Segment Based 3D Object Shape Priors

Rabeeh Karimi Mahabadi, Christian Häne, Marc Pollefeys
Department of Computer Science
ETH Zürich, Switzerland
rabeehk@student.ethz.ch, {christian.haene, marc.pollefeys}@inf.ethz.ch

Abstract

Dense 3D reconstruction still remains a hard task for a broad number of object classes which are not sufficiently textured or contain transparent and reflective parts. Shape priors are the tool of choice when the input data itself is not descriptive enough to get a faithful reconstruction. We propose a novel shape prior formulation that splits the object into multiple convex parts. The reconstruction problem is posed as a volumetric multi-label segmentation. Each of the transitions between labels is penalized with its individual anisotropic smoothness term. This powerful formulation allows us to represent a descriptive shape prior. For the object classes used in this paper the individual segments naturally correspond to different semantic parts of the object. This leads to a semantic segmentation as a side product of our shape prior formulation. We evaluate our method on several challenging real-world datasets. Our results show that we can resolve issues such as undesired holes and disconnected parts. Taking into account a segmentation of the free space, we show that we are able to reconstruct concavities, such as the interior of a mug.

1. Introduction

Despite the continuous advances in dense 3D surface reconstruction there are still many object classes which are a challenge for current algorithms. To tackle such classes shape priors have been proposed. One approach to shape priors is anisotropic surface regularization based on prior knowledge about the shape. Convex multi-label segmentation has been used in dense volumetric reconstruction to jointly infer semantic labels and geometry [10] and to define per object shape priors for challenging classes [9]. The first work uses priors on the directions of transitions between semantic classes, such as a preference of a ground ↔ building transition to be horizontal. This leads to a prior on the shape of semantic classes, which can be weak for general shapes, as we will see later. The latter work defines a shape prior for a given object class by having a spatially varying
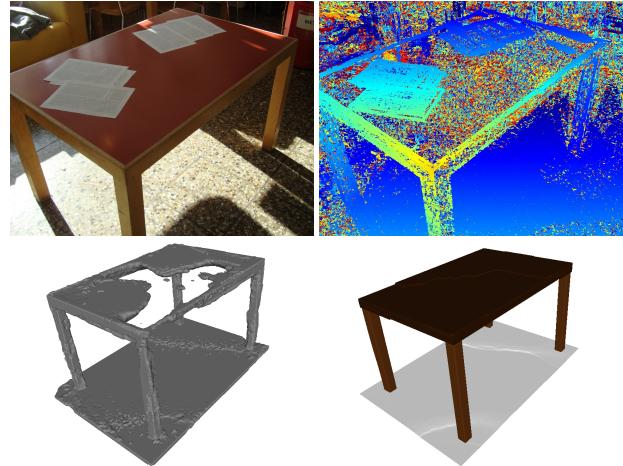

Figure 1. Top: Example input image and depth map. Bottom: Standard volumetric fusion result (left) and our result using the proposed segment based shape prior (right).

anisotropic regularization. This leads to a very descriptive prior, for example a tabletop is always horizontal, but has the drawback that the object needs to be exactly aligned with the bounding box. In this work, we propose an alternative approach to get strong priors, namely splitting the object into multiple simpler parts which we call segments.

To motivate our work we describe the example of a table and observe that a prior based on a single spatially homogeneous anisotropic regularization does not lead to a descriptive shape prior. First, we observe that the main surface area on the top is horizontal and also there is large surface area on the legs which is predominantly vertical. Hence a prior on the surface orientation of a table should penalize those mostly observed directions less than others. In terms of dense volumetric 3D reconstruction the main difficulties in reconstructing a table are the thin leg structures that easily get disconnected and holes appearing in the often texture-less top surfaces. The single anisotropic prior would not help in either of these cases. It would penalize holes in the top and disconnected legs less and therefore make them
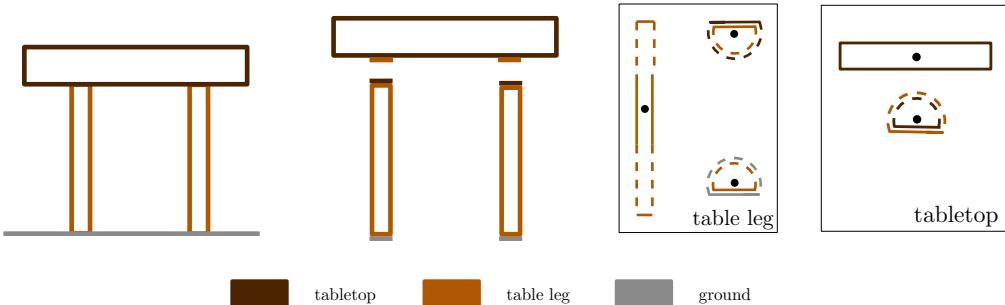
Figure 2. Shape prior formulation for the table: The table gets segmented into convex parts (tabletop, and table legs), the transitions of each part to other labels are taken into account, for each transition the smoothness is defined in terms of Wulff shapes. The dashed lines indicate that the shape is very large in the indicated direction, meaning that this is an unlikely interface in the final reconstruction.

more likely, leading to a very weak shape prior. Our proposed solution is, splitting the object into a top part and legs. Now we can define three different smoothness terms for the surface between the top and the legs, for the top and for the legs. Now each of the surfaces has a strong predominant direction, the top is mostly horizontal, the legs are mostly vertical and the transition between the legs and the top is strictly horizontal. Note that such a prior does not need an exact location of the object, only the main directions need to be known.

Often the segments of the object coincide naturally with the semantic parts of an object. As we will see later, in our formulation we do not make use of any semantic information. But due to the geometric difference of semantic parts of an object we also naturally get a semantic segmentation. Including semantic classifiers to our method would be straight forward using [10] but to underline the strength of our prior, which is able to segment the object solely based on distinctive geometry, we refrained from using any semantic image classifier.

The idea of our shape prior formulation is coming from the study of equilibrium shapes of crystals [23]. We will explain that anisotropic surface regularization can be seen as preferring object shapes which follow the same shape as a convex example shape, named Wulff shape. The Wulff shape is exactly the equilibrium shape of a crystal. Due to the convexity of the Wulff shapes it becomes natural that our input object shapes are split into convex or almost convex segments. An illustration of this idea is given in Fig. 2.

1.1. Related Work

Dense 3D reconstruction from images is a well studied topic in computer vision, we refer the reader to [21] for an overview. Here, we mention only the most related works to our approach. Dense volumetric 3D reconstruction was originally proposed in [4]. The idea is to segment a volume into free space and occupied space and extract the surface as the interface between them. The original approach did not contain regularization. Due to the high quality of the input laser scans this was not necessary. However, in computer vision the data is often contaminated by noise and (strong) regularization is necessary. Traditionally, a prior that penalizes the surface area is utilized, it has been formulated in both the discrete graph-cut based [14] optimization and the continuous convex optimization literature [25]. Anisotropic regularization has been included to improve surface details [13]. In this work, the surface is aligned with an input normal field that is obtained from semi-dense multi view stereo.

Volumetric 3D reconstruction is not limited to the two label case that separates free space and occupied space. Assigning semantic classes to voxels leads to a multi-label formulation that allows for joint reconstruction and semantic segmentation [10]. In this formulation, each of the transitions between two semantic classes is penalized differently with an anisotropic smoothness term. Having a smoothness term which prefers for example ground to be horizontal introduces a shape prior on the semantic classes. This idea has been extended to 3D object shape priors in [9], where a spatially varying anisotropic smoothness term is derived from training data and acts as a shape prior during reconstruction. Using a spatially varying smoothness term leads to very descriptive priors but needs an alignment between the prior and the reconstruction. We argue that we can stick to a spatially homogeneous anisotropic smoothness prior but split the object into multiple convex segments instead, to make the prior descriptive.

Segmenting objects into convex segments is a well-studied problem in computer vision, computational geometry, computer graphics and related fields [17, 3, 15, 16, 18]. The aims for splitting objects into simpler parts in these areas are, for example, to handle tasks such as collision avoidance, simplify manufacturing of objects with 3D printers and similar devices, or reduce a problem to an algorithm which works on convex shapes only.

The idea of splitting objects into parts has also been used for object detectors in images [6]. Different parts of objects

such as hands, head and torso of a body have a different appearance and treating them as individual parts improves the detection quality. For semantic segmentation of images, structured representations have been proposed [7, 22, 5].

In our work we argue that by segmenting an object into multiple convex parts we can derive more descriptive priors without the need of a spatial alignment between the prior and the reconstruction.

2. Formulation

In this section we will present the mathematical formulation of our segment based object shape prior. The traditional dense volumetric 3D reconstruction segments a voxel space into free space and occupied space [4] and extracts the surface as a transition between the free and occupied space. We are using a recently introduced framework which extends these ideas to dense volumetric multi-label segmentation [10]. The formulation originates from continuous multi-label segmentation [2] and has been extended to handle anisotropy and non-metric smoothness, at the same time, in its discretized form [26]. This becomes important for some of our examples later on. In the following we will first describe the convex multi-label segmentation formulation which we are using and then we will explain how we apply this formulation to our proposed segment based 3D object shape prior.

2.1. Convex multi-label segmentation

We directly explain the final convex multi-label formulation in its discretized form. For more information on how it is derived, we refer the reader to [2, 26].

For our purposes, the goal is to assign labels to a volumetric domain. We denote the discretized domain by $\Omega \subset \mathbb{R}^3$ and index the voxels by a position index s. $\mathcal{L} = \{0, \dots, L-1\}$ is a set of labels, where each of the labels corresponds to one of potentially multiple free space labels or one of the occupied space labels. Its meaning is object class specific and will be explained in Section 3. To formalize the label assignment task label indicator variables $x_s^i \in [0, 1]$ are introduced, where $x_s^i = 1$ if label i is assigned to voxel s and $x_s^i = 0$, otherwise. Next we state the convex energy and will explain its interpretation afterwards.

$$E(x) = \sum_{s \in \Omega} \left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j} \phi^{ij}(x_s^{ij} - x_s^{ji}) \right)$$

$$\text{subject to} \quad x_s^i = \sum_j (x_s^{ij})_k, \quad x_s^i = \sum_j (x_{s-e_k}^{ji})_k \quad (1)$$

$$x_s^i \geq 0, \quad \sum_i x_s^i = 1, \quad x_s^{ij} \geq 0$$

The variables $x_s^{ij} \in \mathbb{R}^3$ are used to describe transition gradients of the label indicator functions. The variables are only allowed to be non-negative and hence cannot describe full gradients, but by taking the difference $y_s^{ij} := x_s^{ij} - x_s^{ji}$, the length of the vectors y_s^{ij} describe the amount of change from label x_s^i to label x_s^j in the direction of y_s^{ij}. The functions $\phi^{ij} : \mathbb{R}^3 \to \mathbb{R}^+$ are convex positively 1-homogeneous functions that act as an anisotropic regularizer of the surface area [19]. The $\rho_s^i \in \mathbb{R}$ are the unary data costs. They describe the local preference for label i in voxel s. The index k describes the dimension and e_k is the k-th cannonical basis vector. A first set of constraints, which are called marginalization constraints make sure that the variables x_s^{ij} agree with the label indicator variables x_s^i. Finally, the second line of constraints describes normalization constraints on x_s^i which ensure that a label gets selected in each voxel and the non-negativity constraints on the x_s^{ij}.

To arrive at a convex energy the label indicator variables are allowed to attain values between $[0, 1]$. In an ideal case we would like to get mostly binary assignments for the x_s^i, however, not only the convex relaxation requires non-binary assignments but also around a transition describing a surface we expect non-binary assignments. This is necessary to allow the transition gradients y_s^{ij} to point in arbitrary directions and is introduced into the formulation as an effect of the discretization of the original optimization problem defined on a continuous domain.

After summarizing the main ingredients of the convex multi-label formulation, we can now introduce our novel segment based 3D object shape prior formulation.

3. Segment based shape priors

As we have mentioned earlier our goal is to define spatially homogeneous anisotropic smoothness terms that act as a shape prior. We have observed earlier that the main problem with this approach is that not each object can be represented descriptively in this way. One solution to this problem is to have an anisotropic smoothness prior which is different for each voxel [9]. But this comes with the problem that the volume Ω needs to be exactly aligned with the object. Here, we propose a different solution, namely splitting the object into multiple segments and regularize each of the transitions between the segments differently, but with one single smoothness term for the whole volume. With this approach we can drop the necessity of an exact alignment. Only the main directions of the object need to be known.

3.1. From Wulff Shapes to Convex Segments

The definition of the anisotropic smoothness term ϕ^{ij} that we are using in the following, was presented in terms of a Wulff shape $W_{\phi^{ij}}$ [19] and is defined as

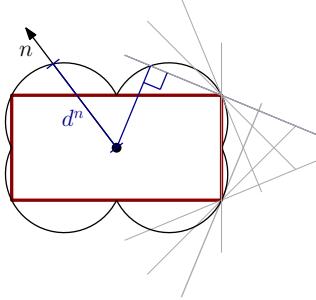$$\phi^{ij}(n) = \max_{p \in W_{\phi^{ij}}} p^T n. \quad (2)$$

Figure 3. Wulff shape illustration in 2D: The Wulff shape induces an anisotropic smoothness term. The distance from the origin to the black line in direction n is the cost d^n for a normal direction n. The construction is indicated with gray lines. Naturally the reconstructed shape using this prior has the same shape as the Wulff shape.

Mathematically, the Wulff shape is defined as convex shape that contains the origin. The name was given to them by George Wulff, who studied the growth of crystals in liquids [23]. His findings were that the equilibrium shape of a crystal is defined by the Wulff shape. This connection between equilibrium crystal shapes and anisotropic surface regularization means that describing the preference for a certain object shape can be achieved by using an anisotropic regularizer that is induced by a Wulff shape having the object's shape. The construction of the shape works as follows: Consider direction n, draw a plane orthogonal to n which has a distance d^n to the origin. d^n corresponds to the amount we want to penalize a surface of direction n. Doing this for all directions leads to a convex shape, the Wulff shape, which is formed by the inner envelope of all the planes. This relationship between distance of planes and cost in a direction is expressed through Eq. 2 . An illustration of how this works for a box Wulff shape is depicted in Fig. 3. Going back to the example of a table we now observe that the shape of the table is not convex and hence a single anisotropic smoothness term does not lead to a descriptive prior. Having made this observation, it becomes clear that in order to derive descriptive priors we need to segment our input object into parts which are sufficiently convex.

Let $m_1, \ldots, m_M$, be a set of M convex segments into which our input object is split. The problem of splitting a shape into convex parts has been well studied (for example [3, 16]). To be able to get a faithful reconstruction we also need to model the supporting ground of the object. It follows that the label space $\mathcal{L}$ is composed out of the labels 0: free space, 1: ground and all the segments, hence $\mathcal{L} = \{0, 1, m_1, \ldots, m_M\}$. With the segments being close to convex, they can be described nicely by convex regularization functions ϕ^{ij}. In order to utilize the convex energy Eq. 1, the data costs ρ_s^i and the regularization functions

ϕ^{ij} need to be defined. The ϕ^{ij} need to be defined for each of the transitions between labels i and j. They can either be derived from training data [10, 9] or manually designed. The ρ_s^i are the unary data costs. They describe a local preference for label i in voxel s and are derived from a set of input depth maps (see Sec. 3.3). There is no semantic classifier involved. The actual labels for the segments are chosen by the optimization based on the regularization term. Our method is illustrated in Fig. 2. First the table gets segmented into convex parts then we determine all the transitions. As a last step the smoothness terms for all the interfaces between the labels are defined in terms of Wulff shapes. In this step the outline of the convex segment gets split into multiple parts which have different neighboring interfaces. The combination of all these interfaces forms the prior for the whole segment.

In this work the label chosen for each of the convex segments coincide with semantic parts of the shape. The label for each segment is determined solely through our shape prior formulation. However, our method is not limited to these kind of segmentations. Semantic classifiers could be naturally included into the formulation as presented in [10]. For this work we chose not to include a semantic classifier to underline the strength of our prior.

3.2. Free Space as Convex Segments

Splitting the geometry into multiple segments is not limited to the occupied space. It is also possible to split the free space into multiple classes. This becomes interesting when reconstructing classes such as mugs which generally have a carved out inner part which is hard to observe in the input data. The general shape of a mug is highly non convex and in general it needs an almost infinite amount of convex segments. However, if we look at the mug with its inner space filled, then it is a convex object and also the filling itself is convex. Introducing a segment for the "inside free space" and interpreting it as a subtraction of a convex segment allows a highly non-convex object such as a mug to be efficiently handled by our approach. By having a preference in the unary term for "inside free space" it can be encoded that the object should be carved out if the input data does not disagree. Moreover, having a horizontal "inside free space" ↔ "free space" transition expensive we can ensure that the mug does not contain any unwanted holes.

Note that this choice of smoothness cost will not form a metric. In horizontal direction a transition "inside free space" ↔ "mug" plus a transition "mug" ↔ "free space" needs to be cheaper than the direct transition "inside free space" ↔ "free space", which violates the condition for a metric. This is not a problem, as the formulation we are using, explicitly allows for non-metric smoothness costs [26].
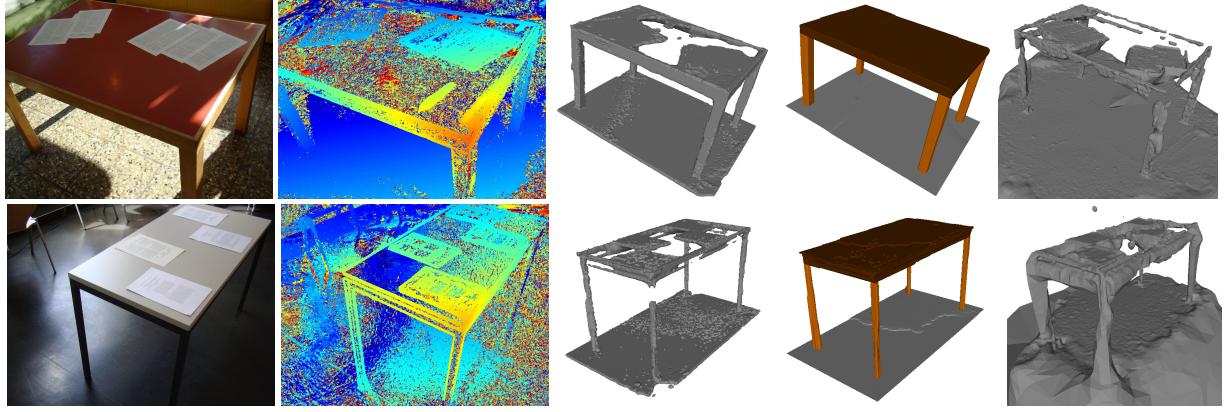
Figure 4. From left to right: example input image, example depth map, baseline TV-flux fusion result, result using our shape prior formulation, PMVS+PSR result.

3.3. Data Term

The method takes as input a set of depth maps. In order to use them in our formulation they need to be transformed into the unary data term ρ_s^i. The depth maps contain information about the geometry but no information about the individual labels i. Therefore our data term is only defined in terms of occupied space labels and free space labels. For a two label problem it is enough to define a data cost for one of the two labels, here we only define the cost for occupied space and denote it as ρ_s. Assigning 0 cost to free space labels implicitly assigns a cost of $-\rho_s$ to free space labels through the optimization. For the occupied label cost ρ_s at voxel s we follow the approach that assumes that in front of an estimated depth $\hat{d}_s$ in the depth map, associated with voxel s, we expect free space and behind we expect occupied space, in a small region [9]. If a voxels' depth d_s lies in front of the estimated depth $\hat{d}_s$ given in the depth map, we add a constant positive weight β to ρ_s and if d_s is behind $\hat{d}_s$ we add $-\beta$ to ρ_s. Along the viewing ray from the camera center up to $\hat{d}_s$ we expect free space. This assumption holds only if there are no outliers in the data. To use the information along the viewing ray and still stay robust against outliers we add only a small positive cost ϵ along the viewing ray to ρ_s.

4. Implementation

In this section we describe some details about our implementation. The smoothness priors for all transitions between all the labels can be derived from training data [10, 9]. The two proposed approaches differ in the way the Wulff shapes are parameterized. While in general any convex shape can be parameterized reasonably well as an intersection of half spaces [9] this has the drawback that shapes such as cylinders need many different half spaces for a faithful representation of the round parts. On the other hand, it is possible to define a catalog of basic shapes, such as a

sphere, a cylinder or a box and let the training procedure decide which of the elements of the catalog fit best with which parameters as a maximum likelihood estimation [10].

For both methods the input data to the training procedure is a set of mesh model of all the transitions. In order to have an efficient optimization we decided to use a slightly modified version of the training procedure given in [10]. The main difference is that we do not model the Wulff shape in terms of a sum of an anisotropic and an isotropic part but directly derive the parameters for one single Wulff shape. With n^{ij} we denote the normal vector for a boundary between label i and j. $i \leftrightarrow j$ denotes a transition between i and j. Now we define the probability of having a transition with normal direction n^{ij} by

$$P(n^{ij}) = \exp(-\phi^{ij}(n^{ij})), \tag{3}$$

subject to $\int_{n\in\mathbb{S}^2} \exp(-\phi^{ij}(n^{ij}))dn = P(i \leftrightarrow j)$. As basic shapes we consider a box, a cylinder and a hemisphere plus an attached spherical cap [10]. The only difference to [10] is that we now need to fulfill the above constraint while training. This, we addressed by doing a grid search for all but one parameter of the basic shape and find the last one using Brent's method [1] such that the constraint is fulfilled. This slightly adapted grid search approach finds the maximum likelihood estimate of the Wulff shape parameters.

After having the smoothness in place, the convex multi-label energy Eq. 1 is minimized with the first order primal-dual algorithm [20]. The energy can easily be transformed into the primal-dual saddle point form by inserting the primal-dual formulation (Eq. 2) and Lagrange multipliers for the constraints (c.f. [9]).

5. Results

We evaluated our method on the challenging real-world objects classes, trees, tables, dumbbells and a mug. We collected images with either a hand held compact camera and
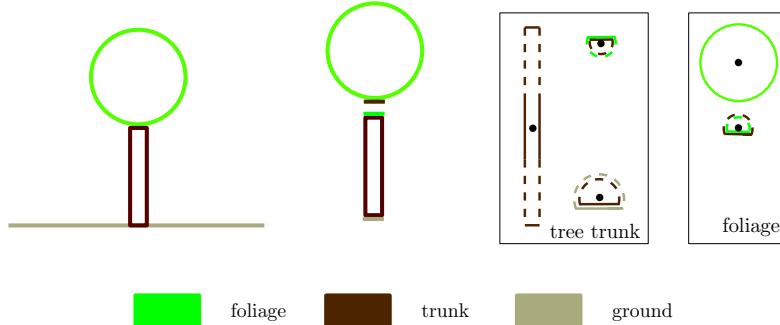
Figure 5. Shape prior formulation for the trees: The tree gets segmented into convex parts (foliage and trunk), the transitions of each part to other labels are taken into account, for each transition the smoothness is defined in terms of Wulff shapes. The dashed lines indicate that the shape is very large in the indicated direction, meaning that this is an unlikely interface in the final reconstruction.
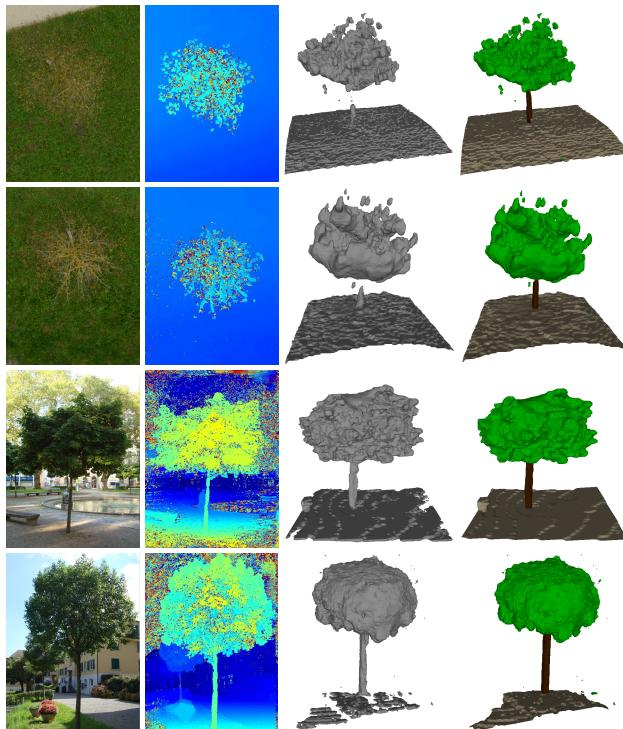

Figure 6. From left to right: Example input image, example depth map, baseline TV-flux fusion result, result using our shape prior formulation.

in addition for the trees aerial images captured by a flying drone. The camera poses are obtained by using the publicly available structure from motion software [27]. As an input to our formulation we computed depth maps using plane sweeping stereo matching [24] for each of the images. As an image dissimilarity measure we used the zero-mean normalized cross-correlation (ZNCC). For occlusion handling, we only accept the best K matching scores [11]. The main directions of the objects where determined by a manual alignment step.

We compare all our results to a baseline approach which reconstructs the same unary terms ρ_s using an isotropic regularization (TV-flux fusion of [25]). This choice of baseline directly shows the effect of using our prior. To illustrate how other state of the art reconstruction methods perform on our datasets we include results using patch based multi-view stereo (PMVS) [8] and poission surface reconstruction (PSR) [12], for some object classes. PMVS was run with the default parameter values and for PSR we hand tuned the parameters for best reconstructions. The results for the individual object classes and their main difficulties are discussed in the following.

5.1. Table

We gathered a training dataset of 11 mesh models of rectangular IKEA dining and coffee tables. As a first step we split the training mesh models into the individual segments by cutting off the table legs. This segments the table into a tabletop and table legs, Fig. 2 Then we extract mesh models for all the individual transitions plus a manually inserted ground plane. This data is then inserted into our training procedure to obtain the Wulff shapes, which defines our shape prior for the table class.

The individual Wulff shapes that are used to reconstruct a table are indicated in Fig. 2. We observe that the leg has a strong preference to be vertical, on the top it can cheaply change to tabletop and on the bottom to ground. This very descriptive nature of the geometry leads to a very strong prior for the table class. The smoothness of the table top prefers a box with a large horizontal extent.

The results we obtained are depicted in Fig. 4. The main problems with the baseline approach not using a shape prior are holes in the tabletop and disconnected legs. Our prior is able to resolve these issues in the reconstruction. We also get a reliable segmentation of the table into its semantic parts.
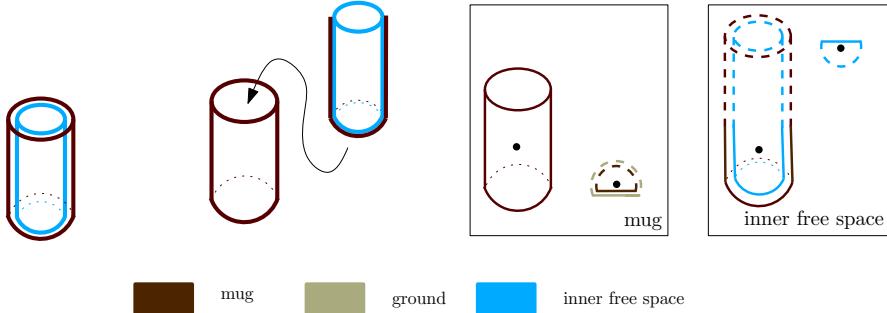
Figure 7. Shape prior formulation for the mug: The mug is segmented as a single convex part (mug) and an interior convex part (inner free space) which gets carved out, the transitions of each part to other labels are taken into account, for each transition the smoothness is defined in terms of Wulff shapes. The dashed lines indicate that the shape is very large in the indicated direction, meaning that this is an unlikely interface in the final reconstruction.
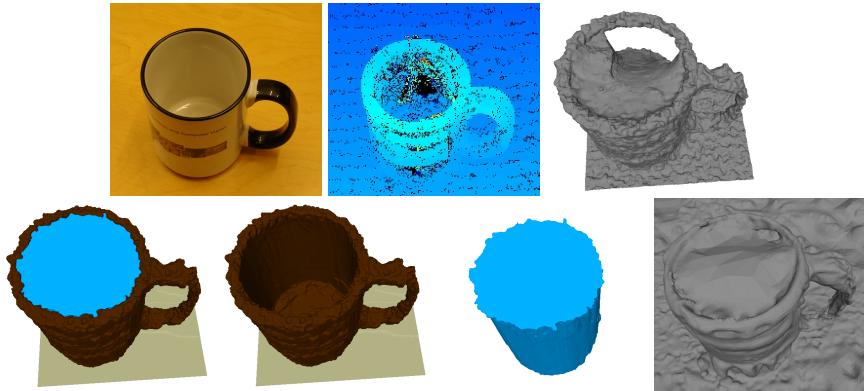

Figure 8. Top row, from left to right: example input image, example depth map, baseline TV-flux fusion result. Bottom row, from left to right: result of our formulation (inner free space inside the mug), result of our shape prior formulation (mug and inner free space separately), PMVS+PSR result.

5.2. Tree

We manually split trees into foliage and trunk. The Wulff shapes and their parameters where hand selected and are depicted in Fig. 5. The behavior of the smoothness is similar as for the table. The trunk is a mostly vertical segment that connects to the foliage on the top and to the ground on the bottom. The foliage is modeled with an isotropic prior. The smoothness between the foliage and the ground is chosen to be a very strong isotropic smoothness as such a transition very rarely occurs.

For the evaluation of the object class trees we used two types of imagery. We used terrestrial images as well as aerial images. The results for both are depicted in Fig. 6. On terrestrial images already the baseline produces faithful geometry but our shape prior formulation gives in addition to the geometry also a segmentation of the object. An important case for trees are aerial images. In this case trunks get often disconnected as they are hard to observe from the air. Our formulation allows for a complete reconstruction of very weakly observed trunks.

5.3. Mug

In order to be able to reconstruct a mug faithfully it is important to reconstruct the inside of the mug correctly. For many methods this is challenging as the concavity is often weakly observed in the depth maps and moreover often false matches on the top lead to a filled mug. With our method we can model the inside free space as a segment of the free space. This allows us to make a concavity cheap. We also include a slight preference for the inside free space label in the data term. The smoothness of the mug and the inside free space is modeled with cylinder shapes and shapes preferring a particular single direction. The shape prior formulation is indicated in Fig. 7. Note, that a transition of inside free space to any other label than mug is only cheap in the vertical direction. This ensures that there are no holes in the mug. For this class we manually defined the Wulff shapes.

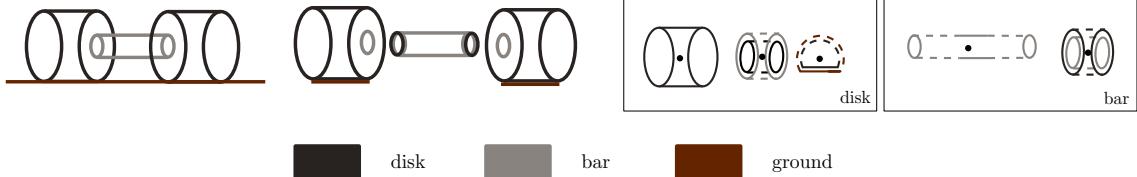Fig. 8 shows an example of a reconstruction of a mug.

Figure 9. Shape prior formulation for dumbbells: Dumbbells get segmented into convex parts (bar and disks), the transitions of each part to other labels are taken into account, for each transition the smoothness is defined in terms of Wulff shapes. The dashed lines indicate that the shape is very large in the indicated direction, meaning that this is an unlikely interface in the final reconstruction.
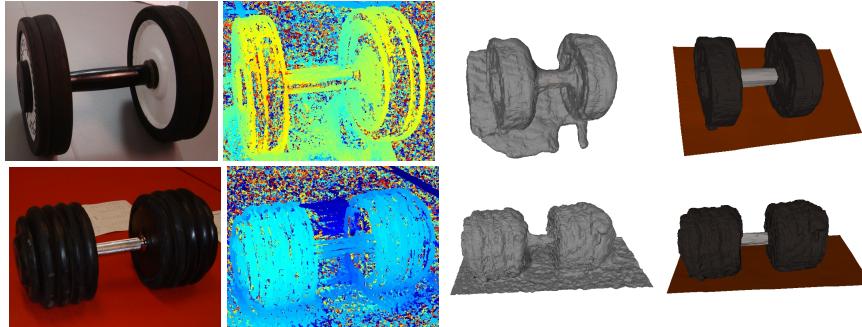

Figure 10. From left to right: Example input image, example depth map, TV-Flux fusion result, result of our shape prior formulation.

The baseline method is not able to reconstruct the mug correctly. Also the additional PMVS+PSR result shows the same behavior. Using our proposed shape prior the interior gets carved out correctly and there are no holes in the mug surface.

5.4. Dumbbell

Also dumbbells can nicely be modeled with our formulation. A dumbbell is geometrically made out of different cylinders. Hence most of the Wulff shapes are cylindrical (see Fig. 9). Our results on this class are depicted in Fig. 10. While the geometry is already faithfully reconstructed using the baseline our formulation outputs a segmentation in addition to the geometry.

6. Conclusion

In this paper we described an approach to shape priors which segments the object into multiple convex segments. The prior is defined in terms of anisotropic smoothness terms for the individual transtions between the labels in a volumetric setting. Because of the spatial homogenity of the smoothness term, our method has the advantage that the prior does not need to be exactly aligned with the object. We only require an alignement with respect to the orientation of the object. However, our prior is still descriptive due to the splitting of the object into multiple segments. For the objects which we used for this work, the segments also naturally agree with the semantic classes.

In our evaluation we demonstrated on multiple object classes that our prior is able to overcome the limitations of the baseline approach that does not use a shape prior. The often appearing holes and disconnected parts are fixed with our shape prior formulation. At the same time a convincing semantic segmentation is obtained as a side product of our method. We also demonstrated that we can correctly carve out concavities on the example of a mug.

Our formulation could also naturally be combined with semantic classfiers, leading to a joint reconstruction and recongnition formulation. In the future we also plan to investigate the reconstruction of larger scenes with multiple different objects present. This poses novel problems to the optimization as the number of variables in the optimization problem grows quadratically with the number of labels.

References

[1] R. P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 1971.

[2] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences (SIIMS)*, 2012.

[3] B. Chazelle and D. P. Dobkin. Optimal convex decompositions. *Computational Geometry*, 1985.

[4] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *International Con-*

ference on Computer graphics and interactive techniques (SIGGRAPH), 1996.

[5] A. Delong, L. Gorelick, O. Veksler, and Y. Boykov. Minimizing energies with hierarchical costs. *International journal of computer vision (IJCV)*, 2012.

[6] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010.

[7] P. F. Felzenszwalb and O. Veksler. Tiered scene labeling with dynamic programming. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[8] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2010.

[9] C. Häne, N. Savinov, and M. Pollefeys. Class specific 3d object shape priors using surface normals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[10] C. Häne, C. Zach, A. Cohen, R. Angst, and M. Pollefeys. Joint 3d scene reconstruction and class segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[11] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[12] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Eurographics symposium on Geometry processing (SGP)*, 2006.

[13] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *European Conference on Computer Vision (ECCV)*, 2010.

[14] V. Lempitsky and Y. Boykov. Global optimization for shape fitting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[15] X. Li, T. W. Woon, T. S. Tan, and Z. Huang. Decomposing polygon meshes for interactive applications. In *Symposium on Interactive 3D graphics*, 2001.

[16] J.-M. Lien and N. M. Amato. Approximate convex decomposition of polyhedra. In *ACM symposium on Solid and physical modeling (SPM)*, 2007.

[17] H. Liu, W. Liu, and L. J. Latecki. Convex shape decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[18] K. Mamou and F. Ghorbel. A simple and efficient approach for 3d mesh approximate convex decomposition. In *IEEE International Conference on Image Processing (ICIP)*, 2009.

[19] S. J. Osher and S. Esedoglu. Decomposition of images by the anisotropic rudin osher fatemi model. *Communications on pure and applied mathematics*, 2004.

[20] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[21] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Conference on Computer vision and pattern recognition (CVPR)*, 2006.

[22] E. Strekalovskiy and D. Cremers. Generalized ordering constraints for multilabel optimization. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[23] G. Wulff. Zur frage der geschwindigkeit des wachstums und der auflsung der krystallflächen. *Zeitschrift für Krystallographie und Mineralogie*, 1901.

[24] R. Yang and M. Pollefeys. Multi-resolution real-time stereo on commodity graphics hardware. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[25] C. Zach. Fast and high quality fusion of depth maps. In *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2008.

[26] C. Zach, C. Häne, and M. Pollefeys. What is optimized in convex relaxations for multilabel problems: Connecting discrete and continuously inspired map inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2014.

[27] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.