPPE = 10/1. All lipids were purchased from Avanti Polar Lipids Inc. (AL, USA) in a purity of at least 99.9%.

Vesicle preparation

Vesicles were prepared by the electroswelling technique [4]. 10 μl of the lipid mixture were carefully deposited on indium tin oxide (ITO) coated glass slides (Praezisionsglas & Optik GmbH, Iserlohn, Germany). The lipid films were then dried under vacuum for at least 1 h. For the electroswelling step, the plates were placed in a chamber containing 2 ml of 130 mM sucrose solution (130 mosm/L) in pure water and separated by a 1 mm teflon spacer. Vesicles were swollen by applying an altering electric field of 1.5 V and 10 Hz for 2 h. After electroswelling, vesicles were incubated in 70 μg/ml protein solution for at least 2 h at pH 5.5 under isoosmotic conditions.

Experimental environment

Increasing osmotic pressure was generated in a special microscopic chamber designed for experiments using gradual osmotic pressure change (see Fig. 6.5). It consisted of two parts. The lower chamber part was filled with 2 ml of 130 mM glucose solution (130 mosm/L) and 50 μl of isoosmotic vesicles suspension. To visualize the model membranes in fluorescence microscopy, fluorescently labeled polymer (dextran-TRITC from Sigma Aldrich Inc., St. Louis,

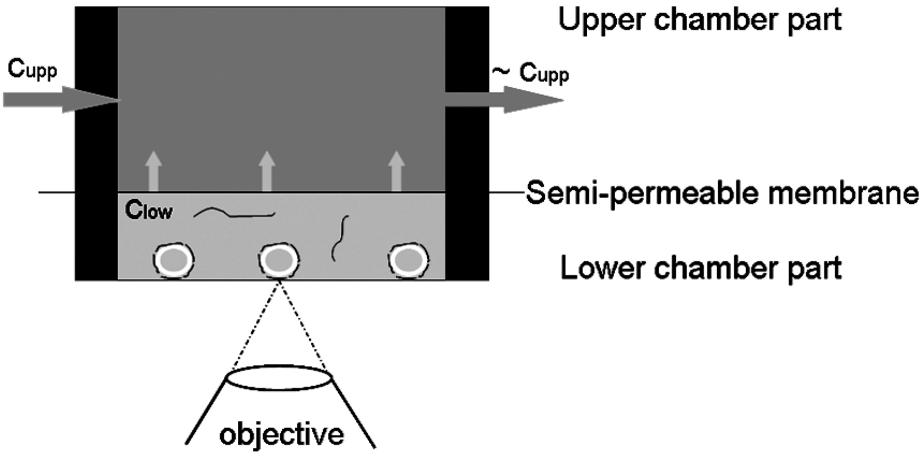
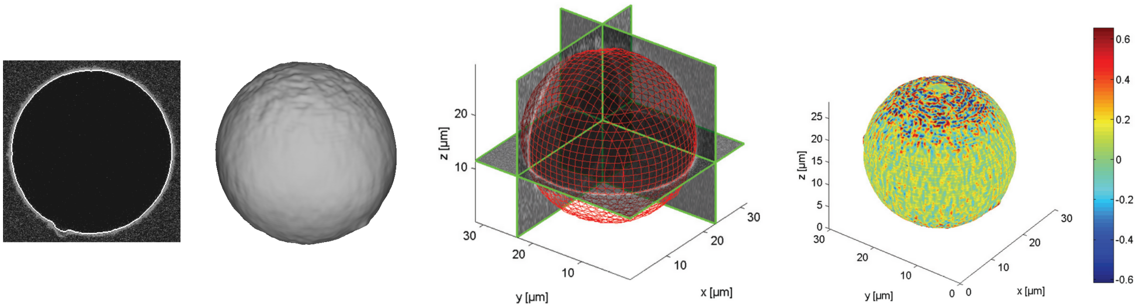


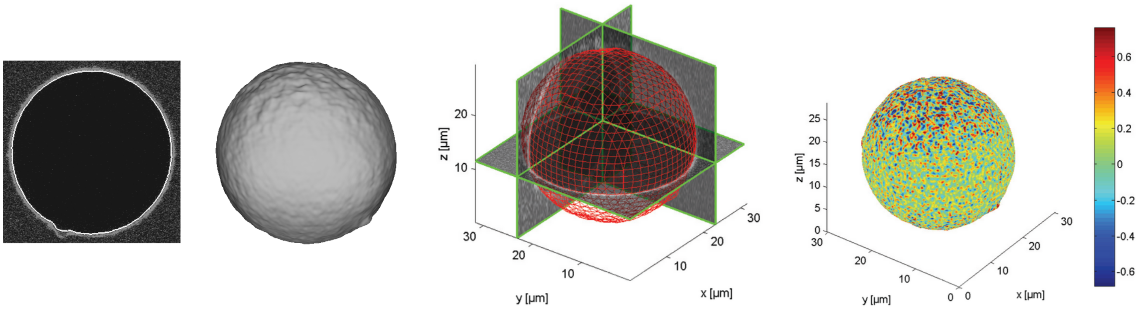
Fig. 6.5: Experimental setup. Dialysis through flow chamber for fluorescence microscopy observation of vesicle folding under continuously increasing hyperosmotic pressures was used. Vesicles were deposited in the lower chamber in glucose solution with a concentration of c_{low}. This chamber part was separated from the upper one by a semi-permeable membrane, which allowed a water flow and simultaneously excluded any other transport between the two chamber parts. The upper chamber had a constantly higher glucose concentration (c_{upp}) than the lower chamber. The concentration difference was reduced by water flow from the lower chamber (see small arrows), including vesicle volumes, into the upper chamber. The changed osmolarity induced vesicle volume loss at constant surface area thereby giving rise to vesicle folding. The process was observed by a laser scanning microscope.

MO, USA) was added to the glucose solution in a final concentration of 1 μg/ml while the vesicle lumen remained unlabeled. The fluorescently labeled polymer was uniformly dispersed in the glucose solution emitting low background intensity, which was bleached during the experiments. Streptavidin coated vesicle membranes are slightly negatively charged at pH 5.5. The electrostatical interaction between protein coating and fluorescently labeled polymer resulted in enrichment of polymer molecules on the protein surface increasing the fluorescence intensity on the vesicle membranes. This process occurred until the accumulation of polymers on the vesicle surfaces was saturated. As lipid bilayers were impermeable to polymers, vesicle lumens remained completely without fluorescent and appeared dark in the fluorescent channel, correspondingly.

The upper chamber part was a flow chamber. Isoosmolar glucose solution streamed through it with a flow rate of approximately 20 mL/min. The two parts were separated and simultaneously contacted with each other by a semi-permeable membrane (Reichelt Chemietechnik GmbH Co, Heidelberg, Germany). This membrane enabled the diffusion of water molecules between the two chamber parts simultaneously but retained the other dissolved components. Under isoosmotic conditions ($c_{upp} = c_{low}$), the two sides were equilibrated. Subsequently, high concentrated glucose solution (400 mosm/L) flowed through the upper chamber part ($c_{upp} \gg c_{low}$) disturbing the equilibrium state. To compensate the new concentration difference between the two sides, water flowed from the lower chamber into the upper flow-chamber until the equilibrium state was reached again. Consequently, the osmolarity in the lower chamber was continuously changed by 0 to 140 mosm/L in 1.5 h. The osmolarities of glucose and vesicle solutions were determined using a freezing point depression osmometer (Osmomat 030, Gonotec, Berlin, Germany).



segmentation of original data



segmentation of interpolated data

Fig. 6.6: Data set #1, vesicle under isoosmotic conditions. *First row:* Segmentation of raw microscopic data. *Second row:* Segmentation of interpolated data. *From left to right:* Superimposed segmentation for one of the volume slices, rendered view of the extracted shape, cross-sections through the volume with superimposed wireframe segmentation result and color-coded mean curvature of the surface $[\mu m^{-1}]$. This experiment demonstrates that upsampling the resolution in the z-direction does not improve the segmentation result, but rather increases the computation time by about a factor of 4.

Microscopic Imaging

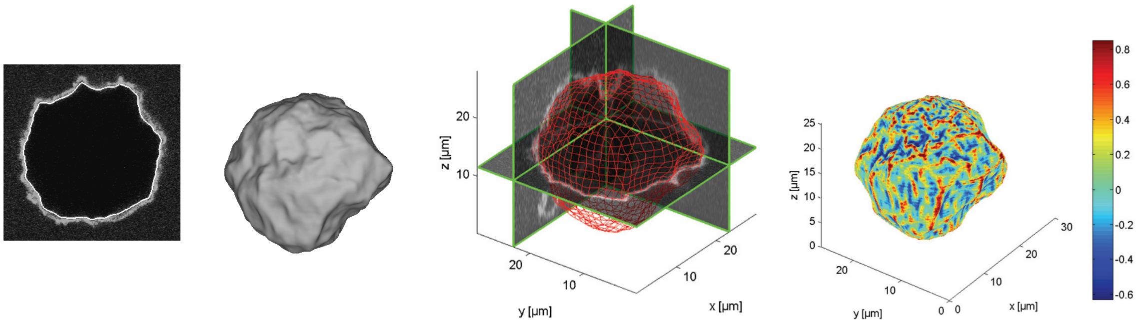
Three dimensional membrane curvatures of streptavidin coated giant vesicles were imaged with a laser scanning microscope (LSM 510, Carl Zeiss MicroImaging GmbH, Jena, Germany) equipped with a helium-neon laser (543 nm). To detect the fluorescent signal of dextran-TRITC, a long pass filter LP600 nm was used. The microscope was focused on the vesicles using a C-Apochromate 40 × /1.20 water immersion objective (Carl Zeiss). Images were recorded with a typical resolution of 512×512 pixels in x-y-plane (pixel size 154 nm) and 60 slices in the z-direction (pixel size 390 nm).

Evaluation on Real Data Sets

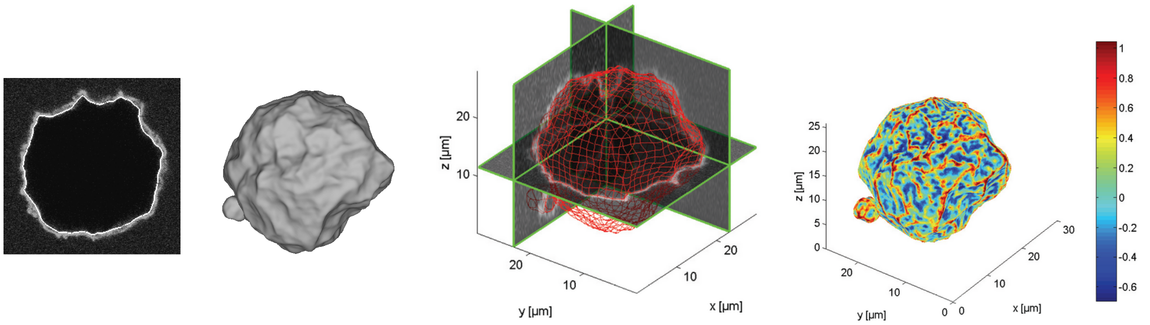
We demonstrate the performance of the proposed approach on four challenging real data sets generated with the described laser scanning methodology. Additionally, we present a comparison to alternative segmentation techniques based on level sets.

Performance

The data sets used in our experiments and the corresponding segmentation results are illustrated in Fig. 6.6, 6.7, 6.8 and 6.9. The data sets, acquired under different osmotic pressures, capture various deformations of the imaged vesicle. The figures depict a superimposed segmentation for one of the volume slices, rendered view of the extracted 3D shape, cross-sections through the volume with superimposed wireframe segmentation result and color-coded mean

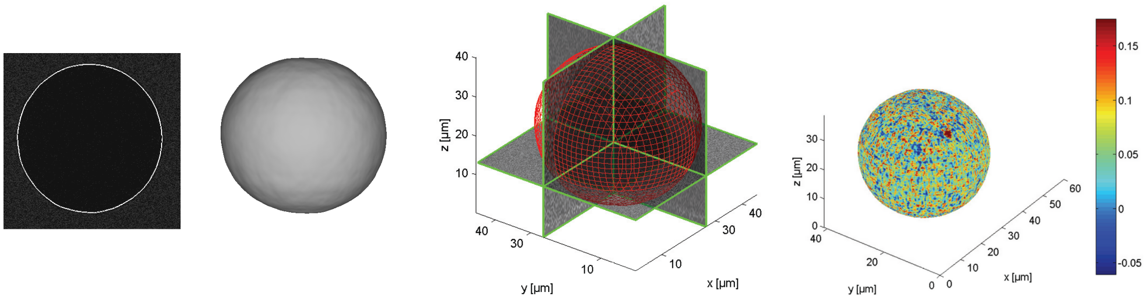


segmentation of original data

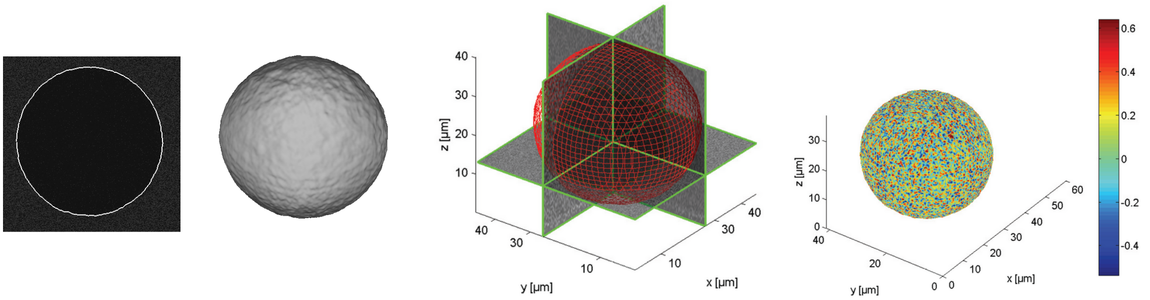


segmentation of interpolated data

Fig. 6.7: Data set #2. See text in Fig. 6.6 for reference. The slight differences in the reconstruction are most likely due to ghosting artifacts introduced in the interpolation process.

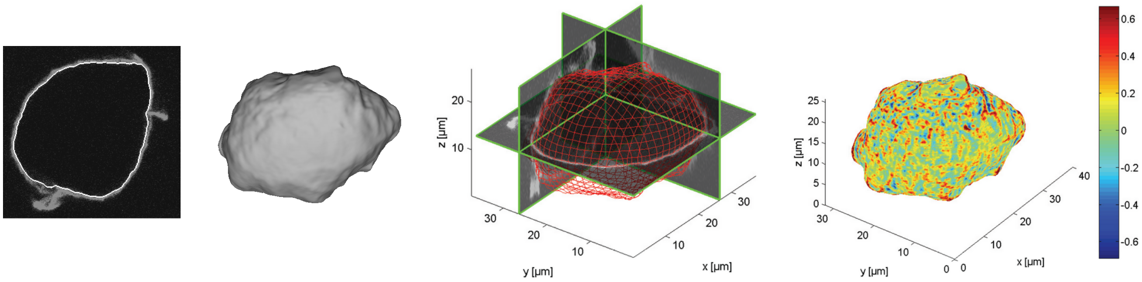


segmentation of original data

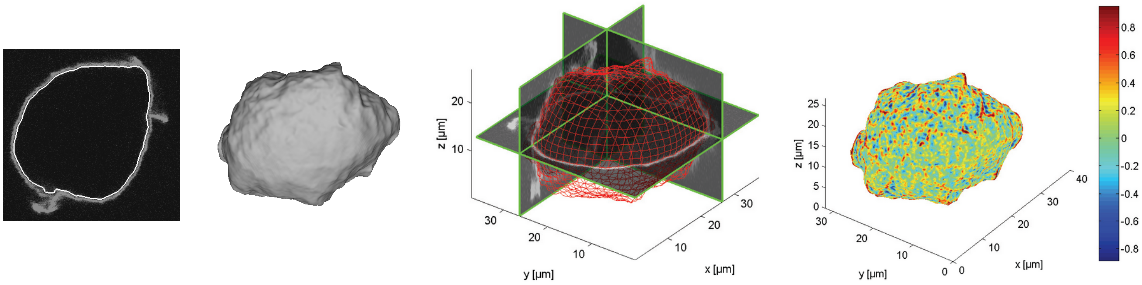


segmentation of interpolated data

Fig. 6.8: Data set #3. See text in Fig. 6.6 for reference. Again the interpolation gives rise to ghosting artifacts.

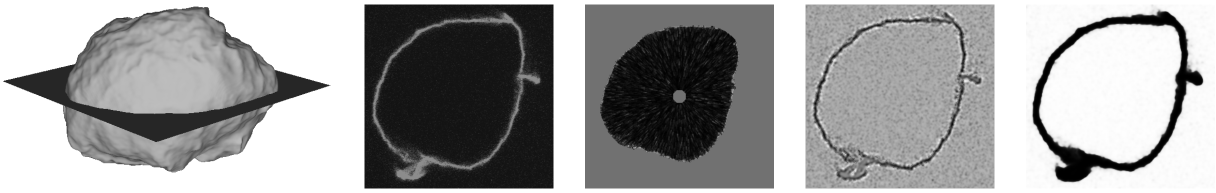


segmentation of original data



segmentation of interpolated data

Fig. 6.9: Data set #4. See text in Fig. 6.6 for reference.



cross-section

input slice

regional map model-based

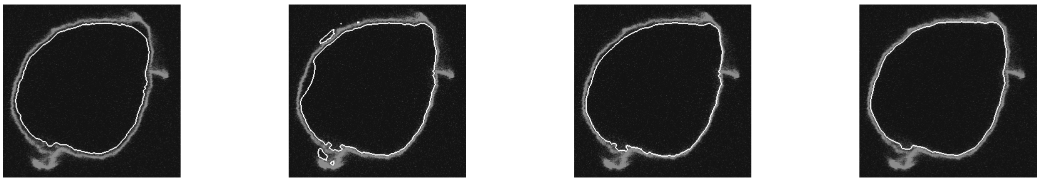
regional map proposed

edge map proposed

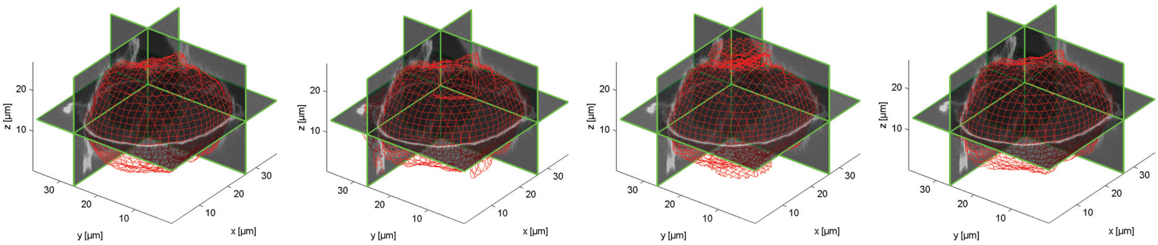
Fig. 6.10: Comparison of the data terms utilized by the model-based approach in [149] and the presented method for slice 104 of the interpolated version of data set #4 (see Fig. 6.9). Although the proposed regional term exhibits lower separability of object and background than the model-based one, it offers higher precision due to the underlying global scheme and the integration of shape orientation. Note that the model-based approach doesn't use explicit edge information in contrast to the proposed one. See Fig. 6.11 for comparison of the segmentation results.

data set	resolution	μ_1	μ_2	β_1	β_2	γ
#1 orig.	$267 \times 274 \times 59$	10^{-4}	10^{-4}	0.2	0.025	0.05
#1 interp.	$267 \times 274 \times 256$	10^{-4}	10^{-4}	0.2	0.025	1.0
#2 orig.	$258 \times 257 \times 56$	10^{-4}	10^{-4}	0.2	0.025	0.05
#2 interp.	$258 \times 257 \times 243$	10^{-4}	10^{-4}	0.2	0.025	1.0
#3 orig.	$345 \times 342 \times 80$	10^{-4}	10^{-4}	1.0	10^{-6}	0.05
#3 interp.	$345 \times 342 \times 305$	10^{-4}	10^{-4}	1.0	10^{-6}	1.0
#4 orig.	$283 \times 275 \times 60$	10^{-4}	$3 \cdot 10^{-4}$	0.2	0.025	0.05
#4 interp.	$283 \times 275 \times 205$	10^{-4}	$3 \cdot 10^{-4}$	0.2	0.025	1.0

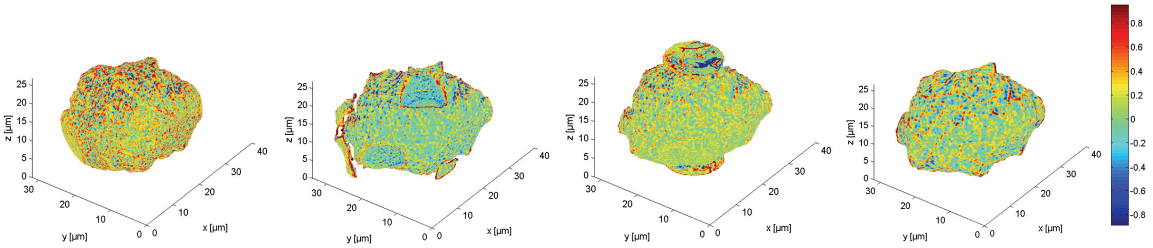
Tab. 6.1: Parameter setting of the proposed approach for all demonstrated experiments.



superimposed segmentation for one representative slice



volume cross-sections and superimposed segmentation



color-coded curvature

Fig. 6.11: Comparison between the proposed approach and alternative segmentation techniques based on level sets. *First row:* Segmentation for one of the volume slices. *Second row:* Cross-sections through the volume with superimposed wireframe segmentation result. *Third row:* Color-coded curvature of the surface. *From left to right:* A variant of the model-based approach in [149], a level set implementation of the proposed energy model, as described in Section 6.5, by using a box-shaped and spherical initialization, respectively, and convex relaxation, as described in Section 6.5. Note that while the formulation in [149] involves only local data term estimation and requires a specific initialization, the presented method is general and relies on global computations.

curvature of the surface. In some cases, intensity variations within the membrane regions caused high gradients which lead to undesired structures within the membrane. We could overcome this problem by limiting the gray values to an adequate range. For all test cases we show segmentation results for both the original raw volumetric data, as obtained with the described laser scanning microscopy technique, and an interpolated version of the data on an approximate cubic grid. The parameter setting used in our experiments is specified in Table 6.1. As can be expected, the anisotropy parameter γ has been adapted appropriately so as to compensate for the sparse volume slicing in the z-direction of the original raw data. Yet, none of the other parameters has been changed. This allows to assess the effect of the interpolation process on the quality of the segmentation results. As can be observed, in both cases the segmentations are quite accurate (see the selected slices). However, they possess different degrees of smoothness, which is confirmed by the curvature measurements. This phenomenon can be explained with the appearance of small-scale ghost structures in the interpolation process. As for the runtimes, on a PC with 2.83 GHz and 8 GB of main memory we measured computational times in the range $10-15$ minutes for the original raw data sets and in the range $40-60$ minutes for the interpolated data sets. The bottleneck of the proposed approach poses the numerical optimization. Yet, note that the applied linearized fixed-point iteration scheme possesses excellent parallelization potential. With a GPU implementation

the computational times shrink to a couple of seconds.[1]
The demonstrate experiments cover a wide range of test scenarios. While some of the data sets (#1 and #3) exhibit little deformation of the imaged vesicle membrane, others (#2 and #4) capture substantial bending and stretching. Furthermore, data set #3 is characterized by completely missing edge information, whereas data set #4 features lacking foreground/background intensity discrimination. The remarkable accuracy of all segmentation results emphasizes the high practical value of the proposed approach.

Comparison to Alternative Approaches

In order to demonstrate the accuracy and robustness of the proposed approach, we present a twofold comparative evaluation (see Fig. 6.11). Firstly, we explore the viability of the energy model by comparing it to a state-of-the-art method. In particular, we implemented a variant of [149] tailored to the application at hand. Secondly, we compare an implementation of the proposed energy model with level sets, as described in Section 6.5, and convex relaxation, as described in Section 6.5.
For the sake of completeness, we briefly overview the approach in [149]. Similar to the proposed formulation, it is based on a variational framework. Concretely, the following energy functional is being minimized

$$\begin{gathered} E(u) = \int_V S(u)\, u(x)\, dx + \int_V |\nabla u(x)|\, dx \\ \textbf{s. t.} \quad u : V \to \{0,1\}\,, \end{gathered} \tag{6.23}$$

where the speed function $S : V \to \mathbb{R}$ is defined as follows

$$S(x) = \frac{I_{max}}{I_{min}} - T. \tag{6.24}$$

Thereby, the function is evaluated only in a tight vicinity around the evolving interface by considering lines of certain length along its local normals. In practice, S is updated only on the surface and at voxels in direct contact with it. I_{max} and I_{min} denote the maximum and minimum intensity along the current line. $T \in \mathbb{R}_{\geq 0}$ is a parameter, specified by the user, that controls the desired intensity contrast and is used to stop the evolution process. Note that due to its greedy nature the described procedure requires a specific initialization. It relies on the assumption that the surface can be identified by locations of high edge contrast while the propagation scheme is initialized within a homogeneous region. Usually, the initial estimate is placed within the object interior and evolves outwards until the speed function reaches values close to zero. In our implementation, the initialization was fixed to a sphere centered in the middle of the volume with radius 10 voxels. Although the above method is able to handle data of weak foreground/background discrimination effortlessly, its precision may suffer. The utilized local update procedure could lead to premature termination and oversmoothing effects (see Fig. 6.11).
In the following, we summarize the main conceptual differences between the approach in [149] and the proposed one. They are revealed by a direct comparison between both variational models, given in (6.17) and (6.23).

- While the method in [149] uses a regional term based on the local intensity contrast, the proposed formulation relies on a more sophisticated computation involving foreground/background distribution modeling and surface orientation alignment. As a result, the estimated regional map is imprecise, susceptible to noise and only locally

---

[1] See Chapter 3 for a more detailed evaluation.

defined in the first case and offers a global range and a high degree of accuracy in the latter case, see Fig. 6.10. It is important to note that while the model-based approach in [149] requires a specific initialization due to the local nature of its evolution scheme, the proposed model is global and does not depend on initialization.

- The method in [149] relies on a classical Euclidean regularizer to impose smoothness of the surface. In contrast, the proposed model uses explicit edge information to avoid oversmoothing effects and increases the accuracy of the segmentation results. See Fig. 6.10 for an example edge map.
- We even take a further step and develop an anisotropic generalization which allows to adequately guide the smoothing along particular directions taking into account the density of the slicing produced by the utilized imaging technology. As a consequence, the proposed approach can be applied to the measured raw volume data without any modifications while the model in [149] is restricted to cubic voxels and requires interpolation as a preprocessing step.
- While [149] is based on the level set framework to solve the underlying minimization problem, we propose the use of convex relaxation which allows for global optimization. However, this is not a crucial issue for the method in [149] since only local data term updates are performed.

Fig. 6.11 compares the segmentations obtained with both methods on data set #4. We used this data since it seems to exhibit the most challenging scenario due to the lack of foreground/background intensity discrimination. It can be observed that the method in [149] gives a relatively good segmentation result for this challenging data set although it is a portent of a clear shrinking bias limiting the precise localization of the vesicle membrane. In contrast, the proposed approach demonstrates clear superiority and provides a quite accurate segmentation.

In order to emphasize the importance of a robust optimization scheme, we compare an implementation of the proposed energy model with level sets, as described in Section 6.5, and convex relaxation, as described in Section 6.5. Note that the level set method leads to a local minimum, which makes it highly susceptible to initialization. To this end, we provide segmentation results with two different initializations - a box centered in the middle of the volume with side length equal to half of the resolution along the corresponding dimension and a sphere centered in the middle of the volume with radius 100 voxels. In contrast, the convex relaxation technique always converges to a global minimum of the underlying energy functional and is completely independent of initialization. Expectedly, both initialization procedures for the level set implementation lead to fairly different segmentations. Note that the optimization scheme is coupled with the process of updating the foreground/background histograms. The box-shaped initialization produces a quite poor segmentation result. In particular, due to the weak foreground/background intensity discrimination and the dominance of the flux term, the generated surface consists of multiple components situated around the vesicle boundary. The spherical initialization provides a closer capturing of the imaged shape and leads to a more accurate segmentation result. The elongated structures at the top and the bottom of the surface may seem surprising at first but not after a more extensive inspection of the input data set (see the visualization of the cross-sections). And still, some oversmoothing effects of the resulting boundary can be observed. Hence, the convex relaxation technique proves essential for the precision of the final segmentation result. Recall that the proposed approach relies on a global data term computation.

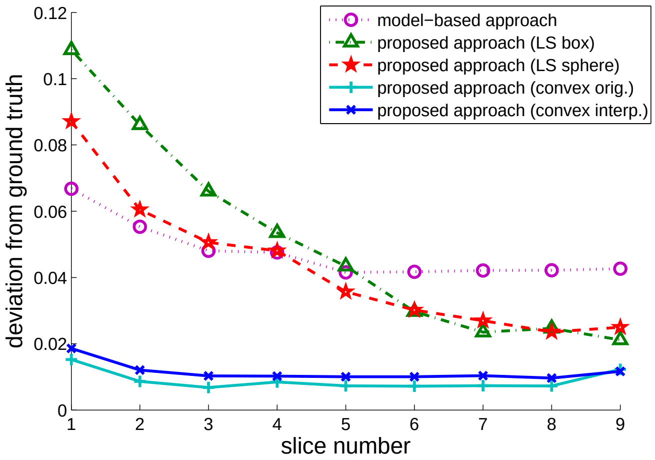


Fig. 6.12: Quantitative evaluation of the segmentation results in Fig. 6.11. The accuracy of the computed segmentations is measured in terms of their deviation from a manually obtained ground truth for a couple of representative slices. Note that the proposed model optimized with level sets exhibits a quite jumpy behavior due to the local minimization scheme in combination with the globally estimated data terms. The proposed approach based on convex optimization proves superior to all other methods.

The discussed conclusions are additionally confirmed by a quantitative evaluation of the above segmentation results, shown in Fig. 6.12. To this end, a few equidistantly spaced slices out of the volume interior have been manually segmented by an expert and used as a ground truth. Note that the slices close to the boundary are quite unclear and a ground truth segmentation is hard to obtain. If $u_{gt} : V \to \{0,1\}$ denotes an implicit labeling representing this ground truth surface (being 0 within the interior region and 1 within the exterior) and $u : V \to \{0,1\}$ the obtained 3D segmentation, we measure the misalignment between them as

$$\epsilon = \frac{\displaystyle\int_V |u_{gt}(x) - u(x)|\, dx}{\displaystyle\int_V u_{gt}(x)\, dx + \int_V u(x)\, dx}. \tag{6.25}$$

In particular, we have $\epsilon \in [0,1]$ with $\epsilon = 0$ if and only if both segmentations are identical and $\epsilon = 1$ if u is the empty set or if both labelings are complementary to each other. As evident from the plot, the proposed model optimized with level sets exhibits a jumpy behavior for both a box-shaped and a spherical initialization due to the local minimization scheme in combination with the globally estimated data terms. In contrast, the model-based method in [149] as well as the proposed approach based on convex relaxation demonstrate a stable behavior, whereas the latter offers substantially increased accuracy. This clear superiority over all other methods validates both the derived energy model and the utilized optimization technique. Note that applying the method to the original volume data without interpolation, which is not possible with the formulation in [149], entails additional accuracy gains.

In addition to the improved accuracy, the proposed approach is substantially faster than the method in [149]. For the interpolated data set, the measured computational time was about an hour for the proposed algorithm as opposed to more than 6 hours for the model-based one. The relatively high computational costs of [149] are mainly due to the time-consuming data term updates within an entire vicinity around the current surface estimate as well as the process of reinitialization of the underlying implicit function, inherent to the level set framework.

6.7 Discussion

In this chapter, we extended the range of applicability of convex optimization to the domain of 3D segmentation. In particular, we proposed a robust and accurate approach with applications to vesicle membrane analysis. The method is designed to operate adequately with noisy data sets with a specific focus on biological applications like vesicle and cell segmentation. Despite this emphasis on the particular application area the approach is general and can be easily adapted to other segmentation tasks. The key idea is to tackle the problem as the minimization of an appropriate energy functional integrating various features that can be extracted from the input volume data like regional statistics, edge identification and surface orientation. Moreover, the proposed model replaces the traditionally used isotropic regularizer by an anisotropic one in order to address the non-uniform resolution and therefore also sampling of the measurements along different spatial directions. We showed how the resulting energy functional can be optimized globally in a continuous setting by means of convex relaxation. Experiments on challenging fluorescence imaging data demonstrated that the proposed approach allows to accurately reconstruct the vesicle membranes despite the limited discriminative characteristics of intensity distributions and despite blurred and noisy surface boundaries. Furthermore, the reconstruction results were systematically improved by incorporating the different energy terms. Experimental results also confirmed the superiority of the proposed convex relaxation approach over state-of-the-art level set methods as it gives rise to robust and highly accurate membrane reconstructions, independent of initialization.

7

Conclusion

The scientist, by the very nature of his commitment,
creates more and more questions, never fewer.

Gordon Willard Allport (1897-1967)

This thesis explored the potential of convex optimization for solving surface reconstruction problems in computer vision like multiview 3D reconstruction and 3D segmentation. Throughout the work, we became acquainted with different energy models, amenable to convex optimization, fusing various cues and thus addressing the tackled problems in manifold ways.

In Chapter 2, an interactive approach for 3D modeling from object outlines was proposed. The underlying convex formulation allowed to find an optimal 3D shape with respect to the specified user input in form of a few scribbles marking object and background in only one of the input images. We built a probabilistic framework taking the entire amount of available color information to deliver the most probable reconstruction. We saw that the proposed probabilistic formulation entails considerable robustness to image noise, camera sensor perturbations and background clutter.

Chapter 3 focused on the multiview stereo problem. We considered three energy models sharing the same convex variational structure. We investigated different strategies for computing respective data terms and compared them qualitatively and quantitatively. Moreover, two numerical schemes for solving the arising minimization problem were examined with respect to runtime, memory requirements and potential for parallel computing.

Chapter 4 tackled the problem of fusing complementary information sources like silhouettes and multiview stereo. To this end, a transparent formulation was proposed, where stereo consistency was imposed by constructing a suitable energy model and silhouette alignment – by means of constraints restricting the domain of feasible shapes. After relaxation, we ended up with a constrained convex optimization problem. A solution of the original non-convex problem was derived by appropriate thresholding, which was shown to lie within an energy gap from the globally optimal one.

Chapter 5 extended the class of functionals amenable to convex optimization by including anisotropic minimal surfaces and minimal ratios. We demonstrated the capability of the anisotropic minimal surface model to incorporate given surface normal information while retaining all globality guarantees of isotropic formulations. Further on, we focused on two particularly interesting properties of the minimal ratio model – absence of a shrinking bias and scale invariance. We saw that both models are amenable to convex optimization. While this is straightforward for the anisotropic minimal surface model as it generalizes previously proposed formulations, the minimal ratio model requires more care. Yet, we showed that its minimization boils down to solving a sequence of convex optimization problems.

Chapter 6 was devoted to the 3D segmentation problem. Once again, convex optimization served as a powerful tool in the design of an appropriate energy model meeting all requirements. The proposed formulation integrates a variety of image cues like regional intensity statistics, edge attraction and orientation information. Although the method was developed with focus on a particular biological application at hand – vesicle membrane reconstruction from fluorescence imaging – it is general enough to be applied to a multitude of different segmentation tasks.

7.1 Main Contributions

In the following, we summarize the main contributions of the thesis.

- The first convex formulation in the field of multiview 3D reconstruction was proposed [73]. Thereby, along with the potential of convex optimization for image-based modeling, the advantages of the technique over discrete counterparts like graph cuts were explored. This opened up new ranges for further investigations in the domain [144, 68, 72, 143] and beyond [103, 120, 102, 97, 99, 25].
- A probabilistic framework for fusing different observations within a multiview setting was developed [66]. While similar formulations have been investigated in the context of background subtraction [38], the proposed approach stands out by modeling intensity/color statistics. It served as a basis for an interactive method which allows to obtain an optimal 3D shape for the specified user input [24].
- It was shown that multiview stereo can be posed as a convex problem [73, 71, 72]. This observation resulted in the design of robust and accurate reconstruction approaches which do not require any initialization and do not depend on the geometrical structure or topology of the imaged object.
- A mathematically elegant method for fusing silhouette and stereo information was proposed [68]. In contrast to alternative formulations (e. g. [116]), the presented scheme avoids premature hard decisions about voxel occupancy, which leads to increased robustness and accuracy. The uniqueness of the proposed model consists also in the guarantee that the computed solution lies within an energetic gap from the globally optimal one. The developed framework inspired other researchers to generalize it to further application domains [82, 80].
- The class of cost functionals amenable to convex optimization was extended to include minimal ratio models [69]. Along with the global optimizability in a continuous setting of the proposed minimal ratio model, some useful properties like absence of a shrinking bias and scale invariance were explored.
- Convex relaxation techniques were generalized to anisotropic metrics [74]. While similar formulations were investigated for other related applications [143, 97], the proposed anisotropic minimal surface model was developed independently.

7.2 List of Publications

- K. Kolev, T. Brox and D. Cremers, “Robust variational segmentation of 3D objects from multiple views”, Pattern Recognition (Proc. DAGM), 2006.
- K. Kolev, M. Klodt, T. Brox, S. Esedoglu and D. Cremers, “Continuous Global Optimization in Multiview 3D Reconstruction”, Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2007.

- K. Kolev, M. Klodt, T. Brox and D. Cremers, “Propagated Photoconsistency and Convexity in Variational Multiview 3D Reconstruction”, ICCV Workshop on Photometric Analysis for Computer Vision (PACV), 2007.
- M. Klodt, T. Schoenemann, K. Kolev, M. Schikora and D. Cremers, “An Experimental Comparison of Discrete and Continuous Shape Optimization Methods”, European Conference on Computer Vision (ECCV), 2008.
- K. Kolev and D. Cremers, “Integration of Multiview Stereo and Silhouettes via Convex Functionals on Convex Domains”, European Conference on Computer Vision (ECCV), 2008.
- K. Kolev and D. Cremers, “Continuous Ratio Optimization via Convex Relaxation with Applications to Multiview 3D Reconstruction”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
- M. R. Oswald, E. Toeppe, K. Kolev and D. Cremers, “Non-Parametric Single View Reconstruction of Curved Objects using Convex Optimization”, Pattern Recognition (Proc. DAGM), 2009.
- K. Kolev, M. Klodt, T. Brox and D. Cremers, “Continuous Global Optimization in Multiview 3D Reconstruction”, International Journal of Computer Vision (IJCV), August 2009.
- K. Kolev, T. Pock and D. Cremers, “Anisotropic Minimal Surfaces Integrating Photoconsistency and Normal Information for Multiview Stereo”, European Conference on Computer Vision (ECCV), 2010.
- D. Cremers and K. Kolev, “Multiview Stereo and Silhouette Consistency via Convex Functionals over Convex Domains”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2011.
- D. Cremers, T. Pock, K. Kolev and A.Chambolle, “Convex Relaxation Techniques for Segmentation, Stereo and Multiview Reconstruction”, Advances in Markov Random Fields for Vision and Image Processing MIT Press, 2011.
- K. Kolev, N. Kirchgeßner, S. Houben, A. Csiszar, W. Rubner, C. Palm, B. Eiben, R. Merkel and D. Cremers, “A Variational Approach to Vesicle Membrane Reconstruction from Fluorescence Imaging”, Pattern Recognition, 2011.
- M. Aubry, K. Kolev, B. Goldluecke and D. Cremers, “Decoupling Photometry and Geometry in Dense Variational Camera Calibration”, IEEE International Conference on Computer Vision (ICCV), 2011.
- K. Kolev, T. Brox and D. Cremers, “Fast Joint Estimation of Silhouettes and Dense 3D Geometry from Multiple Images”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2012.

7.3 Future Work

As convex optimization proved itself as a useful and powerful tool for solving a variety of surface reconstruction problems, this raises the question about further potential application areas. Three research directions are particularly appealing – shape from shading/photometric stereo, large-scale multiview reconstruction and dense camera calibration.

In Chapter 5, we saw that given surface normal information can easily be integrated in a convex optimization process. This observation opens up new ranges for applications like shape from shading [53] and photometric stereo [139]. As the photometric stereo problem is generally well-posed and directly incorporates an estimated normal field, the application of the proposed anisotropic minimal surface model is straightforward. Yet, the shape-from-shading problem is mathematically ill-posed, since the normal at each surface point is not uniquely determined, which entails more difficulties.

Recently, the problem of reconstructing large-scale urban scenes from a collection of images or video data has attracted considerable attention in the community [104, 42, 39]. Severe illumination variations, occlusions and imprecise camera calibration present great challenges and require a robust 3D modeling procedure. As convex optimization manifested itself as a valuable and useful tool to address difficulties of this type, its generalization to large-scale image-based modeling seems quite promising.

This work focused on image-based dense surface reconstruction. In the context of 3D modeling from multiple views, this presumes given camera calibration. Thus, building a unified framework for reconstruction and calibration is of contemporary interest. While such formulations already exist [127, 6], their convexification still remains an open challenge.

A

Visibility Estimation via Implicit Ray Tracing

In Section 1 and 3, we saw that visibility estimation is essential in multiview stereo. Here, we give more details on the utilized technique for visibility reasoning. It is a variant of a method proposed in [126].

Given a set of n fully calibrated cameras and a current surface estimate in an implicit form $\tilde{u} : \tilde{V} \to \{0, 1\}$ within a discrete volume $\tilde{V}$, where 1 denotes interior and 0 exterior, the goal is to compute a global visibility map $Vis : \tilde{V} \to \mathcal{P}(\{1, \dots, n\})$. Thereby, for a grid point $\tilde{x} \in \tilde{V}$, $Vis(\tilde{x})$ should give the indices of all cameras that have a direct visual contact to $\tilde{x}$.[1]

For simplicity, we start with the visibility estimation for one camera, say i, before generalizing the computations to the multiview setting. To this end, we define a visibility function $\psi_i : \tilde{V} \to \{0, 1\}$ encoding this information. Let $O_i \in \mathbb{R}^3$ signify the known camera position. A basic observation is that the visibility along each ray emanating from the vantage point satisfies a causality condition: if a point is occluded, then all other points farther away from the vantage point on the same ray are also occluded. Thus, we can set

$$\psi_i(\tilde{x}) = \max_{\xi \in \mathcal{L}(O_i, \tilde{x})} \tilde{u}(\xi), \tag{A.1}$$

where $\mathcal{L}(O_i, \tilde{x})$ is the line segment connecting O_i and $\tilde{x}$. According to the above definition, if $\psi_i(\tilde{x}) = 1$, then $\tilde{x}$ is occluded. The key idea is to approximate ψ_i by a function $\psi_i^h : V \to [0, 1]$, where V denotes the continuous volume encompassing the scene and thus $\tilde{V} \subset V$, defined as follows

$$\psi_i^h(\tilde{x}) = \max\left(\psi_i^h(\tilde{x}'), \tilde{u}(\tilde{x})\right). \tag{A.2}$$

Thereby, $\tilde{x}' \in V$ is some point immediately before $\tilde{x}$ in the ray direction. As long as the values of $\psi_i^h(\tilde{x}')$ are computed ahead of the estimation of $\psi_i^h(\tilde{x})$, the above procedure will be valid. In particular, for a grid point $\tilde{x}$ an upwind neighbor $\tilde{x}'$ is obtained by considering the cube for which $\tilde{x}$ poses one of the corners and the other corners are neighboring grid points of $\tilde{x}$ on $\tilde{V}$ and which is crossed by the viewing ray passing through $\tilde{x}$. The ray determines a unique intersection point with one of the sides of the cube. Now, this point is taken as $\tilde{x}'$. Note that usually $\tilde{x}'$ does not lie on the grid $\tilde{V}$ and the value $\psi_i^h(\tilde{x}')$ is computed by trilinear interpolation. As a consequence, ψ_i^h takes on not only binary values, but values within the entire unit interval $[0, 1]$. A visibility decision could be taken by thresholding at 0.5.

One additional issue requires specification – determining a processing order of the volume grid points which allows for applying (A.2). A valid traversal scheme is described in the following. In [126], it is referred to as a star-shaped updating sequence. Starting from the vantage point O_i, the grid points close to it are processed first before moving to more distant ones. Note that the camera is usually placed outside the volume V as the volume is aligned with the imaged

---

[1] Note that the terms “voxel” and “grid point” can be used as synonyms. Yet, the interpretation of discrete volume units as grid points is here more convenient.

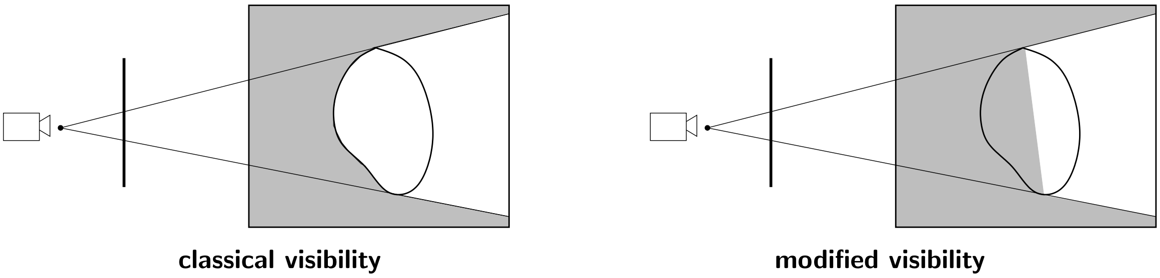


Fig. A.1: Classical vs. modified visibility. While classical visibility estimation procedures classify points within the interior of an object as occluded, the proposed modified scheme extends the visibility region (marked in gray) inside the surface and allows to apply appropriate regional terms.

scene to achieve optimal spatial resolution. In such cases, we take the projection of the vantage point onto the volume $O_i' = \Pi_V(O_i)$ as a starting point for the traversal algorithm. First, we compute the values of ψ_i^h at the eight neighboring grid points and subsequently within the strips along the three coordinate axes. In general, these strips separate the discretized volume $\tilde{V}$ into eight octants[2] which are processed in series slice-wise. See [126] for more details on this sweeping procedure.

So far, we have considered classical visibility estimation for one given camera i. There is one issue that requires more attention. Many of the approaches, proposed in this thesis, involve computing volumetric data terms specifying surface interior and exterior region. Yet, according to the traditional visibility reasoning, all points lying inside the current surface estimate are classified as occluded. Hence, we have to devise a strategy to propagate visibility information inside the surface in order to make the proposed methods applicable. To this end, we modify the definition of ψ_i^h to

$$\psi_i^h(\tilde{x}) = \begin{cases} \psi_i^h(\tilde{x}'), & \text{if } \langle \nabla \tilde{u}(\tilde{x}), \tilde{x} - O_i \rangle \geq 0 \\ \max\left(\psi_i^h(\tilde{x}'), \tilde{u}(\tilde{x})\right), & \text{otherwise.} \end{cases} \tag{A.3}$$

Note that $\nabla \tilde{u}(\tilde{x})$ reflects the local surface normal. Hence, the above formulation allows to propagate visibility information directly if the surface is oriented towards the vantage point. The intuition behind is the fact that accurate regional terms can be computed only for grid points with the given property. See Fig. A.1 for a comparison between the classical visibility estimation and the proposed modified scheme.

Finally, after obtaining the visibility functions for all cameras $\psi_1^h, \dots, \psi_n^h$, the global visibility map Vis is derived from the relation

$$i \in Vis(\tilde{x}) \quad \Longleftrightarrow \quad \psi_i^h(\tilde{x}) \leq 0.5. \tag{A.4}$$

In order to recognize the efficiency of the proposed scheme for visibility estimation, we observe that its computational time is in $\mathcal{O}(n \cdot N)$, where n denotes the number of cameras and N – the number of volume grid points. This is genuinely fast as opposed to a naive procedure which treats all viewing rays separately and has a runtime in $\mathcal{O}(n \cdot N^{4/3})$. Note that in typical application scenarios N is in the order of multiple million.

---

[2] Note that the number of octants could vary between one and eight. In particular, we have eight octants if the vantage point lies within the volume and one octant if the vantage point projects onto one of the corners of the volume.

B

Total Variation Norm

Throughout this thesis, we used the total variation norm as a regularization scheme in practically all derived energy models. Here, we study some of its most prominent properties. A more detailed description can be found in [101].

For a function $u : V \subset \mathbb{R}^3 \to [0,1]$, $u \in BV(V)$, the total variation norm (TV norm) is defined as

$$TV(u) = \int_V ||\nabla u||\, dx, \tag{B.1}$$

where $||.||$ denotes an arbitrary norm. In the sequel, we consider the Euclidean norm $|.|$ for simplicity.

While the definition in (B.1) assumes a certain degree of smoothness of u, i. e. functions with integrable first order derivatives, there exists a more general formulation which is defined for any real-valued function $u \in L^1(V)$, i. e. for any absolutely integrable function

$$TV(u) = \max_{|p|\leq 1} \int_V u \cdot \mathrm{div}(p)\, dx, \tag{B.2}$$

where $p \in \mathbb{R}^3$. Obviously, this definition of the TV norm is also valid for discontinuous functions as long as they are absolutely integrable.

The TV norm has an interesting geometrical property. It can be decomposed by means of the level sets of u. The relation is expressed by the coarea formula

$$\int_V |\nabla u|\, dx = \int_0^1 Per\left(\{x \mid u(x) > \gamma\}\right)\, d\gamma, \tag{B.3}$$

where one integrates the perimeter of the set $\{x \mid u(x) > \gamma\}$ for all values of γ. The coarea formula essentially states that the TV norm can be decomposed into a sum of the length of all level sets of u. This property was already used in the proof of Theorem 6.

Finally, we focus on maybe the most important characteristic of the TV norm – it is discontinuity preserving. To see this, consider an example in 1D depicted in Fig. B.1.[1] All three functions $f_1, f_2, f_3 : [0,1] \to [0,1]$ satisfy the boundary conditions $f(0) = 0$ and $f(1) = 1$. Moreover, all three functions are monotonically increasing in the given interval. Despite the different degree of smoothness, it can be verified that all functions exhibit the same value of the total variation norm which is 1. Note that for discontinuous functions an estimate can be calculated by subsampling the respective signal. We can conclude that the TV norm is not sensitive to discontinuities. This property is of paramount practical importance as it allows for highly detailed reconstructions while maintaining a high level of regularization.

---

[1] The example has been taken from [101].

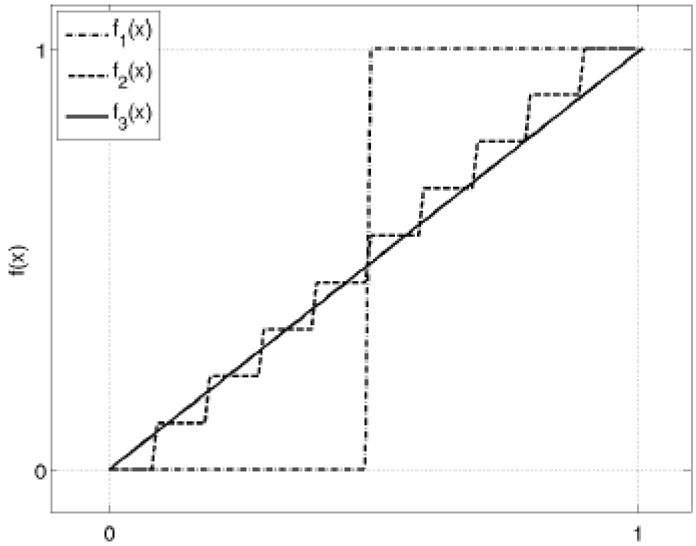


Fig. B.1: Preservation of discontinuities. Despite the different degree of smoothness, all three functions exhibit the same value of the total variation norm. The example has been taken from [101].

C

Notations

dx	volume element (3D)
$d\tilde{x}$	discretized volume element (voxel)
ds	surface element (2D)
dz	image element (2D)
dr	ray element (1D)
$V \subset \mathbb{R}^3$	volume
$\tilde{V} \subset V$	discretized volume
$\Omega_i \subset \mathbb{Z}^2$	image domain of camera i
$I_i : \Omega_i \to \mathbb{R}^3$	i-th color image
$\pi_i : V \to \Omega_i$	projection mapping of camera i
$Sil_i \subset \Omega_i$	object silhouette in image i
$n \in \mathbb{N}$	number of images
$N \in \mathbb{N}$	number of voxels
$Vis : V \to \mathcal{P}(\{1, \ldots, n\})$	visibility map
$S : \Theta \subset \mathbb{R}^2 \to V$	surface estimate
$N_S : \Theta \to S^2$	outward surface normal field of S
$int(S) \subset V$	interior of S
$ext(S) \subset V$	exterior of S
$\tilde{int}(S) \subset \tilde{V}$	discretized interior of S
$\tilde{ext}(S) \subset \tilde{V}$	discretized exterior of S
$u : V \to \{0, 1\}$	implicit labeling function
$u : V \to [0, 1]$	relaxed implicit labeling function
$I : V \to [0, 1]$	volumetric 3D data to be segmented
$\|\|.\|\|$	general metric
$\|.\|$	Euclidean metric
$\|.\|_D$	metric induced by D, i. e. $\|v\|_D = \sqrt{v^T D v}$

Bibliography

[1] `http://phototour.cs.washington.edu/bundler/`.

[2] `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[3] `http://ewokrampage.wordpress.com/`.

[4] M. I. Angelova, S. Soléau, P. Méléard, J. F. Faucon, and P. Bothorel. Preparation of giant vesicles by external AC electric field. kinetics and applications. In *Progress in Colloid & Polymer Science*, volume 89, pages 127–131, 1992.

[5] B. Appleton and H. Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):106–118, 2006.

[6] M. Aubry, K. Kolev, B. Goldluecke, and D. Cremers. Decoupling photometry and geometry in dense variational camera calibration. In *Proc. International Conference on Computer Vision*, Barcelona, Spain, November 2011.

[7] X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *Proc. International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[8] B. Baumgart. *Geometric modeling for computer vision.* PhD thesis, Department of Computer Science, Stanford University, USA, 1974.

[9] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *Proc. European Conference on Computer Vision*, pages 428–441, 2004.

[10] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, New York, NY, USA, 2004.

[11] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proc. International Conference on Computer Vision*, page 26, Washington, DC, USA, 2003. IEEE Computer Society.

[12] Y. Boykov and V. Lempitsky. From photohulls to photoflux optimization. In *Proc. British Machine Vision Conference*, volume 3, pages 1149–1158, 2006.

[13] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proc. International Conference on Computer Vision*, volume 1, pages 105–112 vol.1, 2001.

[14] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Lecture Notes in Statistics*, 37:28–47, 1986.

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

[16] A. Broadhurst, T. W. Drummond, and R. Cipolla. A probabilistic framework for space carving. In *Proc. International Conference on Computer Vision*, pages 388–393, July 2001.

[17] N. D. F. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Automatic 3D object segmentation in multiple views using volumetric graph-cuts. In *18th British Machine Vision Conference*, volume 1, pages 530–539, 2007.

[18] N. D. F. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Automatic 3D object segmentation in multiple views using volumetric graph-cuts. *Image Vision Comput.*, 28(1):14–25, 2010.

[19] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. Fifth International Conference on Computer Vision*, pages 694–699, Cambridge, MA, June 1995. IEEE Computer Society Press.

[20] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 2011.

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

[22] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.

[23] E. K. P. Chong and S. H. Zak. *Introduction to Optimization.* John Wiley & Sons, Inc., New York, second edition, 2001.

[24] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 2011.

[25] D. Cremers, T. Pock, K. Kolev, and A. Chambolle. Convex relaxation techniques for segmentation, stereo and multiview reconstruction. In *Advances in Markov Random Fields for Vision and Image Processing.* MIT Press, 2011.

[26] G. Cross and A. Zisserman. Surface reconstruction from multiple views using apparent contours and surface texture. In *Confluence of Computer Vision and Computer Graphics*, pages 25–47, Norwell, MA, USA, 2000. Kluwer Academic Publishers.

[27] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D shape scanning with a time-of-flight camera. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2010.

[28] A. Delaunoy, E. Prados, P. Gargallo, J.-P. Pons, and P. Sturm. Minimizing the multi-view stereo reprojection error for triangular surface meshes. In *Proceedings of the 19th British Machine Vision Conference, Leeds, UK.* BMVC, sept 2008.

[29] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh–Taylor instability. In R. Rautman, editor, *Approximation Methods for Navier–Stokes Problems*, volume 771 of *Lecture Notes in Mathematics*, pages 145–158. Springer, Berlin, 1979.

[30] W. Dinkelbach. On nonlinear fractional programming. *Management Science*, 13:492–498, 1967.

[31] Y. Duan, L. Yang, H. Qin, and D. Samaras. Shape reconstruction from 3D and 2D data using PDE-based deformable surfaces. In *Proc. European Conference on Computer Vision*, pages 238–251, 2004.

[32] R. L. Dykstra. An algorithm for restricted least-squares regression. *Journal of the American Statistical Association*, 78:837–842, 1983.

[33] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[34] C. H. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:548–554, 2008.

[35] A. X. Falcao, J. Stolfi, and R. A. Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:19–29, 2004.

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

[37] O. Faugeras, Q.-T. Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications.* MIT Press, Cambridge, MA, USA, 2001.

[38] J.-S. Franco and E. Boyer. Fusion of multi-view silhouette cues using a space occupancy grid. In *Proc. International Conference on Computer Vision*, Beijin, China, 2005.

[39] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, pages 1434–1441, 2010.

[40] Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. In *Proc. European Conference on Computer Vision*, pages 564–577, Graz, Austria, 2006.

[41] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, Minnesota, USA, June 2007.

[42] D. Gallup, J.-M. Frahm, and M. Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *CVPR*, pages 1418–1425, 2010.

[43] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil.* IEEE Computer Society Press, 2007.

[44] W. A. Geominy. *Die Florentiner Niobiden.* PhD thesis, University of Bonn, Germany, 1982.

[45] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 2402–2409, Washington, DC, USA, 2006. IEEE Computer Society.

[46] L. M. Goldschlager, R. A. Shaw, and J. Staples. The maximum flow problem is log space complete for p. *Theor. Comput. Sci.*, 21:105–111, 1982.

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

[48] L. Guan, J. S. Franco, and M. Pollefeys. 3D occlusion inference from silhouette cues. In *Proc. International Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007.

[49] M. Habbecke and L. Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Minneapolis, Minnesota, USA, June 2007.

[50] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, Cambridge, UK, second edition, 2004.

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

[52] J.-B. Hiriart-Urruty and C. Lemarechal. *Fundamentals of Convex Analysis.* Springer, Heidelberg, 2001.

[53] B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading.* MIT Press, Cambridge, MA, 1989.

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

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

[56] J. Isidoro and S. Sclaroff. Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *Proc. International Conference on Computer Vision*, pages 1335–1342, Washington, DC, USA, 2003.

[57] I. H. Jermyn and H. Ishikawa. Globally optimal regions and boundaries as minimum ratio weight cycles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1075–1088, 2001.

[58] M. Jung, E. Resmerita, and L. Vese. An iterative method with general convex fidelity term for image restoration. In *Proc. European Conference on Computer Vision*, volume 6311, pages 185–199, 2010.

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

[60] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, SGP '06, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association.

[61] R. Keriven. A variational framework to shape from contours. Technical Report 2002-221, CERMICS, 2002.

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

[63] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *IEEE International Conference on Image Processing*, volume 3, pages 797–800, Rochester, NY, June 2002.

[64] D. Kirsanov and S. Gortler. A discrete global minimization algotithm for continuous variational problems. In *Harvard Computer Science Technical Report: TR-14-04*, July 2004.

[65] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008.

[66] K. Kolev, T. Brox, and D. Cremers. Robust variational segmentation of 3D objects from multiple views. In K. F. et al., editor, *Pattern Recognition (Proc. DAGM)*, volume 4174 of *LNCS*, pages 688–697, Berlin, Germany, September 2006. Springer.

[67] K. Kolev, T. Brox, and D. Cremers. Fast joint estimation of silhouettes and dense 3D geometry from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2012.

[68] K. Kolev and D. Cremers. Integration of multiview stereo and silhouettes via convex functionals on convex domains. In *European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008.

[69] K. Kolev and D. Cremers. Continuous ratio optimization via convex relaxation with applications to multiview 3D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.

[70] K. Kolev, N. Kirchgessner, S. Houben, A. Csiszar, W. Rubner, C. Palm, B. Eiben, R. Merkel, and D. Cremers. A variational approach to vesicle membrane reconstruction from fluorescence imaging. *Pattern Recognition*, 44:2944–2958, 2011.

[71] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Propagated photoconsistency and convexity in variational multiview 3D reconstruction. In *ICCV Workshop on Photometric Analysis for Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[72] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Continuous global optimization in multiview 3D reconstruction. *International Journal of Computer Vision*, 84(1):80–96, August 2009.

[73] K. Kolev, M. Klodt, T. Brox, S. Esedoglu, and D. Cremers. Continuous global optimization in multiview 3D reconstruction. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 4679 of *LNCS*, pages 441–452, EZhou, China, August 2007. Springer.

[74] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *European Conference on Computer Vision (ECCV)*, Heraklion, Greece, September 2010.

[75] V. Kolmogorov and Y. Boykov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[76] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *Proc. International Conference on Computer Vision*, pages 564–571, Beijing, China, 2005. IEEE Computer Society.

[77] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. In *ICCV*, pages 1–8, 2007.

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

[79] A. Laurentini. The visual hull concept for visual-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[80] V. Lempitsky. Surface extraction from binary volumes with higher-order smoothness. In *CVPR*, pages 1197–1204, 2010.

[81] V. Lempitsky, Y. Boykov, and D. Ivanov. Oriented visibility for multiview reconstruction. In *Proc. European Conference on Computer Vision*, volume 3953 of *LNCS*, pages 226–238, 2006.

[82] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

[83] M. Levoy, S. Rusinkiewicz, B. Curless, M. Ginzton, J. Ginsberg, K. Pulli, D. Koller, S. Anderson, J. Shade, L. Pereira, J. Davis, and D. Fulk. The digital Michelangelo project: 3D scanning of large statues. In *SIGGRAPH, Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144, 2000.

[84] J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang. Bundled depth-map merging for multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, USA, June 2010.

[85] Y. Li, J. Sun, and H.-Y. Shum. Video object cut and paste. *ACM Trans. Graph.*, 24(3):595–600, 2005.

[86] J. Liu and J. Sun. Parallel graph-cuts by adaptive bottom-up merging. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2010.

[87] S. Liu and D. B. Cooper. Ray markov random fields for image-based 3d modeling: Model and efficient inference. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1530–1537, 2010.

[88] Y. Liu, X. Cao, Q. Dai, and W. Xu. Continuous depth estimation for multi-view stereo. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:2121–2128, 2009.

[89] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21:163–169, August 1987.

[90] M. Machline, Y. Arieli, A. Sphpunt, and B. Freedman. Depth mapping using projected patterns. Prime Sense Ltd., US patent 20100118123, 2010.

[91] W. N. Martin and J. K. Aggarwal. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):150–158, 1983.

[92] Y. Matsumoto, K. Fujimura, and T. Kitamura. Shape-from-silhouette/stereo and its application to 3D digitizer. In *Proceedings of Discrete Geometry for Computing Imagery*, pages 177–190, 1999.

[93] J. M. Morel and S. Solimini. *Variational methods in image segmentation.* Birkhauser Boston Inc., Cambridge, MA, USA, 1995.

[94] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE Computer Society Press.

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

[96] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.

[97] C. Olsson, M. Byröd, N. C. Overgaard, and F. Kahl. Extending continuous cuts: Anisotropic metrics and expansion moves. In *International Conference on Computer Vision*, 2009.

[98] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[99] 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)*, Jena, Germany, September 2009.

[100] PMVS. `http://www.cs.washington.edu/homes/furukawa/research/pmvs/`.

[101] T. Pock. *Fast Total Variation for Computer Vision.* PhD thesis, Graz, Austria, Jan. 2008.

[102] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

[103] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008.

[104] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision*, 78:143–167, July 2008.

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

[106] M. Potmesil. Generating octree models of 3D objects from their silhouettes from a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40(1):1–29, 1987.

[107] M. J. D. Powell. *Variable Metric Methods for Constrained Optimization*, pages 288–311. Springer Verlag, 1983.

[108] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

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

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

[111] C. Rother, V. Kolmogorov, T. Minka, and A. Blake. Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 993–1000, 2006.

[112] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 519–528, Washington, DC, USA, 2006. IEEE Computer Society.

[113] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 1067–1073, June 1997.

[114] J. A. Sethian. *Level Set Methods.* Cambridge University Press, Cambridge, UK, 1996.

[115] S. Sinha, P. Mordohai, and M. Pollefeys. Multiview stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *Proc. International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[116] S. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *Proc. International Conference on Computer Vision*, pages 349–356, Washington, DC, USA, 2005. IEEE Computer Society.

[117] D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *Proc. International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 345–353, 2000.

[118] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multiview stereo. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 974–981, October 2003.

[119] P. Strandmark and F. Kahl. Parallel and distributed graph cuts by dual decomposition. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2010.

[120] P. Strandmark, F. Kahl, and N. Overgaard. Optimizing parametric total variation models. In *International Conference on Computer Vision*, pages 2240–2247, 2009.

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

[122] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, June 2008.

[123] R. Szeliski. Rapid octree construction from image sequences. *Computer Vision, Graphics, and Image Processing*, 58(1):23–32, 1993.

[124] S. Tran and L. Davis. 3D surface reconstruction using graph cuts with surface constraints. In *Proc. European Conference on Computer Vision*, volume 3952 of *LNCS*, pages 219–231, 2006.

[125] Y.-H. R. Tsai, L.-T. Cheng, S. Osher, P. Burchard, and G. Sapiro. Visibility and its dynamics in a pde based implicit framework. *J. Comput. Phys.*, 199:260–290, September 2004.

[126] Y.-H. R. Tsai, L.-T. Cheng, S. Osher, P. Burchard, and G. Sapiro. Visibility and its dynamics in a pde based implicit framework. *J. Comput. Phys.*, 199:260–290, September 2004.

[127] G. Unal, A. Yezzi, S. Soatto, and G. Slabaugh. A variational approach to problems in calibration of multiple cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1322–1338, 2007.

[128] M. Unger, T. Pock, D. Cremers, and H. Bischof. TVSeg - interactive total variation based image segmentation. In *British Machine Vision Conference (BMVC)*, Leeds, UK, September 2008.

[129] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof. TVSeg - interactive total variation based image segmentation. In *British Machine Vision Conference 2008*, Leeds, UK, Sept. 2008.

[130] A. Vasilevskiy and K. Siddiqi. Flux maximizing geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1565–1578, 2002.

[131] V. Vineet and P. J. Narayanan. Cuda cuts: Fast graph cuts on the gpu. *Computer Vision and Pattern Recognition Workshop*, 0:1–8, 2008.

[132] G. Vogiatzis, C. H. Esteban, P. H. S. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:2241–2246, 2007.

[133] G. Vogiatzis, C. Hernandez, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:1847–1854, 2006.

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

[135] J. Wang, P. Bhat, R. A. Colburn, M. Agrawala, and M. F. Cohen. Interactive video cutout. *ACM Trans. Graph.*, 24(3):585–594, 2005.

[136] J. Weickert. *Anisotropic Diffusion in Image Processing.* PhD thesis, Department of Mathematics, University of Kaiserslautern, Germany, Jan. 1996. Revised and extended version published by Teubner, Stuttgart, Germany, 1998.

[137] R. White and D. A. Forsyth. Combining cues: Shape from shading and texture. In *In Proc. Conf. Computer Vision and Pattern Recognition*, pages 1809–1816, 2006.

[138] A. P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17:17–45, 1981.

[139] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineerings 19*, I:139–144, 1980.

[140] T.-P. Wu, S.-K. Yeung, J. Jia, and C.-K. Tang. Quasi-dense 3d reconstruction using tensor-based multiview stereo. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1482–1489, 2010.

[141] A. Yezzi and S. Soatto. Stereoscopic segmentation. In *Proc. 8th International Conference on Computer Vision*, volume 1, pages 59–66, Vancouver, Canada, July 2001.

[142] A. Yezzi and S. Soatto. Stereoscopic segmentation. *International Journal of Computer Vision*, 53(1):31–43, 2003.

[143] C. Zach, M. Niethammer, and J. M. Frahm. Continuous maximal flows and Wulff shapes: Application to MRFs. In *Proc. International Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 2009.

[144] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV-L1 range image integration. In *Proc. International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[145] C. Zach and M. Pollefeys. Practical methods for convex multi-view reconstruction. In *Proc. European Conference on Computer Vision*, volume 6314, pages 354–367, Heraklion, Greece, 2010.

[146] L. Zhang, B. Curless, and S. M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2003.

[147] J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2008.

[148] S.-C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, Sept. 1996.

[149] A. H. Zhuang, D. J. Valentino, and A. W. Toga. Skull-stripping magnetic resonance brain images using a model-based level set. *Neuroimage*, 32(1):79–92, 2006.