Segment Based 3D Object Shape Priors

Rabeeh Karimi Mahabadi, Christian Häne, Marc Pollefeys
Department of Computer Science, ETH Zürich, Switzerland

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 a prior knowledge about the shape [1]. A shape prior for a given object class is defined by having a spatially varying 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. Often these segments correspond to semantic classes and hence we get a semantic segmentation as a side product of our approach.

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 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.

The idea of our shape prior formulation is coming from the study of equilibrium shapes of crystals [3]. 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.

We use a general convex multi-label optimization framework from [5]. 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 (for a table there are ground, leg, and tabletop). 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.

$$
\begin{aligned}
&E(x) = \sum_{s\in\Omega}\left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j}\phi^{ij}(x_s^{ij} - x_s^{ji})\right)\\
&\text{subject to } \; x_s^i = \sum_j (x_s^{ij})_k, \quad x_s^i = \sum_j (x_{s-e_k}^{ji})_k \qquad (1)\\
&x_s^i \geq 0, \quad \sum_i x_s^i = 1, \quad x_s^{ij} \geq 0
\end{aligned}
$$

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 [2]. 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 set of constraints links the different variables together.

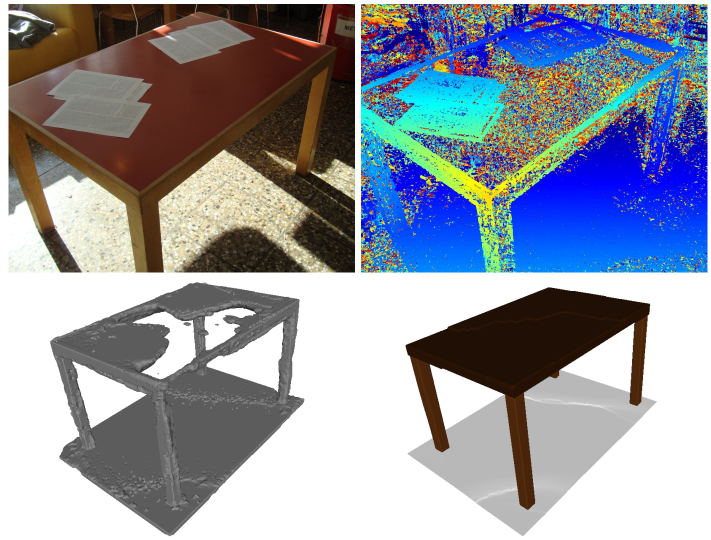

Figure 1: Top: Example input image and depth map. Bottom: Standard volumetric fusion result (left) and our result using the proposed segment based shape prior (right).

We show how the segment based anisotropic smoothness terms ϕ^{ij} are defined and used in the above formulation for several real-world object classes (table, tree, dumbbell and mug) and give a method for inferring the anisotropy from training data. Furthermore, we also demonstrate that parts of the free space can be understood as a convex segment in order to reconstruct concavities such as the inside of a mug. In our results we show qualitative improvements over a baseline approach [4] using standard isotropic regularization of the surface area.

[1] Christian Häne, Nikolay Savinov, and Marc Pollefeys. Class specific 3d object shape priors using surface normals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

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

[3] Geogres Wulff. Zur frage der geschwindigkeit des wachstums und der auflösung der krystallflächen. *Zeitschrift für Krystallographie und Mineralogie*, 1901.

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

[5] Christopher Zach, Christian Häne, and Marc 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.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.