Direction Matters: Depth Estimation with a Surface Normal Classifier

Christian Häne, L'ubor Ladický , Marc Pollefeys
Department of Computer Science, ETH Zürich, Switzerland

The problem of finding a dense disparity map from a stereo rectified image pair is well studied in the computer vision literature. Despite that, in real-world situations, where images contain noise and reflections, it is still a hard problem.

The advances in machine learning approaches have lead to classifiers that are able to estimate surface orientation based on a single image [4]. We argue that information about the surface orientation that is extracted from the input image gives additional important cues about the geometry, exactly in the cases where standard stereo matching algorithms struggle. Therefore we propose a global optimization approach that allows for combining responses of a surface normal direction classifier with matching scores from binocular stereo. Our algorithm is not limited to binocular stereo matching. It can also be applied on scores from a classifier for the single view depth estimation problem [3].

Adding the surface normal directions into a global optimization framework, addresses the problems with standard approaches in stereo matching. In homogeneous areas, such as walls, or on the reflective ground the surface normal directions can often be estimated reliably and hence constrain the depth estimation problem to the desired solution. An important feature of our method is that it is not restricted to use a single surface normal direction per pixel but allows the inclusion of the scores from multiple directions, which is important when the classifier is not able to reliably decide on a specific direction. An example result of our method for the single view depth estimation problem is depicted in Figure 1.

Our method builds on top of the idea of lifting the problem of assigning a depth to each pixel to a volumetric one, where each element of the volume gets assigned whether it is before or after the depth. Using a graph-cut or a convex continuous formulation, the surface attains a globally optimal solution [5, 6]. Depth discontinuities often correspond to image edges, [6] propose to use the anisotropic total variation [1] to align depth discontinuities with image edges. We propose to extend this anisotropic penalization to also include the information from the surface normal direction classifier [3] and hence make a surface direction that has a high likelihood based on the classifier less costly in our energy formulation.

More formally, the goal is to assign to each pixel (r,s) from a rectangular domain $\mathcal{I} = \mathcal{W} \times \mathcal{H}$ a label $\ell_{(r,s)} \in \mathcal{L} = \{0,\dots,L\}$. Instead of assigning labels to pixels directly an indicator variable $u_{(r,s,t)} \in [0,1]$ for each $(r,s,t) \in \Omega = \mathcal{I} \times \mathcal{L}$ is introduced. Using the definition

$$u_{(r,s,t)} = \begin{cases} 0 & \text{if } \ell_{(r,s)} < t \\ 1 & \text{else,} \end{cases} \tag{1}$$

the problem of assigning a label to each pixel is transformed to finding the surface through Ω that segments the volume into an area in front of and behind of the assigned depth. Adding regularization and constraints on the boundary allow us to state the label assignment problem as a convex minimization problem [6], which can be solved globally optimally.

$$\begin{aligned} E(u) = \sum_{r,s,t} \left\{ \rho_{(r,s,t)} \left| (\nabla_t u)_{(r,s,t)} \right| + \phi_{(r,s,t)} \left((\nabla u)_{(r,s,t)} \right) \right\} \\ \text{s.t. } u_{(r,s,0)} = 0 \quad u_{(r,s,L)} = 1 \quad \forall (r,s) \end{aligned} \tag{2}$$

The values $\rho_{(r,s,t)}$ are the data costs or also called unary potential, for assigning label t to pixel (r,s), they for example originate from binocular stereo matching. With the symbol ∇_t we denote the derivative along the label dimension t, and ∇ denotes full 3 component gradient. In both cases we use a forward difference discretization. The regularizer $\phi_{(r,s,t)}$ can be any convex positively 1-homogeneous function. This term allows for an anisotropic penalization of the surface area of the cut surface. The main novelty of our algorithm is the use of a normal direction classifier to define the anisotropic regularization term. The boundary constraints on u enforce that there is a cut through the volume.

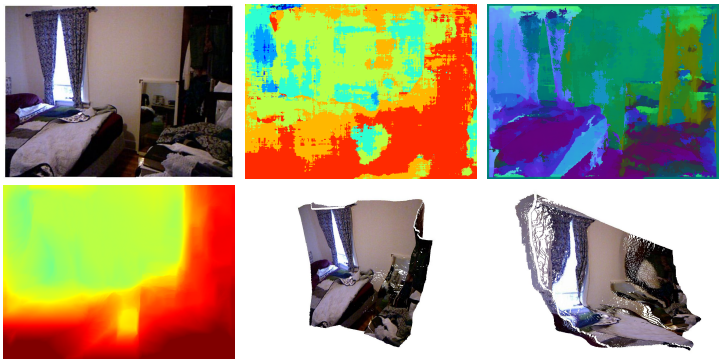

Figure 1: Overview of our method. Top Row: The input to our method is depicted in the top row. On a single input image (left) two classifiers are evaluated, single view depth estimation (middle) and surface normal directions (right). Bottom Row: On the bottom the obtained depth map by our surface normal direction based regularization is shown (left) together with two renderings of the obtained dense point cloud (middle and right).

In order to define the anisotropic smoothness term we follow the approach, that any convex positively 1-homgenoous function ϕ can be defined in terms of a convex shape [1].

$$\phi_{\mathcal{W}}(\nabla u) = \max_{p \in \mathcal{W}} p^T \nabla u, \tag{3}$$

where $\mathcal{W}$ is a convex, closed and bounded set that contains the origin, the so-called Wulff shape. Furthermore, we follow the apporach of [2], which discretizes the space of normal directions to map them to a Wulff shape which is formed as an intersection of half-spaces. This way of defining the regularizer nicely works together with the surface normal direction classifier [4], as already the classifier outputs scores for a discrete set of normals, which is defined during the training of the classifier.

In our experiments we show that we improve over a baseline approach, without using the surface normals, for both, stereo matching and single view depth estimation.

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

[2] 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.

[3] Lubor Ladicky, Jianbo Shi, and Marc Pollefeys. Pulling things out of perspective. In *Conference of Computer Vision and Pattern Recognition (CVPR)*, 2014.

[4] Lubor Ladicky, Bernhard Zeisl, and Marc Pollefeys. Discriminatively trained dense surface normal estimation. In *European Conference on Computer Vision (ECCV)*, 2014.

[5] Sébastien Roy and Ingemar J Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *International Conference on Computer Vision (ICCV)*, 1998.

[6] Christopher Zach, Marc Niethammer, and Jan-Michael Frahm. Continuous maximal flows and wulff shapes: Application to mrfs. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

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