This CVPR2014 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

Class Specific 3D Object Shape Priors Using Surface Normals

Christian Häne, Nikolay Savinov, Marc Pollefeys
ETH Zürich, Switzerland
{chaene, nsavinov, pomarc}@inf.ethz.ch

Abstract

Dense 3D reconstruction of real world objects containing textureless, reflective and specular parts is a challenging task. Using general smoothness priors such as surface area regularization can lead to defects in the form of disconnected parts or unwanted indentations. We argue that this problem can be solved by exploiting the object class specific local surface orientations, e.g. *a car is always close to horizontal in the roof area. Therefore, we formulate an object class specific shape prior in the form of spatially varying anisotropic smoothness terms. The parameters of the shape prior are extracted from training data. We detail how our shape prior formulation directly fits into recently proposed volumetric multi-label reconstruction approaches. This allows a segmentation between the object and its supporting ground. In our experimental evaluation we show reconstructions using our trained shape prior on several challenging datasets.*

1. Introduction

In the last few years a lot of progress has been made in the area of dense 3D reconstruction from images. The main difficulty is imperfect input such as noisy, missing or ambiguous data. Most approaches have in common that they try to find a solution which follows the input data while being sufficiently regular with respect to a given smoothness prior. These priors often penalize surface properties such as *e.g.* area or curvature. Such approaches work well for data which is mostly affected by noise. If it comes to defects such as specularities, translucency or even reflections these methods are prone to produce suboptimal solutions.

To overcome this problem strong domain specific priors are used. These can be priors on semantic classes, *e.g.* building facades have to be vertical, or priors on the structure, *e.g.* the reconstruction has to form a tree. Recently also 3D object shape priors have been proposed.

One powerful approach to dense 3D reconstruction is volumetric depth map fusion, where a set of depth maps is combined into a volumetric representation. The volume is segmented into the exterior (free space) and interior (occupied space) of the object. In this work we propose to exploit anisotropic regularization to embed object shape priors into convex regularization formulations for volumetric depth map fusion. The reasoning behind this is that object classes have locally only a restricted distribution of surface normals over different examples. The outline of a bottle is always close to vertical but on the other hand there can be regions such as the side mirrors of a car where a large set of normals is likely. We capture these normal variations and utilize them to formulate our shape prior.

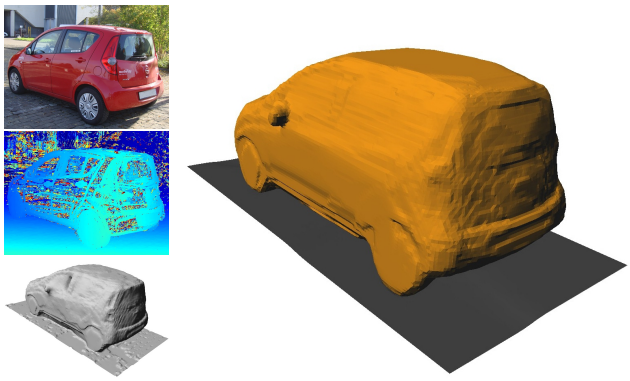


Figure 1. Left: Cropped input image, depth map and baseline result. Right: Result with our shape prior formulation. Object and ground are segmented by our formulation.

1.1. Related Work

There is a broad literature on dense 3D reconstruction from images. We focus on the most related ones to our approach and for a more detailed overview we refer the reader to [15]. The main idea of volumetric depth map fusion is to segment a volume into occupied and free space and extract the surface as boundary between the two regions. The main difference between the various methods is how they regularize the data. Approaches penalizing the surface area have been proposed in the discrete optimization literature [12] and also as continuous convex optimization [22]. The main advantage of such continuously inspired formulations is that they have an underlying continuous representation, which leads to more isotropic behavior compared to discrete approaches using graph-cuts.

Given input images an oriented point cloud can be extracted using structure from motion approaches. Utilizing anisotropic smoothness can then be used to align the final reconstruction with the extracted normal directions leading to a more faithful reconstruction of surface details [11].

While such general smoothness priors lead to faithful reconstructions in the presence of well behaved data they can be too general for very ill-conditioned cases. Therefore other priors have been proposed.

Segmenting elongated structures, such as blood vessels, are prone to be disconnected. To overcome this problem connectivity priors have been introduced [18, 17].

Object shape priors have recently attained much attention in the field of dense 3D reconstruction [3, 5]. Both works try to fit a parametric shape model into input data. This allows the reconstruction of objects which are not well represented in the input data such as highly reflective cars.

Multi-label segmentation methods have only recently emerged. The work of [4] proposes a spatially continuous formulation of multi-label segmentation with a discrete label space. While anisotropic smoothness can be directly applied to this formulation, the main restriction is that the smoothness needs to form a metric over the label space. A comparison of discrete formulations and the discretized version of the continuous formulation revealed that the discretized formulation does not suffer from this problem [20]. The most closely related work to our shape prior formulation is [9]. They propose to introduce semantic information into volumetric 3D reconstruction by utilizing convex multi-label segmentation. The main idea is that the volume is no longer segmented into free and occupied space but rather into multiple semantic classes. This allows the semantic classes to be penalized differently, *e.g*. ground should be less penalized when in a horizontal direction, as this is a more likely direction for ground.

Our approach to shape priors is driven by the observation that, for many real life objects, the local distribution of normals is limited. For example the doors of a car cannot be horizontal but this is a very likely direction for the hood. Therefore, we propose to incorporate a shape prior by locally penalizing the surface directions anisotropically dependent on the distribution of normals given in training data. A more direct approach for volumetric shape priors might be to just derive a location specific prior on the occupancy state of a voxel. However, this does not allow the capturing of that *e.g*. the roof of the car needs to be close to horizontal, which is nicely captured by our direction based prior.

Often the reconstructed object is standing on a supporting ground. To be able to handle the data costs induced by the ground, which do not fit to the shape prior, we propose to use multi-label segmentation to treat the ground differently. This directly enables us to also segment the object and the ground.

1.2. Our Contributions

- We propose a novel way to incorporate shape priors based on surface normal distributions into convex multi-label optimization. The driving force of our method are spatially varying anisotropic smoothness priors (*c.f*. Sec. 2.3). The formulation is based on more general anisotropic smoothness priors as the ones that have been used in recent convex multi-label segmentation approaches.
- An efficient way to derive the shape priors from training data is detailed in Sec. 2.4.
- Due to the multi-label formulation we can segment the object from its supporting ground in a natural way (see Fig. 1).

2. Shape Priors

Our approach is based on recently published convex multi-label segmentation methods [16, 4, 20]. Before we explain how we can incorporate shape priors, we need to summarize the optimization framework.

2.1. Convex Multi-Label Segmentation

The ultimate goal of surface reconstruction is to find the most probable surface which fits the given noisy data. In convex approaches for surface reconstruction, the surface is often implicitly represented in a volumetric form. Traditionally, the volume is separated into two areas, free space and occupied space [12, 19] and the resulting surface is described by the boundary between free and occupied space. An extension of this formulation to multiple semantic classes was proposed in [9]. We follow a similar approach by dividing the reconstruction into the object itself, on which the shape prior should be applied, and the ground, if there is any. This leads to a segmentation of the volume into three regions: free space, object and supporting ground. This allows us to handle the often strongly observed ground which does not fit to the object shape prior. Additionally, this directly gives the segmentation between the reconstructed object and the supporting ground.

The main ingredients of the formulation are: a unary and a smoothness term plus a set of constraints linking them together. The unary potential is computed from the observed data and will be called data term. A smoothness term is required to tackle the noise in the observed data. It can be seen as a prior on the expected surface and is therefore named smoothness prior.

The convex multi-label segmentation formulation presented in the following originates from continuous optimization [4]. The main drawback of the purely continuous formulation is that it does not allow for non-metric smoothness priors. It has been pointed out that this can be overcome by only considering discrete domains in [20]. As

we are inherently bound to the discrete domain and the underlying continuous formulation is not required for the understanding of this manuscript, we directly present the discretized version of the formulation. However some mechanisms need to be understood as a discretization of the underlying continuous formulation.

More formally, a discretized volumetric domain $\Omega \subset \mathbb{R}^3$ is given. The task of the label assignment is to assign one out of the three labels free space, object or supporting ground $\mathcal{L} = \{0, 1, 2\}$ to each location (voxel) $s \in \Omega$. It is represented as soft indicator vectors $x_s \in [0, 1]^{|\mathcal{L}|}$ with $x_s^i = 1$ if and only if label i is assigned at s. In the following, we will use subscripts to denote the spatial position of a variable and superscripts to denote label indices. e_k denotes the k-th canonical basis vector and $(\cdot)_k$ is used to extract the k-th component of a vector. Having this notation in place, we can now state the objective of the utilized convex minimization problem.

$$E(x) = \sum_{\Omega \in s} \left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j} \phi_s^{ij}(x_s^{ij} - x_s^{ji}) \right) \quad (1)$$

In addition to the per voxel soft-indicator vectors x_s^i the objective also contains vectors $x_s^{ij} \in [0, 1]^3$. They are understood as follows: the $k-$th component of the vector x_s^{ij} equals 1 if and only if there is a transition from label i to label j along the k-th canonical direction of the grid. The functions $\phi_s^{ij} : \mathbb{R}^3 \to \mathbb{R}_0^+$ define the smoothness prior and need to be convex and positively 1-homogeneous. Their arguments $x_s^{ij} - x_s^{ji}$ are aligned with the surface normal if they are on the segmentation boundary. They are an anisotropic extension of standard surface area based penalization used in continuous segmentation formulations (c.f. [7, 23]). Note that originating from a discretized continuous formulation, the arguments of the objective are generally non-binary around the optimal surface. This is a necessary behavior to allow the vectors $x_s^{ij} - x_s^{ji}$ to point in arbitrary directions, and the main reason why these formulations are less prone to metrication artifacts than purely graph based approaches. The ρ_s^i are the unary data costs giving a local preference at voxel s for the label i. To ensure that the edge indicator variables x_s^{ij} and the node indicator variables x_s^i are consistent and in a valid configuration, the objective in Eq. 1 is subject to the following marginalization, normalization and non-negativity constraints.

$$x_s^i = \sum_j (x_s^{ij})_k, \quad x_s^i = \sum_j (x_{s-e_k}^{ji})_k \quad k \in \{1, 2, 3\}$$
$$x_s \in \Delta, \quad x_s^{ji} \geq 0 \quad (2)$$

The marginalization constraints ensure that the per node and per edge variables agree on the assigned labels. Constraining the per-node variables to lie in the unit probability simplex $\Delta = \{\sum_i x_s^i = 1, x_s^i \geq 0\}$ ensures that exactly one label is assigned at each position s.

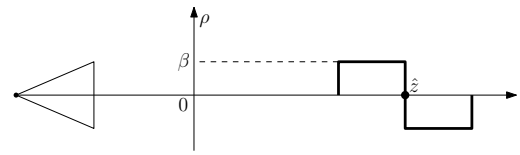


Figure 2. Data term entered into the grid along a viewing ray.

2.2. Data Term

The unary data term gives a local preference for a certain label. In our formulation, we have two different labels denoting occupied space: object and ground. It has been shown that such multi-label segmentation can be done based on per image class scores in [9]. We choose a slightly different approach and do not compute any class specific data term. The segmentation of object and ground is solely based on the shape prior detailed in Sec. 2.3. The only input for the data term is a set of potentially noisy depth maps. This leads to only one data term describing a local preference for occupied space $\rho_s := \rho_s^1 = \rho_s^2$. The local preference for free space is implicitly encoded as penalty for occupied space, therefore $\rho_s^0 = 0$.

In the following we consider a particular voxel s and its associated data cost ρ_s. The projective depth of s with respect to one of the input depth maps is denoted as z_s and its observed depth in the depth map is denoted as $\hat{z}_s$. We then define the data term as

$$\rho_s = \begin{cases} \beta & \text{if } z_s < \hat{z}_s, \hat{z}_s - z_s < \delta \\ -\beta & \text{if } z_s > \hat{z}_s, z_s - \hat{z}_s < \delta \\ 0 & \text{otherwise .} \end{cases} \quad (3)$$

This is based on the assumption that in front of an observed depth $\hat{z}$ a voxel s should be assigned free space and in a neighborhood behind occupied space. The weights ρ along a viewing ray are depicted in Fig. 2. Note that we are minimizing our energy and therefore a negative ρ_s means preference for occupied space. This specific choice for ρ_s corresponds to an exponentially distributed noise assumption for depth inliers (see also [9, 19]). Also note that we have only specified our data term for a single observed depth $\hat{z}_s$. It remains to be mentioned that the final data term is formed by summing up all the weights from all the input depth maps.

2.3. Novel Shape Prior Formulation

Unlike recent works on shape priors that fit a parametric shape model to given data [3, 5] we are taking a somewhat different approach. The main drawback in the aforementioned works is that using a parametric shape model brings the additional difficulty that details not represented by the prior need to be added in an additional refinement step. In this section we will detail our shape prior formulation.

For many real world objects, the distributions of normals in a set of given examples is locally restricted. For example, the outline of a bottle is always close to vertical. We exploit this observation to define our shape priors in terms of surface normal distributions. The main idea is to use a

spatially varying anisotropic penalization of the surface.

This can be included in the convex volumetric multi-label segmentation formulation by utilizing smoothness priors $\phi_s^{ij}(\cdot)$ that penalize the surface according to the expected surface orientation at each location s in our volume Ω. For the remainder of the paper, we assume that the domain Ω is aligned with the object's bounding box. At this point, we would like to mention that the smoothness priors $\phi_s^{ij}(\cdot)$ need to be defined for all the possible transitions between the three labels (free space, object and ground). However, the shape prior is only applied to transitions involving the *object* label, meaning the transitions *object* $\leftrightarrow$ *free space* and *object* $\leftrightarrow$ *ground*. In the remainder of this section we therefore omit the label indices and implicitly refer to the relevant transitions.

Any positively 1-homogeneous convex function can be used as a smoothness prior ϕ_s. Choosing functions fulfilling these properties can be difficult, therefore we do not define them directly but use the primal-dual formulation (c.f. [7])

$$\phi_s(x) = \max_{p \in \mathcal{W}_s} (p^T x). \tag{4}$$

$\mathcal{W}_s$ denotes the so-called Wulff shape and can be any closed, bounded and convex set which contains the origin. Using this formulation, we transformed the problem of specifying a suitable function to specifying a convex shape which guarantees the definition of a valid smoothness prior.

This idea has been used before, but only a restricted set of possible shapes was taken into account [23, 9]. Namely, one that prefers a certain direction and one that prefers all the directions that are orthogonal to a given direction. This may be sufficient in these works, but in our case, in order to capture the distribution of normals present on the object's surface faithfully, we need to allow for more general smoothness priors. Therefore, we propose to use a more general parameterization of Wulff shapes.

We start with the following fact: any convex shape can be described as an intersection of a set of half spaces. We choose a discrete set of half spaces $\mathcal{H}_s$ and call the induced shape $\mathcal{W}_{\mathcal{H}_s}$ the discrete Wulff shape. The half spaces $h_s \in \mathcal{H}_s$ are defined by a normal direction $n \in \mathcal{S} \subset \mathbb{S}^2$ and distances to the origin d_s^n. $\mathbb{S}^2$ denotes the three dimensional unit sphere and $\mathcal{S}$ a discrete subset thereof. The distances d_s^n are indexed by the normal direction n. This is not a restriction as only one half space per direction can be active (having a joint boundary with $\mathcal{W}_s$) in the Wulff shape defined by the intersection of all $h_s \in \mathcal{H}_s$. The set of normal directions $\mathcal{S}$ is chosen to be spatially homogeneous and therefore not indexed by the position s. By using a sufficiently fine grained discretization for the normal directions in $\mathcal{S}$, any valid Wulff shape can be represented with high accuracy. Fig. 3 shows visualizations of some examples of possible shapes.

In our experiments, we use a geodesic sphere, which is generated as a iterative subdivision of an icosahedron, to define the set of normal directions $\mathcal{S}$. The number of directions used is $|\mathcal{S}| = 162$. The actual distances d_s^n used for the individual directions n are determined by training data. This step is detailed in the Sec. 2.4.

Figure 3. Discrete Wulff shapes, from left to right: isotropic, preferring one main direction, more complex shape that penalizes a segment of normals much more strongly than the rest.

For the transitions between ground and free space, which are not covered by the shape prior we use a prior preferring the ground to be horizontal as done in [9].

For objects that only touch the ground at specific places *e.g.* the four wheels of a car, the position of the support points are known from the training data. This allows us to utilize the shape prior only in a neighborhood around the area where the interaction is expected to happen. Outside that region we use a strong isotropic penalization. In Sec. 4 it is shown that this enables the optimization to properly segment cars from the ground.

2.4. Training of the Discrete Wulff Shapes

The aim of training the Wulff shapes is to determine the parameters of the shape prior based on training data. Our training data are mesh models that we downloaded from the Internet. Often realistic models also contain the geometry in the interior of the object. In our case we are only interested in the outline shape of the models, more specifically only in the normals of the outline shapes at all positions $s \in \Omega$. To separate the outline shape from the rest we first render depth maps from the given mesh and run volumetric depth map fusion [19] using the rendered depth maps in order to finally extract the outline shape's normals.

The parameters that need to be determined during training are the parameters of the discrete Wulff shapes, namely all the distances d_s^n for all the normals $n \in \mathcal{S}$ at all the positions $s \in \Omega$. Naturally, the terms ϕ_s^{ij} in the energy function in Eq. 1 correspond to negative log-probabilities. The standard approach for the training would be to find a maximum likelihood estimation of the distances d_s^n given a set of training normals. As we need to estimate a set of parameters at each position s a gradient descent based maximum likelihood estimation would be too slow.

To make the problem tractable we discretize the space of directions $x_s^{ij} - x_s^{ji}$ to the ones that are aligned with the discrete set of normals $\mathcal{S}$. Therefore we have for each unit length normal direction $n \in \mathcal{S}$ at position s a direct mapping between the probability $P_s(n)$ and the corresponding d_s^n by

$$P_s(n) = e^{-\phi_s^{ij}(n)} = e^{-\max_{p \in \mathcal{W}_{\mathcal{H}_s}} (p^T n)} = e^{-d_s^n} \tag{5}$$

and it follows that

$$d_s^n = -\log\left(P_s(n)\right). \tag{6}$$

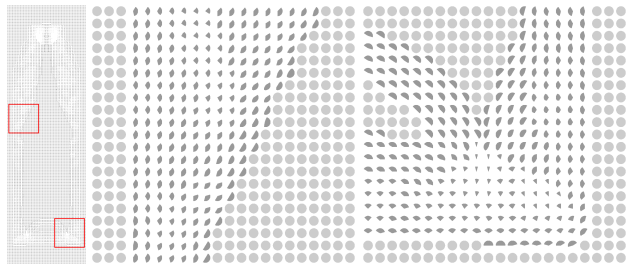
To get an estimate for $P_s(n)$ we generate histograms over the input normal directions weighted by their respective surface area in the training data. The bin centers are set to be $n \in \mathcal{S}$. Now the empirical probabilities can be directly extracted from the histograms and define the d_s^n. For robustness against non observed directions in the training data we impose a maximum on the values d_s^n. At this point we need to make a few remarks.

- In Eq. 5 the arguments to $\phi_s^{ij}(\cdot)$ are unit length normal vectors. However, the arguments $y_s^{ij} := x_s^{ij} - x_s^{ji}$ are in general non-normalized gradient directions from $[-1,1]^3$. From 1-homogeneity it follows that $\phi_s^{ij}(y_s^{ij}) = \|y_s^{ij}\|_2 \phi_s^{ij}(y_s^{ij}/\|y_s^{ij}\|_2)$. And $\|y_s^{ij}\|_2$ corresponds to the area of the surface element in the underlying finite-difference discretization (c.f. [9]). Therefore the surfaces are naturally penalized by their surface area.

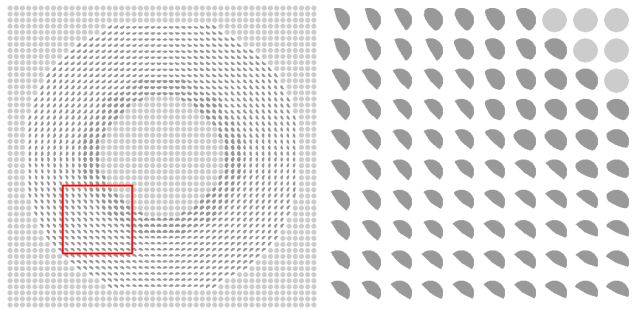
- It can happen that an actual half space h_s is not active (does not share a boundary with $\mathcal{W}_{\mathcal{H}_s}$) in the trained discrete Wulff shape. This means that the corresponding direction n will be penalized less then the training data suggests. This is not an issue as it does not happen for the most probable directions. This behavior can actually be beneficial if a single empty histogram bin is surrounded by much more probable ones.

- In the training process we restricted the directions to the normals $n \in S$. We can still allow for arbitrary directions during the optimization as we generate a valid Wulff shape which assigns a cost to all directions $x_s^{ij} - x_s^{ji}$.

Overall, this way of training the parameters leads to an estimate that penalizes the most probable directions $n \in S$ with the corresponding empirical probability and smoothly blends for directions in between. The training is very efficient as it only involves the computation of a normal histogram while getting an accurate penalization for the most likely directions. In Fig. 4 the shape prior trained on a set of bottles is visualized.



(a) Full slice on the left, closeups in the middle (left side of the bottle neck) and on the right (bottom corner of the bottle).



(b) Full slice on the left, closeup of the lower left part on the right.

Figure 4. Slices through the bottle shape prior in vertical direction (a) and horizontal direction (b). Isotropic Wluff shapes for positions s where there was no training data available are shown in light gray. Trained discrete Wulff shapes are shown in dark gray.

3. Optimization

For the optimization we use the preconditioned first order primal-dual algorithm presented in [14]. In order to apply the algorithm we need to rewrite the optimization problem in its primal-dual saddle point form. This is done by inserting the primal-dual formulation of the smoothness prior from Eq. 4 into Eq. 1 and introducing Lagrange multipliers for the constraints (Eq. 2). The optimization problem can then be stated as

$$\begin{aligned} E_{\mathrm{pd}}(x,p,\lambda,\gamma,\mu) = &\sum_s \left(\sum_i \rho_s^i x_s^i + \mu_s \left(\sum_i x_s^i - 1 \right) \right) \\ &+ \sum_{s,i,k} (\lambda_s^i)_k \left(\sum_j (x_s^{ij})_k - x_s^i \right) + \sum_{s,i,k} (\gamma_s^i)_k \left(\sum_j (x_{s-e_k}^{ji})_k - x_s^i \right) \\ &+ \sum_s \sum_{i,j:i<j} \left((p_s^{ij})^T (x_s^{ij} - x_s^{ji}) \right) \end{aligned} \tag{7}$$

subject to $p_s^{ij} \in \mathcal{W}_s^{ij}$, $\mu_s \geq 0$, $x_s^{ij} \geq 0, x_s^i \geq 0$. The objective is minimized with respect to x and maximized with respect to p, λ, γ and μ. The Wulff shapes $\mathcal{W}_s^{ij}$ define the smoothness prior and describe depending on the labels i and j; either our shape prior, or an isotropic or horizontal direction preferring prior.

The updates of the primal-dual algorithm [14] are gradient steps followed by projections to the feasible area. The main difficulty is the projection to the discrete Wulff shapes $\mathcal{W}_{\mathcal{H}_s}$ which define our shape prior. We detail this step in the following section.

3.1. Projection onto Discrete Wulff Shapes

At each iteration of the algorithm the dual variables p_s need to be projected back to the feasible area. In case of the discrete Wulff shapes $\mathcal{W}_{\mathcal{H}_s}$ this has to be done whenever p_s lies outside one or multiple of the half spaces $h_s \in \mathcal{H}_s$. In the following we denote such half spaces h_s as violating

half spaces. In our procedure we exploit that in practice only a few half spaces per Wulff shape are violated in each iteration of the optimization.

As a preprocessing step that only has to be done once for each shape prior we compute the intersection of all the half spaces $h_s \in \mathcal{H}_s$ and store for each half space, in case it is forming a facet of the intersection, the neighboring half spaces along the boundary of the facet. Computing the intersection of the half spaces can be done by utilizing point plane duality (c.f. [6]). For each half space we take the plane defining its boundary, $\pi = [n \quad -d]$ and map it to a dual point $\bar{\pi} = (1/d)n$. A subsequent computation of the convex hull [1] of the dual points reveals the intersection of the planes. The convex shape can then be extracted by transforming back to the primal representation. Note that this way of computing the intersection of half spaces is only valid because the origin is always contained within the shape.

For the projection

$$p'_s = \arg \min_{p'_s \in \mathcal{W}_{\mathcal{H}_s}} \|p_s - p'_s\|_2 \tag{8}$$

of a dual point p_s to the Wulff shape $\mathcal{W}_{\mathcal{H}_s}$ we use the following observation.

If a point p_s lies outside $\mathcal{W}_{\mathcal{H}_s}$, its projection p'_s lies on the boundary of one of the violating half spaces. This can be seen by case distinction. The projection either lies on a facet, edge or vertex of $\mathcal{W}_{\mathcal{H}_s}$, and for all three cases the area where the point p_s needs to be located to project to p'_s is at least outside of one of the half spaces incident to p'_s.

In our projection we first find the set of violating half spaces. Due to the above observation the projection can then be reduced to an exhaustive check of all the violated half spaces (remember in practice only very few half spaces are violated). At this point we would like to make a few remarks:

- Finding the set of all violating half spaces can be expensive. By storing the closest distance $d_s^{\min} := \min_{n \in \mathcal{S}} d_s^n$ for each $\mathcal{W}_{\mathcal{H}_s}$ we can first check if p_s lies withing the $d_s^{\min}$-ball and only do the expensive check if it does not. This considerably speeds up the optimization.
- At many positions s in the volume there is no training data available because the surface never passes through s. In this case an isotropic prior with a high smoothness cost is utilized. To reduce the memory usage we do not store a discrete Wulff shape but a perfect isotropic prior with the corresponding weight.

4. Experiments

Before presenting the results we obtained on challenging real-world datasets we show a synthetic experiment that shows how our shape prior handles shape variations. Often shape variations within a certain class are still constrained to a set of normals which can be captured by our shape prior. For example looking at a door with respect to the frame, or the door of a car with respect to the car, the different opening angles induce big shape variations that cannot be captured by simply capturing the statistics of the occupancy state of a voxel. To demonstrate that our prior is able to capture such variations we trained a shape prior based on synthetic data.

We use a box which is centered at the origin and rotate it around the z-axis. For the training data we use 32 equally spaced angles. Our shape prior can now reconstruct this box with an fully random rotation angle around the z-axis. This is demonstrated by generating a box using a random rotation angle as test data. We rendered eight depth maps where we only keep 4% of the depth values and add Gaussian distributed noise to the remaining depth values.

For all our experiments we use standard volumetric fusion [19] as a baseline method. The data costs for the result with and without shape prior are identical up to a scaling factor which is necessary to account for the different regularization.

The results of the experiments are shown in Fig. 5. Not using a shape prior we cannot reconstruct the box but using our shape prior the box is nicely reconstructed for an arbitrary rotation angle.

In our experiments using real-world data we computed a shape prior for two classes of objects, bottles and cars. Both of the classes share the difficulty that objects are highly reflective and contain transparent elements such as the windows of cars. We trained our priors on mesh models downloaded from the Internet; for the bottle prior the training set size was 48, and for the car prior we used 32 instances.

To underline that our shape prior can be used on arbitrary shapes within the trained object class, we took several real world datasets of bottles and cars of varying shapes. For cars we used real cars as well as toy cars. As a first step the camera poses are computed using the publicly available structure from motion software [21]. We then use our plane sweep stereo matching implementation to compute a depth map for each of the images as a reference view. The matching is done on 5 images at once. Zero mean normalized cross correlation (ZNCC) matching scores are computed between each of the images and the reference view and occlusions are handled using only the best $K = 3$ matching scores.

In order to be able to apply the shape prior, the domain Ω needs to be aligned with the bounding box of the object. This can be done based on image based object detectors *e.g.* [8] and optimization over all the input images [2, 10, 13]. As our work focuses on the actual shape prior formulation we omit this step and annotate the bounding box manually.

In Fig. 6 we present the results for the bottle datasets. The same shape prior was used for all the results. Due to the specular, reflective and translucent areas, an isotopic prior, as used in the baseline, is not faithfully modeling the shapes of the bottles. Our shape prior is general enough to recover

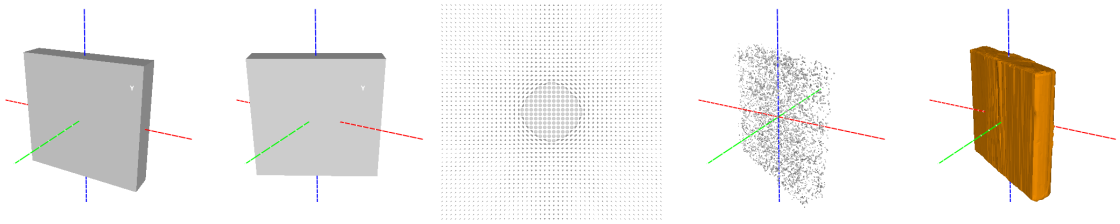


Figure 5. Synthetic experiment. First two images, two out of 32 training shapes. Third image, center part of a horizontal cut through the shape prior. Last two images the reconstruction without and with shape prior, respectively. In the case of not using a shape prior stronger regularization removes the reconstruction entirely.

Figure 6. Reconstructions of the bottle datasets. From left to right: cropped input image, cropped depth map, baseline result [19] and our result using the shape prior with the object and ground segmented. Number of images used, from left to right and top to bottom: 28, 21, 27 and 26.

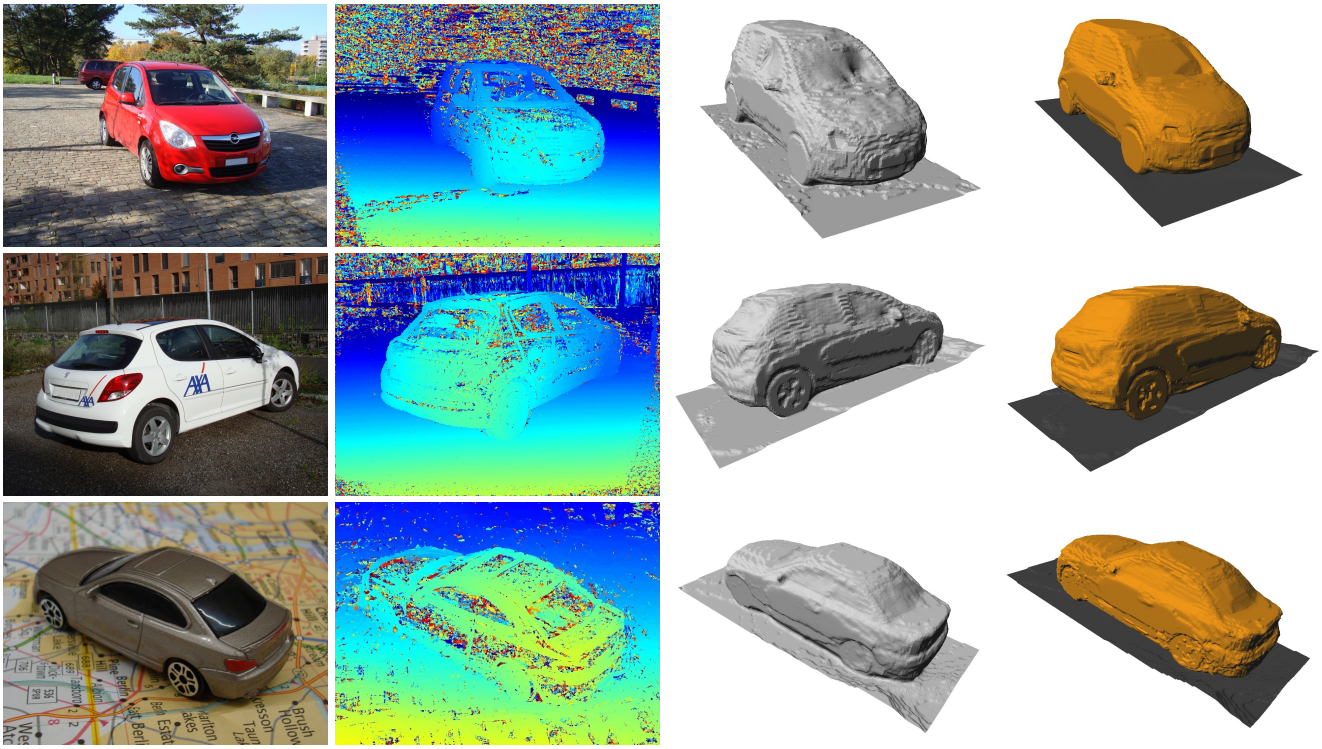


Figure 7. Reconstructions of the car datasets. On the first and second row real cars reconstructed using 80 and 62 images, respectively and on the third row a toy car reconstructed using 27 images. From left to right, cropped input image, cropped depth map, baseline result [19] and our result using the shape prior with the object and ground segmented.

the different shapes present in the input data. Also note that the actual bottle and the ground are segmented through our three label formulation.

Our results on the the car datasets are depicted in Figs. 1 and 7. While a toy car already gives a faithful result in the baseline, applying our shape prior does not change much on the actual shape of the car. Note that details such as the side mirrors and the door handles are present even when utilizing the shape prior. This illustrates that our shape prior is able to allow for a higher variation of normals where they are expected. Additionally, applying the shape prior allows us to segment the ground from the car. Real cars are more challenging then toy cars. Through the transparent windows details in the interior are present in the depth maps, leading to defects on the front wind-shield and the side windows. These issues are fixed by applying our shape prior.

5. Conclusion

We present an approach to include shape priors in convex multi-label segmentation for volumetric 3D reconstruction. It is driven by using spatially varying anisotropic smoothness priors to guide the optimization to the right shape. In order to have a general enough parameterization for anisotropic smoothness we define a discrete Wulff shape, which is formed by the intersection of a discrete number of half spaces. An efficient approach to train the parameters of the discrete Wulff shapes is used to derive the prior from given training data.

On synthetic data we show that our shape prior is capable of capturing big shape variations. Furthermore, we demonstrate the applicability of our formulation on two challenging real world object classes. The prior is able to generalize to different shapes present within a class of objects. We underline this, by showing reconstructions for several examples of each category. Compared to a state-of-the-art baseline not using a shape prior we are able to improve the reconstructions considerably.

Future work needs to examine the possibility of applying such shape priors within reconstructions of larger scenes also containing the surrounding of the object, or scenes with multiple objects of possibly different classes. One step in this direction has been made already in this work by segmenting the ground apart. Also the semantic properties could be investigated further, *e.g.* in the case of a reconstructed car the assigned semantic class to the ground below it is expected to be road.

Acknowledgements: We acknowledge the support of the 4DVideo ERC starting grant #210806 and the V-Charge grant #269916 both under the EC's FP7/2007-2013.

References

[1] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.

[2] S. Y. Bao and S. Savarese. Semantic structure from motion. In *Proc. CVPR*, 2011.

[3] Y. Bao, M. Chandraker, Y. Lin, and S. Savarese. Dense object reconstruction with semantic priors. In *Proc. CVPR*, 2013.

[4] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIIMS*, 5(4), 2012.

[5] A. Dame, V. A. Prisacariu, C. Y. Ren, and I. Reid. Dense reconstruction using 3d object shape priors. In *Proc. CVPR*, 2013.

[6] M. De Berg, M. Van Kreveld, M. Overmars, and O. C. Schwarzkopf. *Computational geometry*. 2000.

[7] S. Esedoḡlu and S. J. Osher. Decomposition of images by the anisotropic rudin-osher-fatemi model. *Comm. PAM*, 57(12), 2004.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Trans. PAMI*, 32(9), 2010.

[9] C. Häne, C. Zach, A. Cohen, R. Angst, and M. Pollefeys. Joint 3d scene reconstruction and class segmentation. In *Proc. CVPR*, 2013.

[10] M. Hödlmoser, B. Micusik, M.-Y. Liu, M. Pollefeys, and M. Kampel. Classification and pose estimation of vehicles in videos by 3d modeling within discrete-continuous optimization. In *Proc. 3DIMPVT*, 2012.

[11] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *Proc. ECCV*, 2010.

[12] V. Lempitsky and Y. Boykov. Global optimization for shape fitting. In *Proc. CVPR*, 2007.

[13] J. Liebelt and C. Schmid. Multi-view object class detection with a 3d geometric model. In *Proc. CVPR*, 2010.

[14] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *Proc. ICCV*, 2011.

[15] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. CVPR*, 2006.

[16] E. Strekalovskiy and D. Cremers. Generalized ordering constraints for multilabel optimization. In *Proc. ICCV*, 2011.

[17] J. Stühmer, P. Schröder, and D. Cremers. Tree shape priors with connectivity constraints using convex relaxation on general graphs. In *Proc. ICCV*, 2013.

[18] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *Proc. CVPR*, 2008.

[19] C. Zach. Fast and high quality fusion of depth maps. In *Proc. 3DPVT*, 2008.

[20] C. Zach, C. Häne, and M. Pollefeys. What is optimized in convex relaxations for multi-label problems: Connecting discrete and continuously-inspired map inference. *Trans. PAMI*, 2014.

[21] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In *Proc. CVPR*, 2010.

[22] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust tv-l^1 range image integration. In *Proc. ICCV*, 2007.

[23] C. Zach, L. Shan, and M. Niethammer. Globally optimal finsler active contours. In *Proc. DAGM*, 2009.