Chapter 16
A Review of Geometry Recovery from a Single Image Focusing on Curved Object Reconstruction

Martin R. Oswald, Eno Töppe, Claudia Nieuwenhuis, and Daniel Cremers

Abstract Single view reconstruction approaches infer the structure of 3D objects or scenes from 2D images. This is an inherently ill-posed problem. An abundance of reconstruction approaches has been proposed in the literature, which can be characterized by the additional assumptions they impose to make the reconstruction feasible. These assumptions are either formulated by restrictions on the reconstructable object domain, by geometric or learned shape priors or by requiring user input. In this chapter, we examine a representative set of state-of-the-art reconstruction approaches, which are applicable to real-world images. We classify the approaches according to their reconstruction objective and compare them based on a variety of important criteria. Finally, we show experimental comparisons for five curved object reconstruction approaches.

16.1 Introduction

Estimating 3D geometry from images has been a core research topic in Computer Vision for several decades. For the case of multiple input images a large variety of methods has been developed which are able to deliver high quality reconstruction results. For the special case that only a single still image is available the problem gets considerably more difficult. For specific assumptions imposed on the image a variety of methods to estimate 3D geometry exist in literature. However, a thorough comparison has not been carried out so far.

The reason for this lies partly in the significant diversity of existing approaches which in turn is due to the inherent ill-posedness of the underlying problem: during image formation, depth is irrecoverably lost. In their effort to make the problem

M.R. Oswald (✉) · E. Töppe · C. Nieuwenhuis · D. Cremers
Department of Computer Science, Institut für Informatik, TU München, Boltzmannstr. 3, Garching bei München, Germany
e-mail: oswaldm@in.tum.de; toeppe@in.tum.de; nieuwenhuis@in.tum.de; cremers@in.tum.de

M. Breuß et al. (eds.), *Innovations for Shape Analysis*, Mathematics and Visualization,
DOI 10.1007/978-3-642-34141-0_16,

tractable, single view methods have come up with an abundance of very different assumptions, methods and priors to infer the geometry of a depicted scene or object. The reconstruction precision of such approaches exceeds that of plausible estimates only in very few cases. Consequently, the reconstruction objectives are of very different nature, which makes a comparison difficult.

In this chapter we give a brief survey on the subject of single view reconstruction. We provide an introduction to the field and examine basic image information and assumptions that are used in order to compensate for ill-posedness. We then review, categorize and compare existing state-of-the-art approaches. Finally, we give a detailed theoretical and experimental comparison of five single view reconstruction methods for curved surfaces.

16.2 Single View Reconstruction

Single view reconstruction has the objective of recovering geometric information from a single photograph or synthetic image. The geometric information that is to be retrieved can be of very different manifestation reaching from purely relational information, sparse metrics or dense depth information to a complete 3D model of a single object or even a scene. This circumstance in combination with the inherent ill-posedness of the problem is the main reason for the strong diversity that is witnessed among the works in single view reconstruction and it is by no means a straightforward task to develop a taxonomy let alone a comparison.

A first approach will therefore be on the one hand to give an overview on the different types of image information ('*image cues*') used in the different reconstruction processes. On the other hand we will review the *priors* that are assumed in order to overcome the ill-posedness. This will also serve as a brief survey on existing works. Later in Sect. 16.3, we will classify a number of high-level approaches and compare them in Sect. 16.4. We will then narrow down further consideration to a subclass of methods that concentrates on reconstructing curved objects which will form the basis for our comparison in Sect. 16.5.

16.2.1 Image Cues

Approaches to single view reconstruction extract specific higher or lower level information contained in the input image either automatically or with the help of user input. This information then is interpreted to infer geometric relations of the depicted object or scene. In the following we list the most important categories and give prominent references.

Shading The problem of *Shape from Shading* (SfS) is to infer a surface (height field) from a single gray level image by using the gradual variation of shading that

is induced by the surface interaction of light. Some approaches also co-estimate lighting conditions and reflection properties. In general, the goal is to find a solution to the following image formation model

$$R(n(x)) = I(x) \ , \tag{16.1}$$

where I is the image, n is the normal field of the surface and R is the reflectance function which is dependent on the object. In most SfS approaches a Lambertian reflectance model is assumed. There are, however, other models which also include the specular case (e.g. Wang et al. [56]). SfS is an ill-posed problem, although there has been progress on deriving conditions for unique solutions by Prados and Faugeras [42].

As shown by Durou et al. [12] and Zhang et al. [62] reconstruction from real world images is limited in the sense that each approach exhibits special and sometimes unrealistic requirements on lighting conditions or reflection properties. Especially the presence of texture is an issue. Work has been done, however, to incorporate interreflection [39], shadowing and perspective projection [7] just to name a few. One of the first minimization approaches to SfS is by Ikeuchi and Horn [24]. For a current survey see Durou et al. [12].

Shadow The shadow that is thrown by objects conveys geometric information relative to the viewpoint of the light source. This information can be used to narrow down possible reconstruction results. Often point light sources have to be assumed as soft shadows do not provide enough information. Furthermore, shadow is not always thrown on known geometry, which makes the problem even more complex. Apart from reconstruction ambiguities, it is not straightforward to extract shadow borders from images. References include works by Daum and Dudek [9], Kender and Smith [29], Yu and Chang [60], and Hatzitheodorou [16].

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

Subclasses of contour edge-based methods are contour-based and silhouette-based reconstruction methods. *Shape from Contour* approaches try to infer geometry given the object contours alone. However, reconstructions are in most cases ambiguous. Especially, smooth surfaces often do not exhibit sufficient contour lines in the image. Shape from Contour approaches based on closed contour drawings include Horaud et al. [19], Ulupinar et al. [53], and Li et al. [33]. Karpenko et al. [27, 28] interpret user line drawings. Other single view reconstruction approaches that use contour edges for reconstruction include [10, 14, 17, 30, 47, 48].

Silhouette Closely related to Shape from Contour are approaches that infer geometry given the object silhouette. The silhouette is the image of the contour generator and the contour generator is the set of visible points on a surface, whose image rays are tangent to the surface.

The goal of silhouette based approaches is to find a geometric reconstruction, whose projection into the image plane agrees with the silhouette. As there are always infinitely many objects that are silhouette consistent this cue suffers from inherent ambiguity if used alone.

There are several silhouette based single view reconstruction algorithms that we will consider in more detail later. These include works by Prasad et al. [44, 45], Oswald et al. [41], and Töppe et al. [52]. Related to these approaches are a class of sketch based modeling tools e.g. by Igarashi et al. [23], Karpenko et al. [27], and Nealen et al. [40].

Texture Besides geometry, the appearance of real world objects is also determined by texture. Although a complete distinction from shading is not possible, texture is considered as an inherent property of an object rather than a result of an interaction of light and geometry.

If one assumes objects to have a regular and known texture it is possible to infer their geometry from the way the texture is deformed after image projection. These *Shape from Texture* approaches, obviously, impose strong constraints on the reconstructable objects. An example constitutes the work of Malik and Rosenholtz [37].

Further single view reconstruction algorithms that use texture cues include Super et al. [50], Hassner and Basri [15], and Vetter et al. [55]. Approaches that combine texture and contour edges for reconstruction by considering so-called 'superpixels' are Hoiem et al. [17] and Saxena et al. [48].

Defocus Due to physical aspects of image formation, the sharpness of a depicted object correlates with its distance to the camera. This fact can be used to infer a dense depth map from an image. However, the accuracy of such methods is limited and camera calibration is necessary. References include works from Levin [32] and Bae and Durand [1].

Location The location of objects in the image can infer semantic knowledge of the objects. For example, ground, floor or sky can be identified more easily from their location in the image. This information can be helpful for 3D reconstructions. Hoiem et al. [17] reconstruct vertical objects by distinguishing them from the ground and the sky. Delage et al. [10] use a Bayesian network to identify floor pixels.

16.2.2 Priors

Priors are of utter importance in single view reconstruction in order to compensate for the problem of ill-posedness. Depending on the ultimate reconstruction goal and the target group of objects, different priors or a combination of them can be applied. Priors can either be chosen in fixed form, or they can be gained by learning. Furthermore, there are low-level and high-level priors. In the following we will list priors that are most frequently assumed in single view reconstruction.

Smoothness Smoothness can be defined as the small spatial change of some property. In single view reconstruction we are often not able to infer a dense reconstruction. It is therefore good practice to choose among the possible reconstruction surfaces those which tend to be smooth. Smoothness naturally plays a significant role in the reconstruction of curved surfaces as in [41, 45, 52, 63].

Smoothness in a wider sense can also be learned as the consistency of object surfaces. Hoiem et al. [17] use a machine learning approach to find image features indicating the assignment of neighboring superpixels to the same object. Saxena et al. [48] use image cues and geometric relations to learn the relative depth of neighboring superpixels.

Geometric Relations Basic geometric relations are often encountered specifically in man-made environments. As a prior they can help to dissolve ambiguities in the reconstruction process. As part of those basic geometric relations we consider e.g. *coplanarity*, *collinearity*, *perpendicularity* and *symmetry*. An early work which makes use of such simple rules is the one of Lowe [36]. By assuming planes to be parallel or perpendicular one can also derive camera parameters (see Criminisi et al. [8]). This is even more important, as perspective projection is not angle-preserving and the image of parallel lines will not necessarily be parallel. We can often assume objects to stand vertically on the ground plane [10, 14, 17].

Symmetric objects exhibit identical sections, which are projected to different locations in the input image. Knowing that these parts possess similar geometric and reflectance properties one can interpret their respective projections as views of the same object part from different observation points. This can be seen as a weak multiview scenario providing more information for reconstruction [18]. Also, occluded geometry can be inferred from this prior [14].

Volume/Area With smoothness as a prior on its own, solutions tend to be trivial or flat depending on the reconstruction approach. Adding a volume prior to the reconstruction process will ensure an inflation of the object surface and will still result in a certain compactness of the solution due to the smoothness assumption. Volume priors can be found in Li et al. [33] and Töppe et al. [52].

Semantic Relations Semantic relations infer high-level knowledge on the relative position and inner structure of different objects and their depth values. Han and Zhu [14], for example, infer occluded points based on semantic human knowledge, e.g. that leaves are connected to the plant. Koutsourakis et al. [30] introduce semantic knowledge to ensure the consistency of different floors. Finally, knowledge on the location of the ground and the sky represents an important cue for 3D reconstruction. The ground is often used as starting point for the reconstruction as objects, especially walls, are usually perpendicular to this plane [10, 14, 17].

Shape Priors Shape priors impose high-level knowledge on the objects to be reconstructed. Among commonly used priors, full shape priors usually impose the strongest restrictions. On the one hand, this leads to a rather limited applicability of the approach. On the other hand, the reconstructions are usually of high quality and work automatically without user input.

Shape priors can be defined or learned. In [30], Koutsourakis et al. define a full shape grammar for the reconstruction of facades. This limits the approach to the reconstruction of buildings in urban environments. In contrast, Rother and Sapiro [46] and Chen and Cipolla [4] shape priors are learned from a database of sample objects. Hence, they are not a-priori limited to a specific object class. However, their choice of samples intrinsically limits their approach to the object classes represented in the database.

The representation of shape priors ranges from specified sets of grammar rules over parametric models to probabilistic priors. In [4], Chen and Cipolla learn depth maps of human bodies by means of principal component analysis. This model imposes strong assumptions on the 3D object, but the dimension of the state space is reduced and only valid 3D reconstructions are obtained. In contrast, Rother and Sapiro [46] impose less strong assumptions on the learned model. For each object class a shape prior is learned as the relative occupancy frequency of each voxel in the object.

16.3 Classification of High-Level Approaches

In this section we will examine selected works in the field of single view reconstruction. Due to the abundance and diversity of approaches we selected algorithms with respect to the following criteria: the chosen approaches are applicable to real world images that are not taken under special or unrealistic material or lighting conditions. We rather concentrate on approaches inferring objects from ordinary photographs where reconstruction plausibility is more important than precision. The selection, furthermore, focuses on works that are representative and state-of-the-art. We provide a classification and examine differences and similarities.

For classification we found several categories ranging from application domain over image cues and shape priors to user input and 3D representation (see below). However, these categories are not suitable to partition the set of approaches due to strong overlap. Instead, we think that the most important information for each single view reconstruction approach is its application domain, i.e. the set of objects and scenes, which can be reconstructed. We introduce the literature sorted by the following reconstruction categories:

- Curved Objects
- Piecewise Planar Objects
- Learning Specific Objects
- 3D Impression from Scenes

More categories will follow in the next subsection. We distinguish between objects and scenes. Reconstructions of scenes aim at producing 3D impressions or depth maps from the whole scene contained in the image. In contrast, object reconstruction approaches concentrate on single objects within the scene. Approaches that reconstruct *curved objects* principally aim at producing arbitrary, mostly smooth objects.

Often, minimal surface approaches are used, which try to minimize the surface of the object given a fixed area or volume. The second class consists of methods that concentrate on *piecewise planar objects* such as buildings and man-made environments. Furthermore, we distinguish arbitrary curved and planar objects from *learning specific objects*. Approaches in this class cannot reconstruct arbitrary objects, but are inherently limited to specific object classes by shape information learned from sample databases. Finally, we discuss methods that do not aim to reconstruct exact or plausible 3D geometry but rather provide a pleasing *3D Impression from Scenes*. In the following, we will present and classify related work on single view reconstruction.

16.3.1 Curved Objects

16.3.1.1 Zhang et al.

Zhang et al. [63] proposed a method for interactive depth map editing based on an input image. The depth map reconstruction is the result of minimizing a thin plate energy [11], which favors smooth surfaces and penalizes bending. User input comes as a variety of constraints on the thin plate energy that can be placed interactively into the depth map. These comprise of position constraints, surface normals, surface or normal discontinuities, planar region constraints or curves on which curvature or torsion is minimized.

From a mathematical point of view the thin plate energy for a continuous function f on a two dimensional rectangular domain $[0,1]^2$ is defined as:

$$E(f) = \int_0^1 \int_0^1 \left[\alpha(u,v) \left\| \frac{\partial^2 f}{\partial u^2} \right\|^2 + 2\beta(u,v) \left\| \frac{\partial^2 f}{\partial uv} \right\|^2 + \gamma(u,v) \left\| \frac{\partial^2 f}{\partial v^2} \right\|^2 \right] du\, dv \ , \tag{16.2}$$

where functions $\alpha, \beta, \gamma : [0,1]^2 \mapsto \{0,1\}$ extend the thin plate model with weighting functions which can be used to define local surface discontinuities. Zhang et al. [63] discretize this minimization problem by introducing a function $g_{i,j}$ that samples values of the depth map function $f : [0,1]^2 \mapsto \mathbb{R}$ on a discrete rectangular grid, that is, $g_{i,j} = f(id, jd)$, with d being the distance between neighboring grid points. For efficiency and accuracy the grid resolution can be locally refined by the user. By stacking all values $g_{i,j}$ into a single vector $\mathbf{g}$ and by discretizing the partial derivatives of g, the energy in Eq. (16.2) can be written in matrix form as

$$\mathbf{g}^T \mathbf{C} \mathbf{g} \qquad \text{subject to} \quad \mathbf{A}\mathbf{g} = \mathbf{b} \ , \tag{16.3}$$

where the condition $\mathbf{A}\mathbf{g} = \mathbf{b}$ may contain any constraints on the surface that can be expressed in linear form. For a detailed description on how the constraints are

incorporated into this quadratic optimization problem we refer to [63]. A description of these constraints from the user's point of view is given later together with the experimental comparison (Sect. 16.5.2).

16.3.1.2 Prasad et al.

The works [45] and [44] of Prasad et al. introduce a framework for single view reconstruction of curved surfaces. The method is related to the one by Zhang et al. [63] but aims at reconstructing closed surfaces.

The main idea involves computing a parametric minimal surface by globally minimizing the same thin plate energy (Eq. (16.2)) as Zhang et al. [63] the difference being, however, that they minimize with respect to a parametrized 3D surface $f : [0,1]^2 \mapsto \mathbb{R}^3$ instead of a depth map. As a result, function domain and image domain are no longer equivalent. For implementation purposes, the discretization of the optimization problem with constraints is done similar to Zhang et al. (see Eq. (16.3)).

The choice of constraints is mostly different from Zhang et al. [63]. The main source of reconstruction information is the silhouette: Prasad et al. [45] use the fact that normals along the contour generator $c(t)$ can be inferred from the 2D silhouette as by definition they are parallel to the viewing plane for a smooth surface. This leads to the constraints

$$\pi(f(u(t), v(t))) = c(t) \tag{16.4}$$

$$n(c(t))\, f(u(t), v(t)) = 0 \qquad \forall t \in [0,1] \ , \tag{16.5}$$

where $n(c(t))$ is the normal at the point $c(t)$ in $\mathbb{R}^3$ and π the orthographic projection function. The user has to determine the coordinates $(u(t), v(t))$ of the contour generator in parameter space. This is done by placing lines within the grid of the parameter space and setting them in correspondence with the parts of the contour generator. Similar to Zhang et al. [63] the user can employ position constraints to define the object inflation locally. Also, surface discontinuities can be optionally specified to relax the surface smoothness along curves in the parameter space.

Important to note is that in order to define the topology of the object, the user has to define which parts of the parameter space boundary are connected. For example, the connection of the left and right boundary defines a cylindrical shape of the function domain.

16.3.1.3 Oswald et al.

The variational single view approach by Oswald et al. [41] computes closed curved minimal surface reconstructions. The input comprises of a side view of an object and its silhouette which is obtained by interactive segmentation of the input image.

An object reconstruction is achieved by searching for a smooth minimal surface that complies with a shape based on the silhouette distance function.

The minimal object surface is represented by indicator function $u : V \mapsto \{0, 1\}$ defining object interior ($u = 1$) and exterior ($u = 0$) and is found by minimizing the following convex energy

$$E(u) = \underbrace{\int_V g(x)|\nabla u(x)|\, d^3x}_{\text{smoothness term}} + \underbrace{\int_V u(x)\, \big(\phi_{\text{vol}}(x) + \phi_{\text{sil}}(x)\big)\, d^3x}_{\text{data term}} \ , \tag{16.6}$$

where $V \subset \mathbb{R}^3$ is the reconstruction domain and the data term promotes volume inflation via ϕ_{vol} and silhouette consistency via ϕ_{sil}. Function $g : V \mapsto \mathbb{R}_+$ is used to locally adapt the smoothness of the surface if desired by the user – its default value is $g \equiv 1$. Given the input silhouette $S \subset \Omega \subset V$ as part of the image plane Ω a height map $h : \Omega \mapsto \mathbb{R}$ is defined

$$h(p) = \min \left\{ \lambda_{\text{cutoff}} \ , \ \lambda_{\text{offset}} + \lambda_{\text{factor}} * dist(p, \partial S)^k \right\} \tag{16.7}$$

with the help of the silhouette distance function $dist(p, \partial S) = \min_{s \in \partial S} \|p - s\|$, which returns the distance of any point $p \in \Omega$ to the silhouette boundary ∂S. The parameters $k, \lambda_{\text{cutoff}}, \lambda_{\text{offset}}, \lambda_{\text{factor}}$ modify the shape of h.

Now, the data term can be expressed in terms of h so that the object locally has a depth proportional to the silhouette distance. Using an implicit surface representation the data term can thus be written as

$$\phi_{\text{vol}}(x) = \begin{cases} -1 & \text{if } dist(x, \Omega) \leq h(\pi(x)) \\ +1 & \text{otherwise} \ , \end{cases} \tag{16.8}$$

where $\pi : V \mapsto \Omega$ is the orthographic projections of 3D points onto Ω. Silhouette consistency is brought forward by

$$\phi_{\text{sil}}(x) = \begin{cases} -\infty & \text{if } \chi_S(\pi(x)) = 1 \text{ and } x \in \Omega \\ +\infty & \text{if } \chi_S(\pi(x)) = 0 \\ 0 & \text{otherwise} \ , \end{cases} \tag{16.9}$$

where characteristic function $\chi_S : \Omega \mapsto \{0, 1\}$ indicates exterior or interior of the silhouette, respectively. The convex energy in Eq. (16.6) is minimized with the help of convex relaxation techniques and the global optimum defines the reconstructed object.

16.3.1.4 Töppe et al.

Very similar to Oswald et al. [41], the variational single view approach by Töppe et al. [52] also computes closed curved minimal surface reconstructions. It follows the same work flow and also requires an object silhouette as input.

The same regularizer as in Oswald et al. [41] makes for the smooth surface, however, instead of using a heuristic data term for surface inflation, the authors suggest that the surface volume complies with a user specified volume V_t. This is achieved by adding a constraint to the minimization of the following convex energy:

$$E(u) = \int_V g(x)|\nabla u(x)|\, d^3x \qquad \text{s.t.} \quad \int_V u(x)\, d^3x = V_t \ . \tag{16.10}$$

Again, the surface is represented implicitly by the indicator function $u : V \mapsto \{0, 1\}$ where $V \subset \mathbb{R}^3$ is the reconstruction domain. Similarly, the purpose of function $g : V \mapsto \mathbb{R}_+$ is to account for optional local smoothness changes (default: $g \equiv 1$). Silhouette consistency is enforced by another constraint

$$u(x) = \begin{cases} 0, & \pi(x) \notin S \\ 1, & x \in S \end{cases} \ , \tag{16.11}$$

where $\pi : V \mapsto \Omega$ is the orthographic projection of 3D points onto the image plane Ω and $S \subset \Omega$ is the input silhouette. Convex relaxation techniques are used to minimize energy in Eq. (16.10) and a GPU implementation allows for interactive rates on standard PC hardware.

The benefit of the implicit surface representation used by Oswald et al. [41] and by Töppe et al. [52] is the topological freedom of the object's surface. In contrast to Prasad et al. [45] and Zhang et al. [63] the surface can contain any number of holes that are induced by the input silhouette. As opposed to [45], however, both reconstruction approaches by Oswald et al. [41] and Töppe et al. [52] assume the contour generator to be planar, which again limits the set of possible reconstructions.

16.3.1.5 Colombo et al.

Another approach to 3D reconstruction are surfaces of revolution (SORs) [6,54,59]. They are common in man-made objects and represent a subclass of Straight Homogeneous Generalized Cylinders. SOR approaches strongly rely on the assumption of rotational symmetry of the objects. A surface of revolution is obtained by revolving a planar curve, referred to as scaling function, around a straight axis, the revolution axis. SORs can be parametrized in polar coordinates:

$$S(\theta, z) = (\rho(z)cos(\theta), \rho(z)sin(\theta), z) \ . \tag{16.12}$$

Important for reconstruction approaches are cross sections and meridians. Cross sections are intersections of planes orthogonal to the revolution axis with the object, which leads to circles. Meridians are intersections of planes containing the revolution axis with the object. This leads to curves which all have the same shape and coincide with the SOR scaling function. Reconstruction approaches based on SORs try to infer the scaling function and the axis of revolution from the image.

In Colombo et al. [6], the task of 3D reconstruction is formulated as the problem of determining the meridian curve from the imaged object silhouette and two given imaged cross sections. Based on the computation of fixed entities such as the vanishing line or the SOR symmetry axis, camera calibration can be done and the SOR is inferred. Texture acquisition is obtained by inverse normal cylindrical projection.

16.3.1.6 Other Approaches

Francois and Medioni [13] present an interactive 3D reconstruction method based on user labeled edges and curves, which are represented by non-uniform rational basis splines (NURBS). The reconstructed objects are either modeled as generalized cylinders or as a set of 3D surfaces. Terzopoulos et al. [51] propose deformable elastic 3D shape models, which evolve around a symmetry axis and whose projection into the image is attracted by strong image gradients. Cohen and Cohen [5] propose a generalization of snakes to 3D objects based on a sequence of 2D contour models for medical images.

16.3.2 Piecewise Planar Objects and Scenes

16.3.2.1 Criminisi et al.

In [8], Criminisi et al. describe how 3D affine measurements can be obtained from a single perspective view. The authors concentrate on scenes containing planes and parallel lines which are obtained by perspective projection. The camera is assumed to be uncalibrated with unknown internal parameters.

The authors assume that a vanishing line of a reference plane as well as a vanishing point can be computed from the image. Given the vanishing line of a reference plane, a vanishing point for a direction not parallel to the plane and a known reference length in the image, the authors derive affine scene structure from the image. In this context, three different types of measurements are computed from the image: distances between planes parallel to the reference plane, measurements on these planes (e.g. length and area ratios) and the camera position. These measurements are obtained from cross-ratios (special numbers associated with an ordered quadruple of collinear points, which is invariant under specific transformations in projective geometry) and specific image mappings.

To simplify computations, Criminisi et al. [8] developed an algebraic approach based on a parametrization of the projection matrix. As the approach relies on feature detection in the image (e.g. for reference points), the resulting measurements can only be of limited accuracy, which is assessed by a confidence measure.

The approach is applied among others to 3D reconstruction. Based on given reference heights (e.g. the true height of a window and a pillar of a house) and a precomputed vanishing line and point in the image the complete 3D reconstruction can be obtained. The position of the camera is also estimated.

16.3.2.2 Delage et al.

Delage et al. [10] describe an approach for the automatic reconstruction of 3D indoor scenes, which mainly consist of orthogonal planes. The following assumptions are made: (1) The image is obtained by perspective projection with known camera calibration. (2) The objects in the scene are composed of planes and edges with only three mutually orthogonal orientations ('Manhattan world assumption'). (3) The camera's axis is vertical to the floor with known height.

To distinguish planes from edges with their orientations, the authors devise a Markov Random Field (MRF) consisting of vertices V and edges E with six labels $y_v = \{p_1, p_2, p_3, e_1, e_2, e_3\}$ for $v \in V$. $\{p_1, p_2, p_3\}$ encodes plane orientation and $\{e_1, e_2, e_3\}$ encodes edge orientation. Let x_v and $x_{u,v}$ denote the feature vector at node v and at node pair (u, v) respectively. The following joint probability distribution is defined over all labels y_v:

$$P_\theta(y|x) = \frac{1}{Z_\theta(x)} \exp\left(-\sum_{v \in V} \psi_1(y_v, x_v, \theta_1) - \sum_{(u,v) \in E} \psi_2(y_u, y_v, x_{u,v}, \theta_2) \right) . \tag{16.13}$$

Here, $Z_\theta(x)$ stands for the partition function and $\theta = [\theta_1, \theta_2]$ indicates the model parameters. The unary term $\psi_1(y_v, x_v, \theta_1) = \theta_1^T \phi(y_v, x_v)$ is a linear combination of image features $\phi(y_v, x_v)$ such as edge orientation or floor affiliation, whereas the pairwise term $\psi_2(y_u, y_v, x_{u,v}, \theta_2) = \theta_2^T \phi(y_u, y_v, x_{u,v})$ encodes features indicating label consistency or if two pixels were members of the same partition after a segmentation of the image.

To obtain a 3D reconstruction from the MRF labeling, a constrained iterative optimization problem is formulated to obtain the location and orientation of planes and edges in 3D space. The constraints yielding 3D points for each image pixel can be formulated from the Manhattan world assumption which ensures a unique solution. To obtain a correct result, occlusion vs. non-occlusion of edges is inferred. Delage et al. [10] can be viewed as a modification and generalization of Sturm and Maybank [49].

16.3.2.3 Koutsourakis et al.

In [30], Koutsourakis et al. generate urban 3D reconstructions from images by estimating the parameters of a 3D shape grammar in a MRF approach, so that the generated building best matches the image.

A shape grammar describes how basic shapes interact together through a set of replacement rules to produce complex structured geometries. It contains basic or atomic shapes, which are modified by operators. The operators either transform the object by means of translation, rotation or scaling, or they perform structural operations such as splitting or mirroring. The main advantages of using a shape grammar are that it always produces well-defined buildings and that the complexity of the optimization as well as the dimensionality of the problem is strongly reduced.

For optimization the authors formulate a MRF approach. The unary terms ensure that object boundaries coincide with image boundaries, whereas the pairwise terms measure the appropriateness of the configuration of atomic shapes and ensure the consistency between the operator and the image.

16.3.2.4 Other Approaches

Kanade [26] recovers shape from geometric assumptions. The world is modeled as a collection of plane surfaces, which allows for a qualitative object recovery. Quantitative recovery is achieved by mapping image regularities into shape constraints. Piecewise planar scenes are computed in Liebowitz et al. [35] based on camera and geometric constraints such as parallelism and orthogonality, e.g. for the reconstruction of buildings. Apart from symmetry and planarity, two additional shape constraints are introduced by Li et al. [33] for object reconstruction: maximum compactness and minimum surface. Instead of computing vanishing lines, Kushal et al. [31] perform 3D reconstruction of structured scenes by registering two user indicated world planes. Hong et al. [18] study the relation between symmetry of objects and the viewer's relative pose to the object. An important principle for the reconstruction of symmetric objects is that one image of a symmetric object is equivalent to multiple images. Li et al. [34] describe a method for reconstructing piecewise planar objects by using connectivity and perspective symmetry of objects.

16.3.3 Learning Specific Objects

16.3.3.1 Han and Zhu

Han and Zhu [14] propose a 3D reconstruction approach based on manually defined shape priors, which can on the one hand be applied to polyhedral objects and on the other hand to grass and tree-like objects. They argue that learning priors is hard in practice, because there is not enough real world training data available. Hence, they revert to manually defined prior models.

The image is represented by two layers, one containing image regions such as sky or planar objects such as polyhedra, the other containing this curved structures such as grass or trees. The full 3D scene is represented by two graphs, one consisting of 3D objects, the other representing the relations between the objects in the scene. The objective of the paper is then to optimally reconstruct the 3D scene given the layer representation of the image in a Bayesian approach. To this end, the authors manually introduce prior knowledge. For polyhedral objects they assume planarity of faces as well as similar angles and edge lengths. For grass and tree-like objects they assume smooth and evenly spread curves. The relation graph is used to impose regularity constraints on touching objects, e.g. on angles. Furthermore,

hidden information is inferred based on human geometric reasoning on missing vertices, parallelism, rotational or mirror symmetry and stability assumptions.

Optimization is done by Markov Chain Monte Carlo Methods with reversible jumps which define rules for altering sub-graph structures.

16.3.3.2 Rother and Sapiro

Rother and Sapiro [46] present a general framework for pose estimation, 2D segmentation, object recognition and 3D reconstruction from a single image. The approach is well-suited to reconstruct bounded objects, but not for elaborate scenes.

The 3D object to be reconstructed is represented by voxels having either state full or empty: $V = \{V_i\}_{i=1}^{M},\ V_i \in \{empty, full\}$. The corresponding segmentation of the image is given by the pixel states $Q = \{Q_j\}_{j=1}^{N},\ Q_j \in \{background, foreground\}$. It is assumed that the camera matrix is given, which relates 3D voxels to 2D image pixels.

To obtain the most likely 3D reconstruction from the image, the authors formulate a graphical model based on two fidelity terms: (1) The object fidelity $P(V|K)$, which means its conformity with a learned 3D shape prior for a given shape class K. The shape prior $P(V|K)$ is learned for each class separately from sample images as the relative frequency of the voxel for being part of the object. (2) The data fidelity, which defines the probability that the given input image is generated by the occupied voxels under projection onto the image plane. This term is again composed of two probabilities for each of the N pixels in the image plane: (a) The color probability given by the user specified foreground and background color model for each pixel with color C_j, $P(C_j|Q_j)$, and (b) the probability of obtaining a particular pixel state (foreground or background) based on the number of full voxels projected onto pixel j along its corresponding projection ray $R(j)$, $P(Q_j|V_{R(j)})$.

The likelihood for a given hypothesis (shape class and affine transformation or pose) is then defined as follows:

$$L(Q,V) = \left(\prod_{j=1}^{N} P(C_j|Q_j) P(Q_j|V_{R(j)}) \right) P(V|K) \ . \tag{16.14}$$

To select the best hypothesis for each shape class and pose, their likelihood must be compared. In order to reduce computational complexity, an efficient branch and bound algorithm is proposed to discard suboptimal hypotheses and refine only plausible ones.

16.3.3.3 Chen and Cipolla

Chen and Cipolla [4] propose to infer 3D information directly from learned shape priors. They assume a number of given training shapes each consisting of the silhouette and the corresponding depth map. In a first step, the training silhouettes

are registered. To ensure the independence of different data dimensions and to reduce the complexity of the approach, Principal Component Analysis is applied to the silhouettes and the depth maps separately in order to find lower-dimensional subspaces of the training shapes. In this way, the authors obtain 2D lower-dimensional training feature pairs consisting of position and depth information. To model the low-dimensional sub-manifold structure underlying the feature pair space, Gaussian Process Latent Variable Models (GPLVM) are used. They assume Gaussian processes for the position and depth information as well as for the latent variables (the PCA coefficients). Then they estimate the underlying parameters in an optimization problem, which leads to a shape model learned from sample data.

Given an unknown silhouette, 3D information can now be inferred from the learned shape model. First, the silhouette is registered and projected into the lower-dimensional PCA subspace yielding a position feature vector. Given this feature vector and the shape model, we ask for the most likely depth estimate at each point. Since there is no closed-form solution to this problem, the authors revert to a two-stage approach: (1) They find the most likely latent variables generating the given silhouette feature. (2) From these latent variables the most likely depth estimate is inferred as the mean value of a Gaussian distribution. The final 3D reconstruction at each point is the sum of the most likely depth value and the linear combination of the PCA eigenvectors determined by the latent variables.

16.3.3.4 Hassner and Basri

Hassner and Basri [15] aim at depth reconstruction from a single image based on examples. The samples are given in a database S containing mappings of images to their corresponding depth maps $S = \{M_i\}_{i=1}^n = \{(I_i, D_i)\}_{i=1}^n$. For a given image patch W_p centered on p its depth is inferred from known depth values of the most similar patches V in the database by maximizing its plausibility: $Plaus(D|I, S) = \sum_{p \in I} maxSim(W_p, V)$. Similarity Sim between patches is measured in the least squares sense. The image patches overlap leading to several depth estimates for each pixel. These are combined by averaging. To ensure consistency of neighboring patches a global optimization procedure is proposed which iteratively refines depth estimates.

16.3.3.5 Other Approaches

Vetter [55] learned a parametric model for the reconstruction of faces by applying PCA to a database of registered 3D faces. Then the model parameters can be found, which best explain the given image of a face. In Nagai et al. [38], objects are learned from a sample database. A Hidden Markov Model is used to model the correspondence between intensity and depth.

16.3.4 3D Impression from Scenes

16.3.4.1 Hoiem et al.

In [17], Hoiem et al. propose a fully automatic approach for creating 3D models from single photographs, which is similar to the creating of pop-up illustrations in children's books. They divide the world into ground, sky and vertical objects, which they call geometric classes. The appearance of these classes is described by image cues, which are learned from sample images.

In a first step the image is automatically segmented into superpixels, which are grouped to constellations. Constellations consist of superpixels, which are likely to have the same label. This probability is expressed by

$$S(C) = \sum_{k=1}^{N_c} \frac{1}{n_k(1-n_k)} \sum_{i,j \in C_k} \log P(y_i = y_j \,|\, |z_i - z_j|) \ , \tag{16.15}$$

where N_c is the number of constellations, n_k the number of superpixels in constellation k and $P(y_i = y_j \,|\, |z_i - z_j|)$ is the estimated probability that two superpixels have the same label y given the absolute difference of their feature vectors z.

The likelihood of a superpixel label given its feature vector x is estimated by marginalizing over the constellation likelihoods

$$P(y_i = t|x) = \sum_{k:s_i \in C_k} P(y_k = t|x_k, C_k) P(C_k|x_k) \ , \tag{16.16}$$

where both probabilities are learned from training data. By means of a machine learning approach the constellations are labeled as ground, sky or vertical objects. To reconstruct a 3D model from the data, the boundary of the bottom of the vertical regions is fit with the ground, the sky is removed from the model, and vertical pixels are assumed to belong to objects sticking out of the ground. Line segments are then fit to the ground-vertical label boundary and the segments are formed into poly-lines. The image is finally 'folded' along the poly-lines and 'cut' upward at the endpoints of these lines as well as at ground-sky and vertical-sky boundaries. A reasonably scaled camera model can be obtained by estimating the horizon line and setting the remaining parameters to constants.

16.3.4.2 Saxena et al.

In [48], Saxena et al. propose another approach for obtaining 3D structure from a single image of an unstructured environment. The only assumption the authors make is that the world consists of small planes, whose 3D position and orientation is to

be estimated. Similar to Hoiem et al. [17], the authors start out from a superpixel segmentation of the image. In this way, they obtain N superpixels (miniature planes) each containing S_i image pixels. But instead of grouping superpixels into constellations defining ground, sky and vertical object classes, for each single superpixel the depth and orientation is inferred. This is done by a Markov Random Field (MRF) model. Each plane or superpixel is represented by a vector α such that all plane points q fulfill the equation $\alpha^T q = 1$. For a plane pixel lying on a normalized camera ray vector R at depth d we thus have $\alpha^T(dR) = 1$. The authors use a learning approach to obtain parameters θ which identify feature vectors X with depth estimates $\hat{d} = X^T\theta$. The authors maximize the following probability in an MRF approach

$$P(\alpha|X, Y, R, \theta) = \frac{1}{Z}\prod_{i=1}^{N} f_\theta(\alpha_i, X_i, Y_i, R_i) \prod_{i,j\in\{1,..,N\}} g(\alpha_i, \alpha_j, Y_{ij}, R_i, R_j) \ , \tag{16.17}$$

where α_i is the plane parameter of the superpixel i, X_i its corresponding feature vector, Y_i indicates feature confidence and R_i is the set of rays connecting the camera viewpoint to the pixels of the superpixel. Then f_θ relates image features to estimated depth by

$$f_\theta(\alpha_i, X_i, Y_i, R_i) = \exp\left(-\sum_{s_i=1}^{S_i} y_{i,s_i}|\alpha_i^T(X_{i,s_i}^T\theta R_{i,s_i}) - 1|\right) \ . \tag{16.18}$$

The pairwise term g can contain different definitions which capture features such as coplanarity, connectedness and colinearity.

The MRF represents two important aspects of depth estimation: it learns, (1) how visual cues such as color and texture are associated with depth, and (2) the relative depth of nearby superpixels based on geometric relations such as connectedness, coplanarity and collinearity. They also give estimates for occlusions and folds of the object.

Learning can be done only approximately due to the complexity of the model. To this end, the graphical model is approximated by a product of several marginal conditional likelihoods. MAP inference to infer the plane position and orientation of each superpixel is done by solving a linear program. The result is a polygonal mesh representation of the 3D model.

16.3.4.3 Other Approaches

In Horry et al. [20], simple 3D scenes are reconstructed based on user input such as vanishing points and foreground objects. The background of the scene is then modeled by rectangles, the foreground by hierarchical polygons. Barinova et al. [2] propose a reconstruction approach for urban scenes yielding visually pleasant results. The method is based on fitting 3D models containing vertical walls and ground plane to the scene.

16.4 General Comparison of High-Level Approaches

In the previous section we have presented important high-level single view reconstruction approaches. In this section, these approaches will be compared with respect to several categories, which we find important for their successful application. Table 16.1 compares the presented approaches with respect to these categories. It also indicates related image cues and shape priors for each approach, which were described in the previous section.

Applicability The applicability of a reconstruction approach depends on several aspects. First, methods are usually limited to some **reconstruction domain** which refers to the set of objects or scenes which can be reconstructed "successfully". These domains range from architectural such as buildings over man-made objects and piecewise planar environments to natural scenes. The examined single view approaches are therefore listed with respect to the categories introduced in Sect. 16.3 (first column of Table 16.1).

The applicability of an approach is also characterized by the **assumptions** made by the method. If specific assumptions are not met, the reconstruction process easily fails. Assumptions for each method are given in column five of Table 16.1. Typical assumptions are a calibrated camera [10], a simplified scene composition [10, 17], an object database containing samples for learning shape priors [4, 46], a specific viewpoint [41, 45, 52] or given geometric properties such as vanishing lines of reference planes [8].

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

Reconstructions can be exact, if the computed lengths and orientations of the inferred 3D objects accurately correspond to the true object. This is usually only possible from a single image if strong assumptions are made, e.g. piecewise planarity with only three orientations (Manhattan assumption) [10] or known reference heights and a calibrated camera [8]. Since such strict assumptions strongly limit the applicability of the approach, most approaches revert to computing the most likely solution to the ill-posed reconstruction problem without guaranteeing accuracy. The probability of a solution is usually measured by means of manually defined [14] or learned shape priors [4, 46]. We call this a *plausible* precision. Finally, there are approaches, which do not aim at reconstructing the real object. Instead, they find solutions which look good to the viewer when animated [17, 20, 48] or can be used to synthesize approximate new views of a scene. We call these reconstructions *pleasing*. The reconstruction precision is indicated in the third column of Table 16.1.

Table 16.1 Detailed overview of single view methods: for each approach the most important characteristics are indicated: Precision of the method (exact '$=$', plausible '$\simeq$', pleasing '$\approx$'), the representation of the 3D object, important assumptions made by the approach, the necessary user input and image cues as well as priors which are used in the reconstruction process. The 'L' indicates a prior which is not assumed but learned by the approach. Terms in brackets are optional.

						Image Cues				Priors				
	Method	Precision	Surface Representation	Assumptions	User Input	Silhouette	Edges	Location	Texture	Smoothness	Volume / Area	Semantic Relation	Geom. Relation	Shape
Curved Objects	Prasad et al. [45]	$\simeq$	[closed] parametric	charateristic sideview, max. genus 2	contours, [creases]	x			x	x				
	Zhang et al. [63]	$\simeq$	depth map	none	constraints					x				
	Oswald et al. [41]	$\simeq$	closed implicit	sideview, symmetry	silhouette, [creases], [data term]	x				x				
	Töppe et al. [52]	$\simeq$	closed implicit	sideview, symmetry	silhouette, [creases], [volume]	x				x	x			
	Colombo et al. [6]	$\simeq$	closed parametric	rotational symmetry	silhouette, cross sec.	x							x	
Piecewise Planar	Criminisi et al. [8]	$=$	non-closed polygonal	vanishing line, refer. height, ground plane	all edges to be measured		x						x	
	Delage et al. [10]	$\simeq$	non-closed polygonal	calibration, Manhattan	none		x	x	x	x		x	x	
	Koutsourakis et al. [30]	$=$	closed polygonal	rectified image, buildings	none		x		x			x		shape grammar
Learning Specific Objects	Han & Zhu [14]	$\simeq$	closed polygonal	polyhedra, plants	none		x			x		x	L	proba-bilistic
	Rother & Sapiro [46]	$\simeq$	closed implicit	calibration, color models, database	none			x						learned voxel model
	Chen & Cipolla [4]	$\simeq$	depth map	database	silhouette	x								learned PCA model
	Hassner & Basri [15]	$\simeq$	depth map	fixed view, spec. object database	none				x	x				learned example based
Scenes	Hoiem et al. [17]	$\approx$	pw. planar depth map	simple scene: sky, vertical walls&ground	none		x	x	x	L		x		
	Saxena et al. [48]	$\approx$	pw. planar depth map	world consists of planes	none		x	x	x	L			x	

'$=$' indicates exact precision, '$\simeq$' plausible precision and '$\approx$' a pleasing approach. Surely there are smooth transitions between these classes.

Representation The form of representation is closely connected to the reconstruction algorithm. Firstly, only those objects are reconstructable that can be adequately represented. Seen the other way, the representation has to reflect the reconstruction domain well. And secondly, the representation model has to conform to the reconstruction process.

Different representations have been used for single view reconstruction approaches. We distinguish between parametric and implicit representations. Each point on a *parametric surface* can be uniquely described by a coordinate. Finding a good parametrization for an object is not straightforward and generally does not allow for arbitrary topology. *Implicit surfaces* are a remedy to this problem. In this case, the surface is a level set of a function defined on $\mathbb{R}^3$. In contrast to parametric surfaces, single points on the surface are not easily addressed. *Polygonal* surface representations are neither parametric nor implicit and can be described as a planar graph with nodes, edges and faces. Note that polygonal surfaces often describe piecewise planar objects but are also used for approximating curved parametric surfaces. Finally, representations can describe *closed* and *non-closed* 3D surfaces. As a special case we also regard *depth maps*, which assign a depth to each pixel.

User Input and Runtime Completely automatic reconstruction on a single input image is often not feasible. Therefore, the user may be required to give cues on important image features. Most single view approaches aim to keep user input simple. User input can convey low-level and high-level information. High-level input is of semantic quality which helps to dissolve ambiguities, e.g. the object silhouette.

This stands in contrast to tools, where the user *models* the reconstruction with the help of low-level operations, e.g. by specifying surface normals or cutting away object parts. Many of these *modeling tools* [3, 25, 58] are not image-based and therefore only remotely related to single view reconstruction. In *Sketch-based modeling tools* [23,27,40,61] such modeling operations are driven by user indicated lines. The Teddy tool will be examined in more detail in Sect. 16.5. A pioneering work on free-form modeling was done by Welch and Witkin [57].

There is 2D and 3D user input. Most approaches use 2D input which in most cases is directly applied to the input image [52]. This involves tracing contour edges such as creases or vanishing lines. 3D input is directly applied to the reconstruction surface and is often more involved for the user as he needs to navigate in 3D space (e.g. specifying normals).

For some approaches the user input stage is separate from the reconstruction stage [4, 8]. Other methods compute a first reconstruction, then the user can add further input and the process is continued [41, 45, 52, 63]. For approaches of the latter kind, runtime is obviously an important factor of the reconstruction approach.

16.5 Comparison of Approaches for Curved Surface Reconstruction

In this section we concentrate on methods that aim for the reconstruction of curved surfaces. In particular, we discuss the methods by Zhang et al. [63], Prasad et al. [45], Oswald et al. [41], Töppe et al. [52], and Igarashi et al. [23]. Although Igarashi et al. is a pure modeling tool it inflates object contours and is thus related to silhouette based single view methods. Comparison is done theoretically and experimentally. The user input is discussed separately.

16.5.1 Theoretical Comparison

In the following we will compare the aforementioned approaches with respect to four topics which are important in surface reconstruction.

The Inflation Problem A common problem of approaches for curved surface reconstruction is that reconstructions tend to be flat since – by default – there are no inflation forces present due to a lack of depth information. A remedy is to let the user specify the depth of certain constraint points of the reconstruction which are then interpolated by the minimal surface [41, 45, 63]. This is tedious for the user. The depth constraints can be estimated fully automatically from the silhouette only for cylindrical objects as is done in some examples by Prasad et al. [43]. Several heuristics are conceived for more complicated cases. Igarashi et al. [23] automatically set the depth by a heuristic based on a signed distance function of the silhouette outline. A similar heuristic is used by Oswald et al. [41] in order to define a data term for their variational minimal surface approach. However, in contrast to Igarashi et al. [23] the user is able to adapt the parameters of this data term and thus the final surface. Töppe et al. [52] use a constraint on the volume of the reconstruction, which in many cases leads to natural inflation behavior.

Surface Representation and Topology The reconstructability of curved surfaces with arbitrary topology depends on the surface representation. Implicit representations [41, 52] are generally better suited for this task than parametric ones [45, 63], since the parameter space has to reflect the topology. The same holds for mesh-based approaches such as the one by Igarashi et al. [23]: during modeling operations it can be tedious to keep the mesh consistent, especially during topology changes.

The parametric representation by Prasad et al. [45] has other implications. Firstly, uniformly distributed points in the parameter space are not uniformly distributed on the surface. This property may lead to oscillations, especially in the case of higher genus. Further, the relation between points in parameter space and points on the surface is non-trivial for inexperienced users.

Silhouettes Silhouettes are used by Prasad et al. [45], Oswald et al. [41], and Töppe et al. [52] for surface inference. Full silhouette consistency of the reconstruction, however, is only enforced in the latter approaches as Prasad et al. [45] derive merely local constraints from the silhouette.

View Constraints Finally, view constraints are of practical importance. Oswald et al. [41] as well as Töppe et al. [52] assume symmetric objects. Reconstructions work best if symmetry and viewing plane are parallel. This implies that the contour generator is planar. The approach by Prasad et al. [45] allows for non-planar contour generators and, thus, in some cases for slightly more general view points than just a side-view.

16.5.2 Experimental Comparison

In this section we experimentally compare the five methods discussed in the previous subsection. For all experiments, we used the single view modeling tool by Zhang et al. [63] and the software called SmoothTeddy which incorporates results of several works by Igarashi et al. [21–23]. Both of them are publicly available for download. The reconstruction results by Prasad et al. are taken from the works [43–45]. For the reconstruction results of the method by Oswald et al. [41] and Töppe et al. [52] we used our own C++ and CUDA-based implementations.

In Figs. 16.1–16.3 we compare the reconstruction results of all five methods on ten different examples, covering various issues such as object shape, topology, viewing angle and image type. Instead of explaining every example, we rather concentrate on the examples which demonstrate properties, advantages or drawbacks discussed in the theoretical comparison as well as issues we identified during the experiments. Since the necessary amount of user input and thus the simplicity of the modeling process is of particular interest for practical purposes, we also explain and compare the user input for each method. The user inputs for each method is summarized in Table 16.2. Necessary user input is printed in bold. The other inputs are either optional or the program provides a heuristic to initialize these values reasonably well.

For each method we discuss the user input separately and illustrate it in several Figures for the teapot example, which is the most sophisticated one. Since we identified significant differences in the necessary amount of time a medium experienced user needs to generate the above examples, we quantitatively compare the amount of user input by listing the modeling times for every method and each example in Table 16.3.

The modeling times for Zhang et al. [63] include all user defined constraints for a medium experienced user. The respective times for Igarashi et al. [23] only contain the modeling time without model coloring as an automated model texturing could easily be integrated. The modeling times for Oswald et al. [41] and Töppe et al. [52] include the user scribbles for the silhouette extraction, which is rather simple

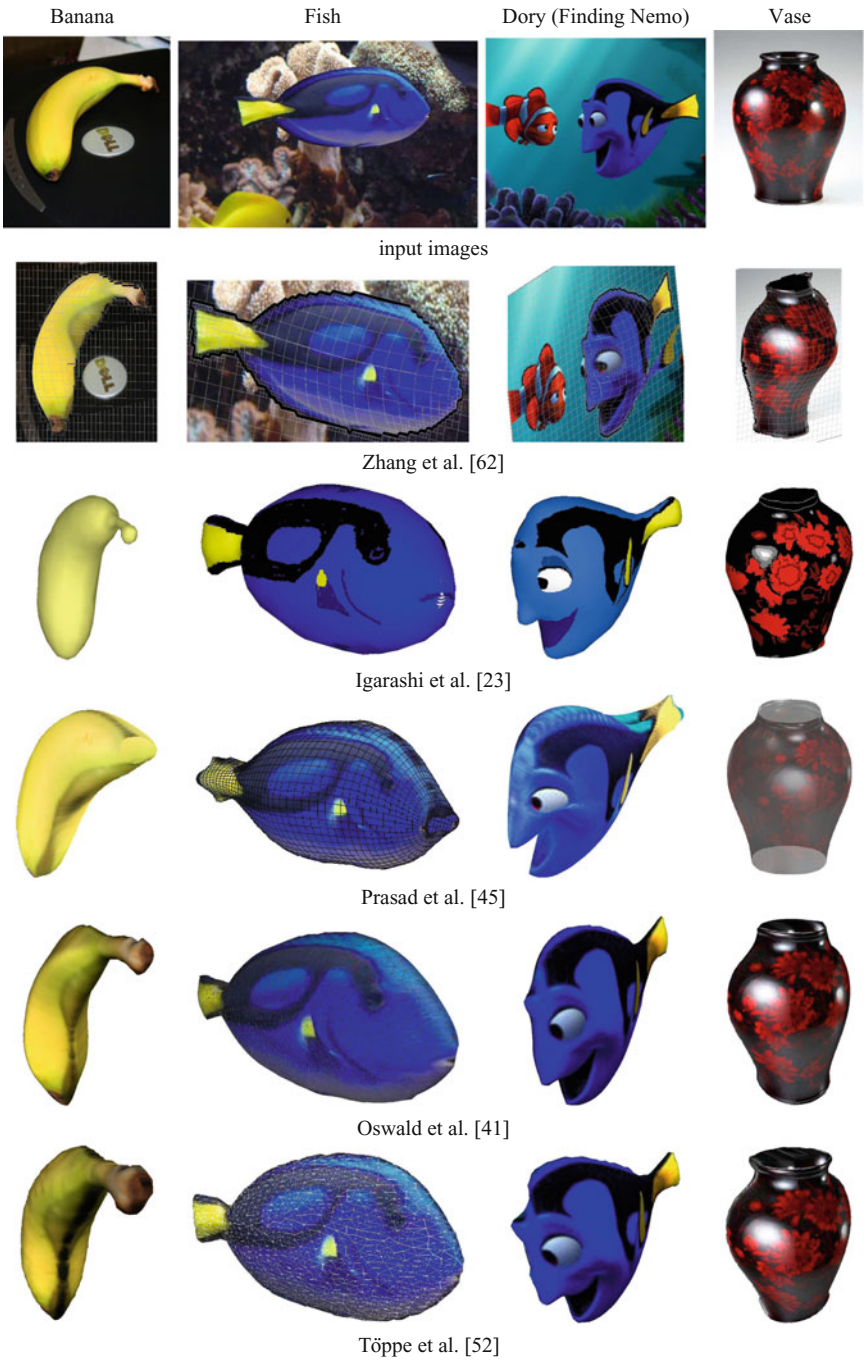

Fig. 16.1 Experimental comparison of several methods for curved object reconstruction (The figures for Prasad et al. are taken from [43])

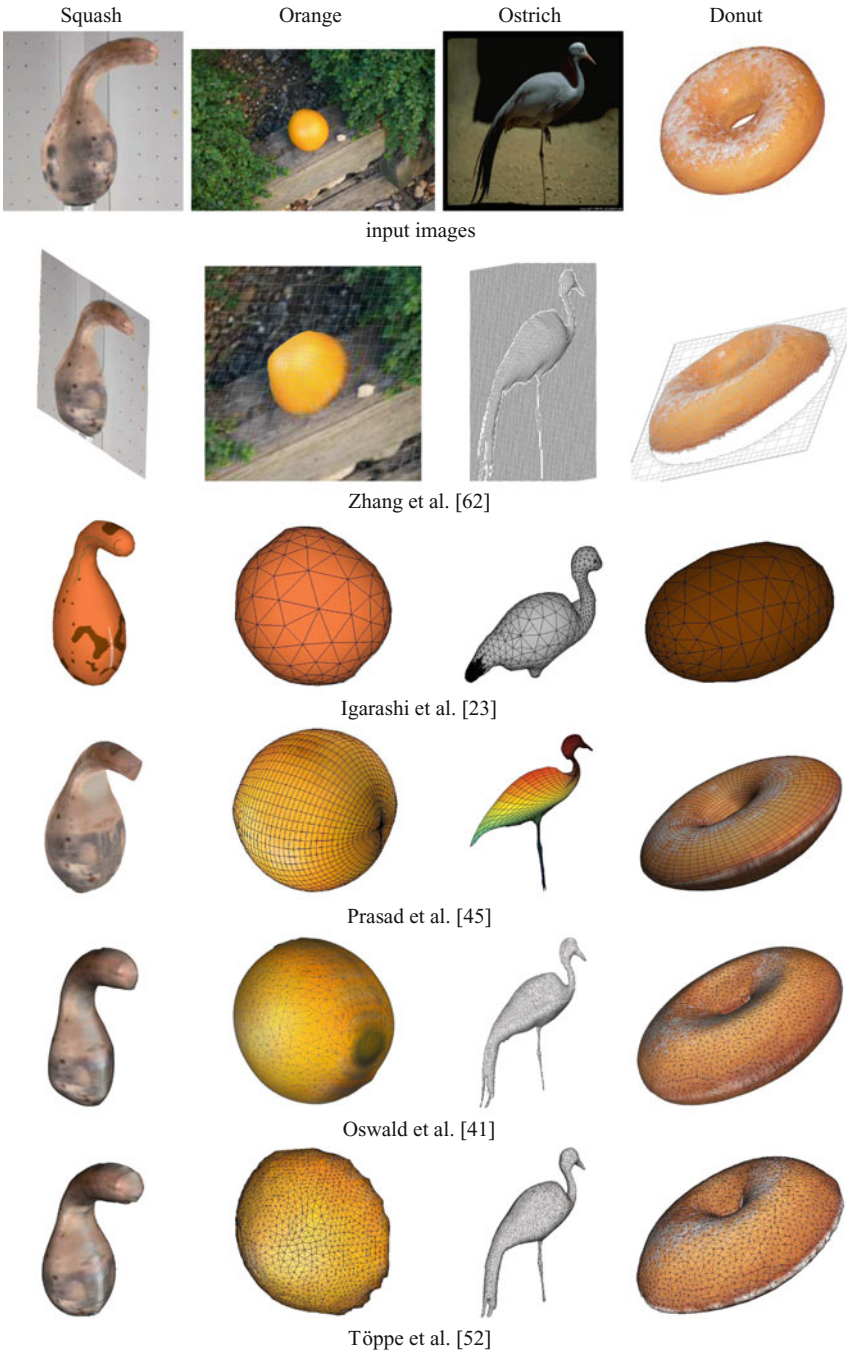

Fig. 16.2 Continuation of Fig. 16.1: experimental comparison of several methods for curved object reconstruction (The figures for Prasad et al. are taken from [43])

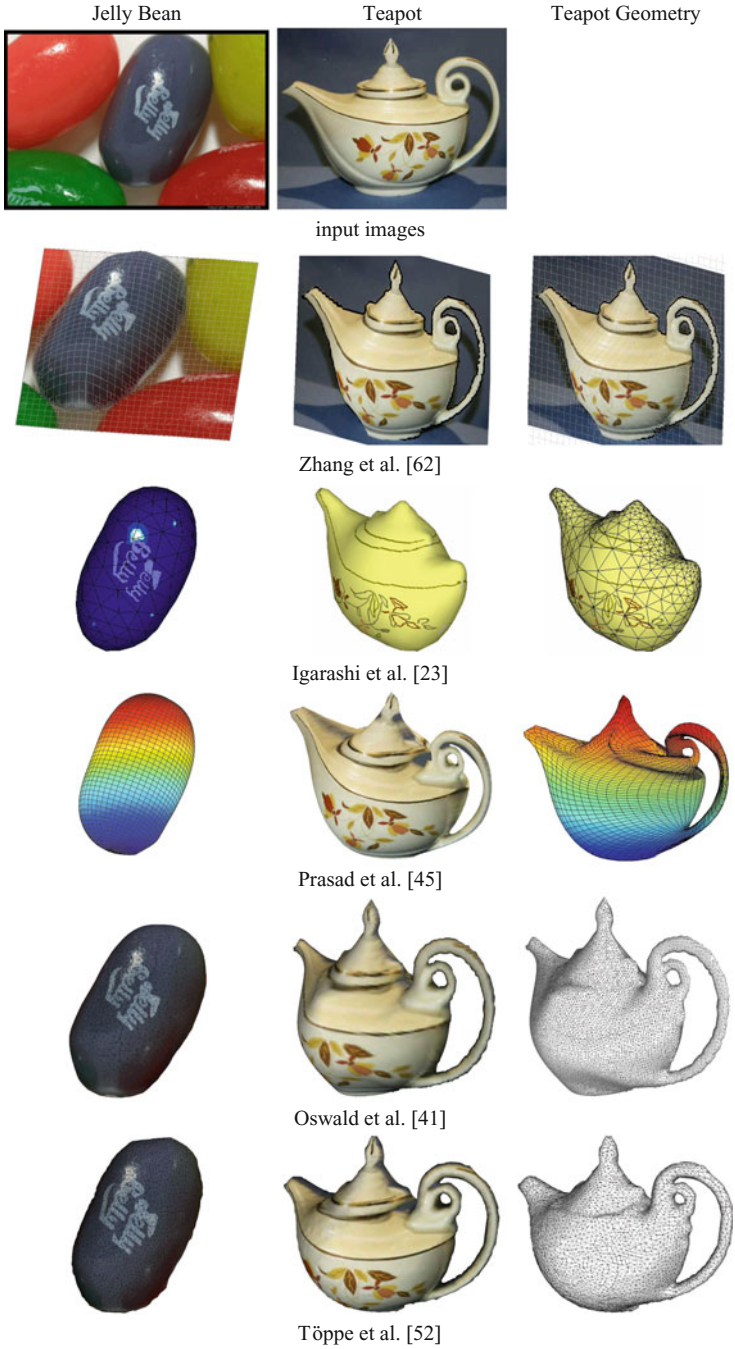

Fig. 16.3 Continuation of Fig. 16.2: experimental comparison of several methods for curved object reconstruction (The figures for Prasad et al. are taken from [43])

Table 16.2 Necessary (**bold**) and optional user inputs and modeling steps for several methods in comparison. Optional user inputs are still required algorithm inputs but they can be predefined by default values or simple heuristics and later on changed by the user if desired. Note that the variety of user input shown in this table does not reflect the amount or complexity of the input that is necessary for a reconstruction

Method	User input (optional and **required**)
Zhang et al. [63]	• Depth map dimensions • Normal/position constraints • Discontinuity lines (position/normal discontinuity) • Planar region constraint • Curvature/torsion minimizing fairing curve constraints • Manual mesh-subdivision
Igarashi et al. [23]	• **Rough contour lines** • Union or cut operations between objects • Object coloring
Prasad et al. [45]	• Mesh resolution • **Silhouette extraction** • **Define corresponding parameter space boundaries** (defines topology) • **Assign parts of the contour to lines in the parameter space** • Choose **inflation heuristic** (cylindrical, cylindrical by parts, distance transform, approximation constraints) + further inflation input • Spillage correction (correct silhouette consistency violated through optimization) • Surface creases
Oswald et al. [41]	• Volume dimensions • **Silhouette extraction** • Define data term shape interactively (4 parameters) • Surface creases
Töppe et al. [52]	• Volume dimensions • **Silhouette extraction** • Define target volume interactively • Surface creases

Table 16.3 Approximate modeling times for all five methods and all examples in Figs. 16.1–16.3 for a medium experienced user. Considering the very similar reconstruction results, this table reveals significant differences for the time of their modeling

Example	Zhang et al. [63]	Igarashi et al. [23]	Prasad et al. [45]	Oswald et al. [41]	Töppe et al. [52]
Banana	20 min	<1 min	10 min	5 min	<1 min
Fish	15 min	<1 min	2 min	8 min	1 min
Dory	40 min	<1 min	5 min	7 min	1 min
Vase	20 min	<1 min	2 min	13 min	4 min
Squash	12 min	<1 min	2 min	2 min	1 min
Orange	14 min	<1 min	<1 min	3 min	<1 min
Ostrich	30 min	<1 min	15 min	7 min	2 min
Donut	55 min	<1 min	10 min	3 min	1 min
Jelly Bean	15 min	<1 min	2 min	4 min	1 min
Teapot	35 min	<1 min	20 min	15 min	4 min

with any graph-cut based method for most of the examples. The reconstructions by Prasad et al. [45] are taken from their respective publications and the authors provided modeling times which include the time taken to annotate contours and all other constraints.

16.5.2.1 Zhang et al.

Evaluation of Experiments This method is more a single view modeling tool rather than a reconstruction tool. Every detail, every extrusion or inflation has to be modeled by the user. Due to the depth map approach it is easy to ensure silhouette consistency when using discontinuity constraints at the silhouette boundary because points are only allowed to change their elevation. These discontinuity constraints have been used for all examples in Figs. 16.1–16.3. The modeling process with this tool can be cumbersome because most constraints only have a local influence on the surface shape. Usually many constraints are necessary for modeling 3D structures from a single image. Further, the oblique position of the donut with respect to the image plane (Fig. 16.2) is difficult to model with local constraints only. The same difficulties emerged with the banana example in Fig. 16.1. Fine scale structures can hardly be modeled and elongated structures such as the handle of the teapot (Fig. 16.3) are often tedious to be modeled properly. The leg of the ostrich (Fig. 16.3) could not be modeled at all due to the limited mesh resolution.

An advantage of this method is the variety of modeling possibilities. The user can fully control each part of the mesh and can even disconnect surface parts. For that reason, the user is able to model details like the round shaped eye of Dory in Fig. 16.1 or the side fin of Dory which bends away from the fish body. Such details cannot be modeled with the other four methods in this comparison. However, the freedom in modeling incurs a larger amount of user input.

User Input Figure 16.4 illustrates the variety of different constraints listed in Table 16.2 and their general effects on the surface shape. None of these constraints is required by the method, but for most reconstructions a reasonable amount of constraints will be necessary. An example input of constraints for the teapot can also be seen in Fig. 16.4. Yellow crosses depict position constraints and the red curve is a position discontinuity constraint.

Although we observed a decrease of user input and modeling time with increasing user experience the amount of modeling time remains comparatively high even for simple reconstructions such as the banana and vase (Fig. 16.1), squash and orange (Fig. 16.2) and jelly bean (Fig. 16.3). The large amount of user input results in higher modeling times which are shown in Table 16.3. The difficulty of modeling a non-side view considerably increased the modeling time for the donut example (Fig. 16.2).

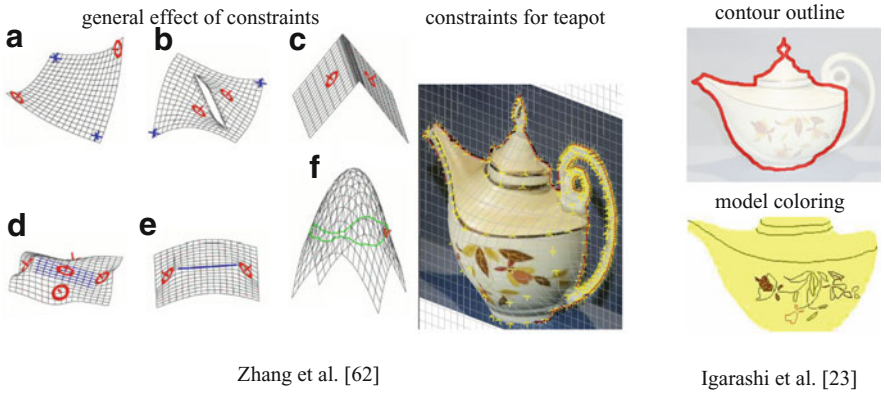

Fig. 16.4 User input for the methods of Zhang et al. [63] and Igarashi et al. [23]. The first image shows general effects of different constraints and is taken from [63]. In particular, the constraints are: (**a**) position (*blue*) and normal (*red*) constraints, (**b**) depth discontinuity constraint, (**c**) crease constraint, (**d**) planar region constraint, (**e**) curvature minimizing fairing curve and (**f**) torsion minimizing fairing curve (see [63] for further details)

16.5.2.2 Igarashi et al.

Evaluation of Experiments This method generally over-smoothes the input silhouette which can be seen in many examples but especially at the peak of the bird in the ostrich example in Fig. 16.2. Although part of the input silhouette the leg of the ostrich is totally missing. Similarly, the grip of teapot lid in Fig. 16.3 is not reconstructed although it is part of the input silhouette which is shown in Fig. 16.4 (top right). Consequently, the reconstructions resulting from this method are not silhouette consistent. Further, the mesh generation in this approach is not straightforward and may lead to mesh inconsistencies which can be observed in the reconstruction of the ostrich example in the lower neck area (Fig. 16.2).

The main advantage of this approach is the fast and intuitive modeling of geometrically simple objects. One of the drawbacks is the restricted topology, the hole in the donut example in Fig. 16.2 cannot be reconstructed. A further disadvantage is the limited influence of the user during the modeling process. The fact that surface discontinuities like sharp edges are not allowed largely decreases the class of reconstructable objects. For instance, the tail fin of Dory in Fig. 16.1 cannot be modeled to end with a sharp edge. The same holds for the top and bottom parts of the vase example in the same figure and for the bottom part of the teapot in Fig. 16.3. Only very simple roundish objects like the banana (Fig. 16.1), squash and orange (Fig. 16.2) or the jelly bean (Fig. 16.3) can be easily and reliably reconstructed.

User Input None of the user input in Table 16.2 needs much experience or even expert knowledge: From a given closed contour line the method instantly inflates a 3D mesh. For better user experience the user interface of the tool provides simple

paint operations to color the model. See Fig. 16.4 for the exemplary user input of the teapot example. In all experiments this method needed the least user input (cf. Table 16.3) at the price of producing the least accurate reconstructions with respect to the given input silhouette (see Figs. 16.1–16.3).

16.5.2.3 Prasad et al.

Evaluation of Experiments The main benefit of this method stems from a few extensions over the method by Zhang et al. [63] which facilitate the object inflation and allow the reconstruction of closed surfaces. Apart from the silhouette outline the user can assign any contour line to a line in the parameter space. For instance, this has been used to relate a contour edge of the banana (Fig. 16.1) to a crease in the reconstructed geometry (see [45] for further details).

For closed surfaces a correspondence of parameter space boundaries has to be defined. One the one hand, this facilitates the reconstruction of objects with a simple geometry such as the orange example in Fig. 16.2 or the jelly bean in Fig. 16.3. The orange and the jelly bean get naturally inflated only be defining the silhouette contour and the correspondence of parameter space boundaries. On the other hand, the definition of the parametrization can be sophisticated for objects of higher genus such as in the teapot example (Fig. 16.3) which has genus 2. However, not only the object topology limits the applicability of this method: it is also difficult to define a proper parametrization for objects which have many elongated structures such as legs, arms, tails or spikes. A limiting example may be the ostrich in Fig. 16.2 which is parametrized as a cylinder starting from the peak, going through the neck down to the leg. The body of the ostrich is modeled as a 'bump' in this cylinder which already leads to very sparsely sampled regions at the back of the ostrich. A further modeling of the tail feathers is done with the other methods would need an enormous increase of the mesh resolution in the tail part of the ostrich in order to have enough quads that can align to the silhouette. Thus, finding a good parametrization is a demanding task despite a simple object genus of 0 as for example when modeling a cow with 4 legs and a tail. Moreover, an undesired effect of such a parametrization is the non-uniform distribution of points in the parameter space on the surface as shown in the ostrich (Fig. 16.2) and teapot (Fig. 16.3) examples. Together with further constraints this may lead to non-controllable surface oscillations of the object surface. This can be observed in the top part of the teapot which is connected to the teapot handle (Fig. 16.3).

Another disadvantage is the fact that silhouette consistency is only incorporated with boundary conditions which may get violated during the optimization process. The user may need to add further constraints to ensure a silhouette consistent reconstruction. This process is called spillage correction in Table 16.2 and Fig. 16.5f. Nonetheless, this method generated the most accurate results for the non-side-view examples (banana and teapot).

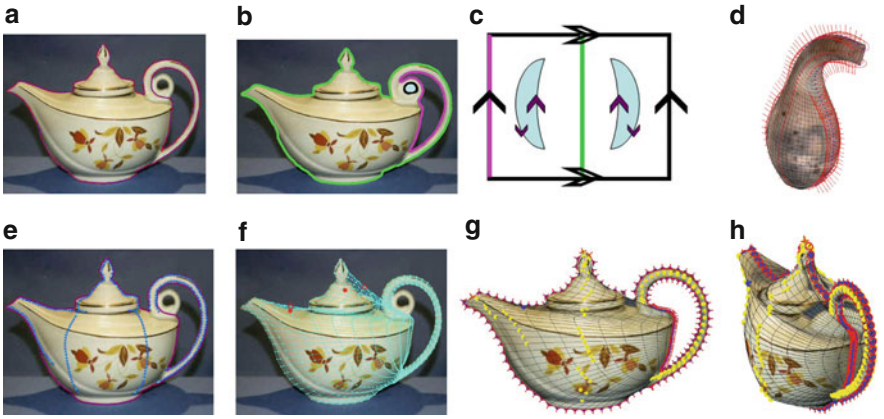

Fig. 16.5 Necessary and optional steps and user input for Prasad et al. [45]: (**a**) contour extraction; (**b**) lines of the contour have to be related to lines in the parameter space (**c**); (**d**) and (**e**) demonstrate different inflation heuristics; (**f**) during the optional spillage correction, the user can account for silhouette inconsistencies by adding further constraints; (**g**) and (**h**) show the final model and generated interpolation constraints as *yellow dots*. Note that (**b**) and (**c**) show a genus 2 reconstruction, while the other teapot images show a genus 1 reconstruction (All figures are taken from [43])

User Input Most of the user input listed in Table 16.2 is illustrated in Fig. 16.5. After the *contour extraction* (Fig. 16.5a) normal constraints are generated along the contour shown as red discs with needles in Fig. 16.5d, g, and h. The definition of corresponding parameter space boundaries and the assignment of contour line parts to lines in the parameter space is shown in Fig. 16.5b and c.

Figure 16.5d shows generated constraints as red dots for a *cylindrical inflation*. Objects with cylindrical shape about a virtual free-form 3D curve or spine can be inflated by generating interpolation constraints along the spine with a depth value equal to the minimal distance between the point of the spine and the contour line. This inflation heuristic is generalized for more complex objects as object parts can be independently inflated with the same technique. To this end the user defines pairs of *inflation curves* (Fig. 16.5e) for which interpolation constraints are generated along the (virtual) medial spine (Fig. 16.5g and h).

The necessity and complexity of each single user input step depends on the object to be reconstructed leading to very different modeling times for the presented experiments (see Table 16.3).

16.5.2.4 Oswald et al.

Evaluation of Experiments Single view modeling with this method is mostly intuitive and many examples did not need much effort. For instance, the banana, fish, dory (Fig. 16.1), squash, orange, ostrich and donut examples (Fig. 16.2) or the

Fig. 16.6 Both methods, Oswald et al. [41] and Töppe et al. [52] have a similar workflow: a silhouette can be estimated with a few user scribbles (*red* = foreground, *blue* = background). Optionally, the user can add creases (*green line*) to locally change the smoothness properties of the final surface. The final geometry shows the discontinuity around the lid and is a result of the method by Töppe et al.

jelly bean (Fig. 16.3) example are easy to accomplish, especially in comparison to the method by Zhang et al. [63]. The shape of the data term mainly defines the shape of the final reconstruction and a proper modeling needs a little user experience. However, the edged shape of the data term which is governed by the silhouette distance function is sometimes hard to conceal. This can be observed in the squash (Fig. 16.2) and the teapot (Fig. 16.3) examples in comparison to Töppe et al. [52]. This characteristic shape, however, can also be beneficial as shown in the banana (Fig. 16.1) and ostrich examples (Fig. 16.2). In contrast to the methods discussed previously, this one assumes objects to be symmetric to the image plane. That is, it assumes a side view of the object, which in turn restricts the applicability of this method to a smaller class of objects that can be reconstructed. This can be observed in the donut example in Fig. 16.2. The input image is not a side view and the resulting reconstruction underestimates the size of the hole.

User Input Although the silhouette extraction is the only necessary user input an adaption of the data term shape by changing parameters $k, \lambda_{\text{offset}}, \lambda_{\text{factor}}, \lambda_{\text{cutoff}}$ from Eq. (16.7) is necessary in most cases. The effect of optional creases to locally relax the smoothness regularization and the general workflow of this method is depicted in Fig. 16.6.

16.5.2.5 Töppe et al.

Evaluation of Experiments Reconstructions with this method are easily obtained with even less user input in comparison to Oswald et al. [41], that is, by an adaption of a single parameter: the object volume. Only the teapot (Fig. 16.3) and vase (Fig. 16.1) examples needed additional user input for surface creases. Similar to Oswald et al. [41] this method assumes symmetry of objects and requires a representative silhouette from a side view. Therefore, the hole of the donut (Fig. 16.2) is underestimated in the same way. For the banana example (Fig. 16.1), the method still yields good results because the oblique object view did not change much on the characteristic silhouette properties.

Table 16.4 Overview of advantages and disadvantages for each method. Note that the number of advantages and disadvantages is not important in this listing since each point weights differently depending on the desired application of each method

Method	Advantages (+) and disadvantages (−)
Zhang et al. [63]	+ Large variety of constraints allows for flexible user modeling + User has full control of every surface detail − Reconstructions are restricted to a depth map − Occluded object parts cannot be modeled, synthesized views from different angles will reveal those areas − Large amount of user input is often necessary
Igarashi et al. [23]	+ Very easy to use and fast interactive modeling − Over-smoothes the input silhouette − Smoothness properties cannot be changed by the user − Not silhouette consistent − Topology is limited to genus 0
Prasad et al. [45]	+ Objects can also be modeled from oblique view points + Apart from the silhouette the user can also use contour edges for modeling − Parametric surface representation limits topology and object shape (many long elongated structures are difficult to model) − Higher complexity of user input (requires expert knowledge) − Silhouette consistency is not guaranteed and may require additional user input
Oswald et al. [41]	+ Moderately fast modeling + Reconstructions are silhouette consistent − Objects need to be symmetric, a side view is required
Töppe et al. [52]	+ Fast modeling + Very little user input + Reconstructions are silhouette consistent − Objects need to be symmetric, a side view is required − User can barely influence the surface shape − Limited possibilities to add surface creases

User Input Similar to Oswald et al. [41] the object silhouette is the only necessary user input (see also Fig. 16.6). Apart from the silhouette the object shape is mainly defined by the adapting the object volume. Also similar to Oswald et al. [41] surface creases can be added to create non-smooth surface parts. However, changing the smoothness locally may behave differently at different locations within the object silhouette, because there is no data term that governs the object shape in non-smooth surface areas. In these areas the surface shape is more defined by the distribution of volume which minimizes the overall surface area for given smoothness properties.

16.5.2.6 Summary

The theoretical and experimental comparison of the five methods for curved object reconstruction identified several advantages and disadvantages of the presented approaches which are listed in Table 16.4. In general, the performance of a method

highly depends on the application. Each method has its strengths and weaknesses when applied to a specific class of single view reconstruction problems.

The results of our experiments support the hypothesis that *generality and flexibility* of a reconstruction method is traded for the *amount of user input* or *expert knowledge*. Expert knowledge refers to the *variety* or *complexity* of the user input. The flexibility of modeling fine details with the method by Zhang et al. [63] either requires the user to know and understand a variety of modeling constraints or it needs a large amount of user input. On the other hand, the method by Prasad et al. [45] needs less user input, but increases its complexity such as the definition of a suitable surface parametrization. The comparatively simple and small amount of user input for the methods by Igarashi et al. [23], Oswald et al. [41], and Töppe et al. [52] comes along with the limited generality and flexibility of these methods.

16.6 Conclusion

Single view reconstruction approaches infer the structure of 3D objects from 2D images. In this chapter, we discussed a representative set of existing algorithms and grouped them into four classes: curved objects, piecewise planar objects, learning specific objects and 3D impression from scenes. These groupings are based on the different reconstruction objectives of the algorithms. We identified several important categories for comparing the approaches: the reconstruction precision, the 3D object representation, the assumptions made by the algorithm, the required user interaction, as well as image cues and priors used by the algorithms. Moreover, we have concentrated on five methods aiming for curved object reconstruction and provided an extensive experimental comparison of established and recent methods.

Acknowledgements We thank Li Zhang and Steve Seitz for sharing their data and providing their single view modeling tool. Further, we are grateful to Mukta Prasad and co-authors for providing test data and results of their single view reconstruction method.

References

1. Bae, S., Durand, F.: Defocus magnification. In: Proceedings of the Eurographics. Blackwell, Oxford (2007)
2. Barinova, O., Konushin, V., Yakubenko, A., Lee, K., Lim, H., Konushin, A.: Fast automatic single-view 3-d reconstruction of urban scenes. In: Proceedings of the European Conference on Computer Vision, pp. 100–113. Springer, Berlin/Heidelberg (2008)
3. Botsch, M., Kobbelt, L.: An intuitive framework for real-time freeform modeling. In: ACM Transactions on Graphics (Proc. SIGGRAPH), vol. 23, pp. 630–634. ACM/Addison-Wesley Publishing Co., New York (2004)
4. Chen, Y., Cipolla, R.: Single and sparse view 3d reconstruction by learning shape priors. Comput. Vis. Image Underst. **115**, 586–602 (2011)

5. Cohen, L.D., Cohen, I.: Finite-element methods for active contour models and balloons for 2-d and 3-d images. IEEE Trans. Pattern Anal. Mach. Intell. **15**(11), 1131–1147 (1993)
6. Colombo, C., Del Bimbo, A., Pernici, F.: Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 99–114 (2005)
7. Courteille, F., Crouzil, A., Durou, J.-D., Gurdjos, P.: Towards shape from shading under realistic photographic conditions. pp. 277–280 (2004)
8. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Comput. Vis. **40**(2), 123–148 (2000)
9. Daum, M., Dudek, G.: On 3-d surface reconstruction using shape from shadows. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, USA, pp. 461–468 (1998)
10. Delage, E., Lee, H., Ng, A.Y.: Automatic single-image 3d reconstructions of indoor Manhattan world scenes (2005)
11. Duchon, J.: Splines Minimizing Rotation-Invariant Semi-norms in Sobolev Spaces. In: Schempp, W., Zeller, K. (eds.) Constructive Theory of Functions of Several Variables, vol. 571/1977, pp. 85–100. Springer, Berlin (1977)
12. Durou, J.-D., Falcone, M., Sagona, M.: Numerical methods for shape-from-shading: a new survey with benchmarks. Comput. Vis. Image Underst. **109**, 22–43 (2008)
13. Francois, A.R.J., Medioni, G.G.: Interactive 3d model extraction from a single image. Image Vis. Comput. **19**(6), 317–328 (2001)
14. Han, F., Zhu, S.-C.: Bayesian reconstruction of 3d shapes and scenes from a single image. In: Proceedings of the Higher-Level Knowledge in 3D Modeling and Motion Analysis, Nice, France, pp. 12–20. (2003)
15. Hassner, T., Basri, R.: Example based 3d reconstruction from single 2d images. In: Beyond Patches Workshop at IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, p. 15. (2006)
16. Hatzitheodorou, M.: The derivation of 3-d surface shape from shadows. In: Proceedings of a Workshop on Image Understanding Workshop, pp. 1012–1020. Morgan Kaufmann, San Francisco (1989)
17. Hoiem, D., Efros, A.A., Hebert, M.: Automatic photo pop-up. ACM Trans. Graph. **24**(3), 577–584 (2005)
18. Hong, W., Yang, A.-Y., Huang, K., Ma, Y.: On symmetry and multiple-view geometry: structure, pose, and calibration from a single image. Int. J. Comput. Vis. **60**, 241–265 (2004). doi:10.1023/B:VISI.0000036837.76476.10
19. Horaud, R.P., Brady, M.: On the geometric interpretation of image contours. Artif. Intell. **37**(1–3), 333–353 (1988). Special Issue on Geometric Reasoning
20. Horry, Y., Anjyo, K.-I., Arai, K.: Tour into the picture: using a spidery mesh interface to make animation from a single image. In: SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 225–232. ACM/Addison-Wesley, New York (1997)
21. Igarashi, T.: Adaptive unwrapping for interactive texture painting. In: Proceedings of the ACM Symposium on Interactive 3D Graphics, pp. 209–216. ACM, New York (2001)
22. Igarashi, T.: Smooth meshes for sketch-based freeform modeling. In: Proceedings of the 2003 Symposium on Interactive 3D Graphics, pp. 139–142. ACM, New York (2003)
23. Igarashi, T., Matsuoka, S., Tanaka, H.: Teddy: a sketching interface for 3d freeform design. In: ACM Transactions on Graphics (Proceedings of the SIGGRAPH), pp. 409–416. ACM/Addison-Wesley, New York (1999)
24. Ikeuchi, K., Horn, B.K.P.: Numerical shape from shading and occluding boundaries. Artif. Intell. **17**, 141–185 (1981)
25. Joshi, P., Carr, N.: Repoussé: automatic inflation of 2d art. In: Proceedings of the Eurographics Workshop on Sketch-Based Modeling, Annecy, France (2008)
26. Kanade, T.: Recovery of the three-dimensional shape of an object from a single view. Artif. Intell. **17**, 409–460 (1981)

27. Karpenko, O.A., Hughes, J.F.: Smoothsketch: 3d free-form shapes from complex sketches. ACM Trans. Graph. **25**(3), 589–598 (2006)
28. Karpenko, O., Hughes, J.F., Raskar, R.: Free-form sketching with variational implicit surfaces. Comput. Graph. Forum **21**(3), 585–594 (2002)
29. Kender, J., Smith, E.: Shape from darkness. In: Proceedings of the International Conference on Computer Vision, London, pp. 539–546 (1987)
30. Koutsourakis, P., Simon, L., Teboul, O., Tziritas, G., Paragios, N.: Single view reconstruction using shape grammars for urban environments. In: Proceedings of the International Conference on Computer Vision, Kyoto. IEEE (2009)
31. Kushal, A.M., Sanyal, S., Bansal, V., Banerjee, S.: A simple method for interactive 3d reconstruction and camera calibration from a single view. In: Proceedings Indian Conference in Computer Vision, Graphics and Image Processing, Ahmedabad, India (2002)
32. Levin, A.: Analyzing depth from coded aperture sets. In: Proceedings of the European Conference on Computer Vision, pp. 214–227. Springer, Berlin (2010)
33. Li, Y., Pizlo, Z., Steinman, R.M.: A computational model that recovers the 3d shape of an object from a single 2d retinal representation. Vis. Res. **49**, 979–991 (2009)
34. Li, Z., Liu, J., Tang, X.: A closed-form solution to 3d reconstruction of piecewise planar objects from single images. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, pp. 1–6 (2007)
35. Liebowitz, D., Criminisi, A., Zisserman, A.: Creating architectural models from images. In: Proceedings of the EuroGraphics, vol. 18, pp. 39–50 (1999)
36. Lowe, D.G.: Three-dimensional object recognition from single two-dimensional images. Artif. Intell. **31**, 355–395 (1987)
37. Malik, J., Rosenholtz, R.: Computing local surface orientation and shape from texture for curved surfaces. Int. J. Comput. Vis. **23**(2), 149–168 (1997)
38. Nagai, T., Naruse, T., Ikehara, M., Kurematsu, A.: Hmm-based surface reconstruction from single images. In: Proceedings of the International Conference on Image Processing, Rochester, NY, USA, vol. 2, pp. II–561–II–564 (2002)
39. Nayar, S.K., Ikeuchi, K., Kanade, T.: Shape from interreflections. In: Proceedings of the International Conference on Computer Vision, Osaka, Japan, pp. 2–11 (1990)
40. Nealen, A., Igarashi, T., Sorkine, O., Alexa, M.: Fibermesh: designing freeform surfaces with 3d curves. ACM Trans. Graph. **26**(3), 41 (2007)
41. Oswald, M.R., Töppe, E., Kolev, K., Cremers, D.: Non-parametric single view reconstruction of curved objects using convex optimization. In: Pattern Recognition (Proceedings of the DAGM), Jena (2009)
42. Prados, E., Faugeras, O.: Shape from shading: a well-posed problem? In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 870–877. IEEE, San Diego/Etats-Unis (2005)
43. Prasad, M.: Class-based single view reconstruction. Ph.D. thesis, University of Oxford (2009)
44. Prasad, M., Zisserman, A., Fitzgibbon, A.W.: Fast and controllable 3D modelling from silhouettes. In: Eurographics, Short Papers, pp. 9–12 (2005)
45. Prasad, M., Zisserman, A., Fitzgibbon, A.W.: Single view reconstruction of curved surfaces. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, New York, pp. 1345–1354 (2006)
46. Rother, D., Sapiro, G.: Seeing 3d objects in a single 2d image. In: Proceedings of the International Conference on Computer Vision, Kyoto, pp. 1819–1826 (2009)
47. Saxena, A., Chung, S.H., Ng, A.Y.: 3-d depth reconstruction from a single still image. Int. J. Comput. Vis. **76**, 53–69 (2007)
48. Saxena, A., Sun, M., Ng, A.Y.: Make3d: learning 3d scene structure from a single still image. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 824–840 (2009)
49. Sturm, P.F., Maybank, S.J.: A method for interactive 3d reconstruction of piecewise planar objects from single images. In: Proceedings of the BMVC, pp. 265–274. British Machine Vision Association, England (1999)

50. Super, B.J., Bovik, A.C.: Shape from texture using local spectral moments. IEEE Trans. Pattern Anal. Mach. Intell. **17**, 333–343 (1995)
51. Terzopoulos, D., Witkin, A., Kass, M.: Symmetry-seeking models and 3d object reconstruction. Int. J. Comput. Vis. **1**, 211–221 (1987)
52. Töppe, E., Oswald, M.R., Cremers, D., Rother, C.: Image-based 3d modeling via cheeger sets. In: Proceedings of the Asian Conference on Computer Vision, Queenstown (2010)
53. Ulupinar, F., Nevatia, R.: Shape from contour: straight homogeneous generalized cylinders and constant cross section generalized cylinders. IEEE Trans. Pattern Anal. Mach. Intell. **17**, 120–135 (1995)
54. Utcke, S., Zisserman, A.: Projective reconstruction of surfaces of revolution. In: Pattern Recognition (Proceedings of the DAGM), pp. 93–102. Springer, Berlin/New York (2003)
55. Vetter, T.: Synthesis of novel views from a single face image. Int. J. Comput. Vis. **28**, 103 (1998)
56. Wang, G., Su, W., Song, Y.: A new shape from shading approach for specular surfaces. In: Proceedings of the Third International Conference on Artificial Intelligence and Computational Intelligence – Volume Part III. Lecture Notes in Computer Science, pp. 71–78. Springer, Berlin/Heidelberg (2011)
57. Welch, W., Witkin, A.P.: Free-form shape design using triangulated surfaces. In: ACM Transactions on Graphics (Proceedings of the SIGGRAPH), pp. 247–256 (1994)
58. Williams, L.: 3d paint. In: Proceedings of the 1990 Symposium on Interactive 3D Graphics, Snowbird, pp. 225–233 (1990)
59. Wong, K.-Y.K., Mendonça, P.R.S., Cipolla, R.: Reconstruction of surfaces of revolution from single uncalibrated views. In: Proceedings of the British Machine Vision Conference, Cardiff, pp. 265–272 (2002)
60. Yu, Y., Chang, J.: Shadow graphs and surface reconstruction. In: Proceedings of the European Conference on Computer Vision, Copenhagen, pp. 31–45 (2002)
61. Zeleznik, R.C., Herndon, K.P., Hughes, J.F.: Sketch: an interface for sketching 3d scenes. In: ACM Transactions on Graphics (Proceedings of the SIGGRAPH), pp. 163–170. ACM, New York (1996)
62. Zhang, R., Tsai, P.-S., Cryer, J.E., Shah, M.: Shape from shading: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **21**(8), 690–706 (1999)
63. Zhang, L., Dugas-Phocion, G., Samson, J.-S., Seitz, S.M.: Single view modeling of free-form scenes. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, Kauai, HI USA, pp. 990–997 (2001)