Variational Regularization and Fusion of Surface Normal Maps

Bernhard Zeisl
Computer Vision and Geometry Group
ETH Zurich, Switzerland

Christopher Zach
Toshiba Research Europe
Cambridge, UK

Marc Pollefeys
Computer Vision and Geometry Group
ETH Zurich, Switzerland

Abstract—In this work we propose an optimization scheme for variational, vectorial denoising and fusion of surface normal maps. These are common outputs of shape from shading, photometric stereo or single image reconstruction methods, but tend to be noisy and request post-processing for further usage. Processing of normals maps, which do not provide knowledge about the underlying scene depth, is complicated due to their unit length constraint which renders the optimization non-linear and non-convex. The presented approach builds upon a linearization of the constraint to obtain a convex relaxation, while guaranteeing convergence. Experimental results demonstrate that our algorithm generates more consistent representations from estimated and potentially complementary normal maps.

I. INTRODUCTION

Noisy vectorial images representing surface normals appear in several computer vision applications, and subsequent processing of normal fields relies on cleaner estimates with much of the noise removed. While denoising of standard single-channel intensity or multi-channel color images and also depth maps received a lot of interest in the literature (e.g. [1]), removing noise from surface normals is usually addressed via mesh processing on surfaces (e.g. [2]) but typically not by image denoising applied on normal maps. However, in many applications a surface normal is estimated per pixel and a natural representation is based on normal maps, i.e. a mapping from a regular image grid to surface normals, and without knowledge of the actual 3D scene geometry.

Noisy normal maps are the intermediate result of shape-from-shading [3], [4] and photometric stereo techniques (e.g. [5], [6]). Further, surface orientations are the natural output of single-image relative depth estimation [7], [8] and direct regression of normals from a single image [9], [10], [11]; a research direction which has gained quite some interest recently [12], also because the estimated, coarse information about the underlying scene geometry has shown to help other computer vision tasks such as object detection, semantic reasoning or scene understanding. In addition to the denoising task, regularized fusion of normal maps is important if several complementary estimates of surface normals are given, which should be merged to return one consistent normal map. In contrast to multi-channel image denoising [1], where the unknown vector per pixel is either unconstrained or has simple box constraints (max. and min. value), processing of surface normals is more difficult due to the unit length constraint on normals. This non-linear and non-convex constraint means that a global minimum of a denoising objective over surface normals is hard to find.

In this work we propose a vectorial optimization and demonstrate that by linearizing the constraint we can guarantee a decrease in the objective in each iteration — and consequently that our optimization method is convergent. The observation that man-made environments exhibit predominantly piece-wise planar geometries is incorporated as a total variation based regularization, which favors piece-wise constant normals (or piece-wise smooth normals, as discussed in Sec. II).

The remainder of our manuscript is structured as follows. In Sec. II we state our problem formulation and show how to cope with the norm constraint on normals. Then Sec. III details our iterative optimization approach, while particular considerations for 3D surface normals and the implementation itself are discussed in Sec. IV and V, respectively. Finally, experiments and their results are presented in Sec. VI.
II. DENOISING OF ORIENTATION FIELDS

We will first present our algorithm for the general n-dimensional case; i.e. the optimization of surface normals of hyperplanes, parameterized as points lying on the respective hyperplane. In Sec. IV we will then concentrate on the important case of 3D surface normals and point out related geometric properties.

A normal vector could be efficiently represented in spherical coordinates; however, this choice of parameterization leads to non-linearities in any optimization procedure, due to the periodicity of the angular measure. We take a different approach and chose an (over-)parameterization in vector space. Such vector fields e.g. occur as 2D displacement fields in optical flow or multi-channel color values. However, surface normals are constrained to be of unit length, which is a difficult non-convex constraint. In the following we will linearize this constraint and demonstrate the resulting convex formulation is a majorizer of the original non-convex objective. Thus, the majorize-minimize principle [13] can be applied to determine an (approximate) minimizer. Via this relaxed representation we implicitly obtain a minimal parameterization of our optimization problem.

Let us start by stating the original objective function to minimize in n-dimensional space. Given an image domain Ω we search for surface normals n which minimize the objective function

\[
E(n) = \int_{\Omega} \lambda \sum_i \phi_l(x, n(x)) + \sum_{d \in \{h,v\}} w_d(x) \psi(∥\nabla_d n(x)∥) \, dx \quad \text{s.t. } ||n(x)||_2 = 1,
\]

where \{h, v\} denote the horizontal and vertical direction in the image domain. We use a (weighted) anisotropic regularizer, \[ \int \sum_d w_d(x) \psi(∥\nabla_d n(x)∥) \, dx \quad \text{where } \psi : R^+ \rightarrow R \]

is a monotonically increasing convex function. The weights \( w_d \geq 0 \) are induced by texture gradients — which serve as strong cues for geometric edges — of the underlying image. By this we aim for a discontinuity preserving, image-driven regularization. The weighting has the effect to facilitate regularization among homogeneous image regions, but suppresses regularization along texture gradients. The data term \( \phi_l : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R} \) is discussed below.

In Eq. 1 we explicitly stated the dependence of functions and unknowns on the image location \( x \), but for clarity we will drop the respective argument in the following. We will also use \( n \) etc. to denote both, a single normal or an entire normal map, and the meaning should be clear from the context. We abbreviate the regularization part of Eq. 1 by

\[
R(n) \overset{\text{def}}{=} \int_{\Omega} \sum_{d \in \{h,v\}} w_d \psi(∥\nabla_d n∥) \, dx. (2)
\]

Our choice of regularization is motivated by the assumption that we expect piecewise smooth normal fields. A convex regularizer favoring such piecewise smooth solutions is given by a “smoothed” total variation, where the non-differentiable norm is replaced by the Huber cost [14],

\[
\psi(∥\nabla_d n∥) = \begin{cases}
\frac{1}{2} ||\nabla_d n||^2 & ||\nabla_d n|| < \epsilon \\
\frac{1}{2} \epsilon & ||\nabla_d n|| = \frac{1}{2} \epsilon \\
\frac{1}{2} \epsilon & \text{else}.
\end{cases}
\]

Clearly, this regularizer approaches the total variation for \( \epsilon \rightarrow 0 \). Thus for generality in the remainder of this work we will just consider the Huber norm, but piecewise constant regularization behavior is achieved by setting \( \epsilon = 0 \) (cf. Fig. 2 for an example of 2D normal field denoising).

The data terms \( \phi_l(x, n) \) measure the data fidelity between measured observations and optimized normals \( n \), and the exact form of \( \phi_l \) depends on the assumed noise model for the observations, but in this work we assume convex \( \phi_l \). We allow several data terms per pixel (indexed by \( l \)) in order to allow several estimates of normals to be merged. Suitable choices in our application are the squared and unsquared \( L_2 \) distance between the observation \( f_l \) and the unknown \( n \).

\[
(i) \quad \phi_l(x, n) = \frac{w_l(x)}{2} ||n - f_l(x)||^2_2 \quad \text{or} \quad \\
(ii) \quad \phi_l(x, n) = w_l(x)||n - f_l(x)||_2.
\]

The unsquared distance is known to be more robust to outlier measurements due to the heavier than Gaussian tails in the underlying noise model. It has been used e.g. in image [15]

Figure 2: Denoising of a 2D normal map. (a) Ground truth orientation field. (b) Same orientation field perturbed by noise; i.e. \( n' = d/∥d∥_2 \). \( d = n + (r_x, r_y)^T \), where \( r_i \) are random variables are from \( N(0,0.4) \). (c) Denoised result with TV regularization. (d) Denoised result with Huber-TV regularization. Note the smoother transition of direction vectors compared to (c), e.g. in the highlighted region. For both denoised results the weight \( \lambda \) of the data term was set to 0.5.
and shape denoising [16]. For normalized vectors it holds that

\[ \frac{w_l}{2} \| \mathbf{n} - \mathbf{f}_l \|_2^2 = w_l(1 - \mathbf{n}^T \mathbf{f}_l). \] (5)

Therefore the squared $L_2$ distance equals the (weighted) cosine distance. If we regard the data fidelity in terms of the angular error $\theta = \arccos(\mathbf{n}^T \mathbf{f}_l)$, Fig. 3 illustrates that our considered choices of error measurements are robust cost functions wrt. the angular deviation. In the remainder of this work we will focus on $\phi_l(\mathbf{n}) = w_l \| \mathbf{n} - \mathbf{f}_l \|_2$ due to the increased robustness. With our choice of convex $\phi_l$ the only source of non-convexity in the overall formulation comes from the pixel-wise unit-length constraint on the normals, which is addressed below.

A. Linearization of norm constraint

As pointed out previously, the objective function of Eq. (1) is a non-convex optimization problem, solely due to the non-convex constraint on the normal vectors. One can obtain an obvious convex relaxation by replacing the non-convex set $\{ \mathbf{n} : \| \mathbf{n} \|_2 = 1 \}$ by its convex hull, $\{ \mathbf{n} : \| \mathbf{n} \|_2 \leq 1 \}$, but such an approach would suffer from a strong tendency to shrink the unknown normals due to the regularization term.

As an alternative we propose to use the tangent plane approximation of the non-linear manifold, $\{ \mathbf{n} : \| \mathbf{n} \|_2 = 1 \}$, and thus the non-linear constraint is replaced by a linear one. The tangent plane at a linearization point $\mathbf{n}_0$ is given by

\[ \| \mathbf{n} \|_2^2 \approx \mathbf{n}_0^T \mathbf{n} + 2 \mathbf{n}_0^T (\mathbf{n} - \mathbf{n}_0) = 1. \] (6)

Since the unit sphere is a smooth manifold, the resulting tangent plane equation $\mathbf{n}_0^T \mathbf{n} = 1$ gives a very good approximation of the original constraint near $\mathbf{n}_0$. Fig. 4 illustrates the linearization for the 2D case. Effectively we have transformed the feasible set from a hyper-sphere to the hyper- (tangent-) plane.

It allows to perform the optimization in a local parameterization within the tangent space $\mathcal{T} \triangleq \{ \mathbf{n} : \mathbf{n}_0^T \mathbf{n} = 1 \}$. The orthogonal complement to the working point $\mathbf{n}_0$ builds a basis $\mathbf{B} \in \mathbb{R}^{n \times n-1}$ for $\mathcal{T}$. It corresponds to the null space of $\mathbf{n}_0^T$, i.e. $\mathbf{B} = \text{null}(\mathbf{n}_0^T)$. Consequently the minimal parameterization for surface normals under a linearized constraint is given by $\mathbf{s} \in \mathbb{R}^{n-1}$, where $\mathbf{s}$ lives in a n-1 dimension subspace. The needed transformation from local parameterization to global normals and vice versa is given by the linear maps

\[ \mathbf{n} = h(\mathbf{s}) = \mathbf{B} \mathbf{s} + \mathbf{n}_0 \quad \text{and} \quad \mathbf{s} = h^{-1}(\mathbf{n}) = \mathbf{B}^T(\mathbf{n} - \mathbf{n}_0). \] (7)

For the 2 and 3 dimensional case a basis $\mathbf{B}$ is directly computed via orthogonal vector and cross product expansion, respectively. In the n-dimension case the Gram-Schmidt process is an efficient algorithm compared to a full singular value decomposition to retrieve the null-space.

What remains is our transformed linear, convex but non-smooth objective function

\[ \min_{\mathbf{s}} \int_{\Omega} \lambda \sum_l \phi_l(h(\mathbf{s})) \, dx + \mathcal{R}(h(\mathbf{s})). \] (8)

which is now optimized within the coordinates $\mathbf{s}$ in the tangent space $\mathcal{T}$ of the current valid linearization.

B. Majorize-minimize method framework

According to standards in non-linear optimization with non-linear constraints, we also solve a sequence of surrogate problems in order to minimize the original objective. In contrast to very general approaches such as sequential quadratic programming, we have the guarantee of non-increasing the objective by solving the convex surrogate problem. This is the consequence of

\[ \bar{E}(\mathbf{n}) \triangleq \int_{\Omega} \lambda \sum_l \| \mathbf{n} - \mathbf{f}_l \|_2 \, dx + \mathcal{R}(\mathbf{n}) \quad \text{s.t.} \quad \mathbf{n}^T \mathbf{n}_0 = 1 \] (9)

being a majorizer [13] of our objective

\[ E(\mathbf{n}) \triangleq \int_{\Omega} \lambda \sum_l \| \mathbf{n} - \mathbf{f}_l \|_2 \, dx + \mathcal{R}(\mathbf{n}) \quad \text{s.t.} \quad \| \mathbf{n} \|_2 = 1 \] (10)

The authors in [17] propose to obtain a minimal parameterization for 3D normals by considering only 2 of the components. However, we claim that such a representation leads to a non-uniform, data-dependent regularization, because the resulting gradient and thus the regularization strength correlates with the normal direction – which is an undesired property.
This can be seen by noting that (i) the tangent plane \( \{ \mathbf{n} : \mathbf{n}^T \mathbf{n} = 1 \} \) is outside the unit sphere (which means that the projection into the set \( \{ \mathbf{n} : \|\mathbf{n}\|_2 \leq 1 \} \) is always on its boundary \( \{ \mathbf{n} : \|\mathbf{n}\|_2 = 1 \} \), and (ii) that in general the projection into a convex set is non-expansive. More formally, define the sets
\[
\mathcal{B} \equiv \{ \mathbf{n} : \|\mathbf{n}\|_2 \leq 1 \} \quad \text{and} \quad \partial \mathcal{B} \equiv \{ \mathbf{n} : \|\mathbf{n}\|_2 = 1 \},
\]
and the projection into \( \mathcal{B} \)
\[
\Pi_{\mathcal{B}}(\mathbf{n}) = \arg \min_{\mathbf{n} \in \mathcal{B}} \|\mathbf{n} - \tilde{\mathbf{n}}\|_2 = \begin{cases} \tilde{\mathbf{n}} & \text{if } \|\mathbf{n}\|_2 \leq 1 \\ \tilde{\mathbf{n}}/\|\tilde{\mathbf{n}}\|_2 & \text{otherwise.} \end{cases}
\]
It is well-known that projections into convex sets are (firmly) non-expansive (see e.g. [18]), hence \( \Pi_{\mathcal{B}} \) is non-expansive due to the convexity of the set \( \mathcal{B} \). Thus, we have
\[
\|\Pi_{\mathcal{B}}(\mathbf{n}) - \Pi_{\mathcal{B}}(\hat{\mathbf{n}})\| \leq \|\mathbf{n} - \hat{\mathbf{n}}\| \quad \text{and} \quad f(\|\Pi_{\mathcal{B}}(\mathbf{n}) - \Pi_{\mathcal{B}}(\hat{\mathbf{n}})\|) \leq f(\|\mathbf{n} - \hat{\mathbf{n}}\|)
\]
for any monotonically increasing function \( f \). Therefore, we have for the discretized gradient
\[
\psi(\|\nabla_d \mathbf{n}\|) = \psi(\|\hat{\mathbf{n}}(x) - \hat{\mathbf{n}}(x')\|) \\
\geq \psi(\|\Pi_{\mathcal{B}}(\mathbf{n}(x)) - \Pi_{\mathcal{B}}(\hat{\mathbf{n}}(x'))\|) \quad (11)
\]
for suitable neighboring pixels \( x \) and \( x' \) and our choice of a monotonically increasing \( \psi(\cdot) \) in Eq. 3. For the \( L_2 \) data fidelity term we obtain analogously,
\[
\|\hat{\mathbf{n}}(x) - \mathbf{f}_i(x)\|_2 \geq \|\Pi_{\mathcal{B}}(\mathbf{n}(x)) - \Pi_{\mathcal{B}}(\mathbf{f}_i(x))\|_2 \\
= \|\Pi_{\mathcal{B}}(\mathbf{n}(x)) - \mathbf{f}_i(x)\|_2, \quad (12)
\]
since the noisy observation \( \mathbf{f}_i(x) \) has unit length.

Let \( \hat{\mathbf{n}}^* \) be a minimizer of our relaxed objective function in Eq. 9 and \( \mathbf{n}^* = \Pi_{\mathcal{B}}(\hat{\mathbf{n}}^*) \) (its pixel-wise projection). By applying the above fact term-wise, we see that inserting \( \mathbf{n}^* \) into the original objective of Eq. 10 has a lower or equal value than the objective in Eq. 9 at \( \hat{\mathbf{n}}^* \), and the objective values coincide at \( \mathbf{n}_0 \). Overall, by repeatedly minimizing Eq. 9 and subsequent linearization of the norm constraint we obtain a non-increasing sequence of objective values (which is converging, since the energy is always non-negative i.e. bounded from below).

### III. Optimization

For optimization of our objective function in Eq. (8) we build upon the primal-dual algorithm, proposed in [19]. In order to apply it the primal problem needs to be transformed into a convex-concave saddle-point problem, which can be shown (using convex conjugacy [20]) to have the form
\[
\min_{\mathbf{s}} \int_\Omega \sum_{d \in \{h,v\}} \max_{\mathbf{p}_d} \left\{ \langle \mathbf{p}_d, \nabla_d h(s) \rangle - \frac{w_d}{2} \|\mathbf{p}_d\|_2^2 \right\} \\
+ \sum_i \max_{\mathbf{q}_i} \langle \mathbf{q}_i, (h(s) - \mathbf{f}_i) \rangle \ dx \quad (13)
\]
\[
= \min_{\mathbf{s}} \int_\Omega \sum_{d \in \{h,v\}} \max_{\mathbf{p}_d} \left\{ \langle -\nabla_d^T \mathbf{p}_d, \mathbf{B} \mathbf{s} + \mathbf{n}_0 \rangle - \frac{w_d}{2} \|\mathbf{p}_d\|_2^2 \right\} \\
+ \sum_i \max_{\mathbf{q}_i} \langle \mathbf{q}_i, (\mathbf{B} \mathbf{s} + \mathbf{n}_0 - \mathbf{f}_i) \rangle \ dx \quad (14)
\]
subject to constraints \( \|\mathbf{p}_d\| \leq w_d \) and \( \|\mathbf{q}_i\|_2 \leq \lambda w_l \) (applied pixel-wise). The optimization follows an iterative procedure exhibiting gradient descent steps for the primal variables and gradient ascent steps for the dual variables followed by appropriate projection steps. As such in each iteration the dual variables are updated via
\[
\mathbf{p}_d^{k+1} = \mathcal{P}_P \left( \mathbf{p}_d^k + \sigma \nabla_d h(s) \right) \\
\mathbf{q}_i^{k+1} = \mathcal{P}_Q \left( \mathbf{q}_i^k + \sigma (h(s) - \mathbf{f}_i) \right), \quad (15)
\]
where \( \sigma \) denotes the dual time step. The projection operators for the dual variables are defined as
\[
\mathcal{P}_P(\hat{\mathbf{p}}_d) = \frac{\hat{\mathbf{p}}_d/(1 + \sigma \epsilon)}{\max\{1, \|\hat{\mathbf{p}}_d\|_2/w_d\}} \\
\mathcal{P}_Q(\hat{\mathbf{q}}_i) = \frac{\hat{\mathbf{q}}_i}{\max\{1, \|\hat{\mathbf{q}}_i\|_2/\lambda w_l\}}. \quad (16)
\]
The update of the primal variables follows the gradients of regularization and data terms to
\[
s^{k+1} = s^k + \tau \mathbf{B}^T \left( \sum_{d \in \{h,v\}} \nabla_d^T \mathbf{p}_d - \sum_i \mathbf{q}_i \right) \quad (17)
\]
where \( \mathbf{B} \) is the primal step size. The parametrization in tangent space has the effect that both sets of dual variables are inversely mapped to the local coordinate system of \( s \). The algorithm maintains an overrelaxation of the primal variables in addition to \( s^k \), which are updated via \( s^{k+1} = 2s^{k+1} - s^k \), and \( \mathbf{n}^{k+1} = \hat{h}(s^{k+1}) \) respectively. This extrapolation step from the previous iteration enables faster convergence of the procedure.

For a given linearization point, and thus current local parameterization, the algorithm is run for a few iterations until convergence. Then we apply a projection
\[
\mathbf{n}_0 \leftarrow \mathbf{n}_0^{k+1} = h(s^{k+1})/\|h(s^{k+1})\|_2 \quad (18)
\]
on the hypersphere such that the original norm-constraint is fulfilled. As indicated this result then serves as the new working point \( \mathbf{n}_0 \) and a new linearization, local parameterization and orthogonal complement for the tangent space are computed. The overall optimization is run until global convergence.
IV. APPLICATION TO 3D SURFACE NORMALS

In the 3-dimensional case we are able consider visibility information and by this constraint the feasible set of possible normal directions. Without loss of generality let us assume that surface normals are defined in the local camera coordinate system and that they point towards the camera. For a given image location \( x = (x, y) \) the point ray originating from the camera center and passing through \( x \) is defined as \( r = K^{-1}(x, y, 1)^T \), where \( K \) defines the camera intrinsics. This setup limits surface normals to lie on a half-sphere, since only surfaces which are oriented towards the camera are visible. As a result the constraint

\[
    n^T r \leq 0 \quad \forall x \in \Omega \quad (19)
\]

must be fulfilled for each image location individually.

One can utilize this prior information as a preprocessing step for the orientation field. It is encoded via the data term weights, i.e. \( w_1 = 1 \) if the observed surface normal is visible and 0 otherwise. As a result the optimization procedure will fill in the missing values. Fig. 5(b) illustrates an exemplary mask for infeasible normals.

Visibility consistent regularization: The previous preprocessing makes a hard decision about feasible normals. If the constraint is just slightly violated due to noise in the normals — but the direction is assumed to be largely correct — we can encode the visibility constraint of Eq. (19) directly in the optimization. It has the effect that the observation still exhibits a driving force, but the optimized normals are guaranteed feasible, i.e. visible. In order to achieve this effect the iterative optimization procedure of Sec. III just needs to be modified slightly: The original primal problem in Eq. (8) and accordingly the primal-dual formulation of Eq. (13) are augmented by the visibility constraint. Within the optimization procedure it is simply handled by an additional projection step. In each iteration after the primal update and projection step of Eq. 18 an additional projection

\[
    n \leftarrow P_N(n) = \begin{cases} 
    n & n^T r \leq 0 \\
    \frac{r \times n \times r}{\|r \times n \times r\|^2} & \text{else} 
    \end{cases} \quad (20)
\]

to the feasible half-sphere, defined by point ray \( r \) for normal \( n \), is performed. The cross products create a 3D vector which is closest to \( n \) and orthogonal to \( r \) and thus on the boundary of the feasible set.

V. IMPLEMENTATION

So far our optimization has been formulated in the continuous domain. For implementation, operation on a discrete image grid is required. In this regard, derivative operators have been discretized as forward differences with reflecting (Neumann) boundary conditions. As a result the divergence operator is based on backward differences with Dirichlet boundary conditions. For algorithm initialization each normal is assigned the respective data term, i.e. \( n = f_0 \). For infeasible 3D normals in the data terms, \( n \) is initialized analogous to Eq. (20) or in case of missing data defined to point towards the camera. In case large portions of missing data are present in the orientation field, we have created a coarse to fine framework for faster convergence.

The algorithm itself has been implemented as a multi-threaded C++ application. Average run-times for the 3D surface normal denoising results in Sec VI (640 × 480 images, 1000 iterations) are around 5 seconds on a current desktop computer with 8 cores. Due to the decomposition into independent per pixel computations for both, the local parameterization and primal and dual update steps, the algorithm would lend itself to highly parallel implementations on a GPU. Image gradients are computed via \( 3 \times 3 \) Sobel filters in horizontal and vertical image direction. In addition we pre-smooth the image with a Gaussian kernel of standard deviation of 1.2 pixels to suppress image noise.

VI. EXPERIMENTS AND RESULTS

A. Denoising Performance

In a first experiment we evaluate the denoising performance of our algorithm. Required ground truth measurements have been obtained from 4 different synthetic 3D models (one shown in Fig. 5). We rendered several depth maps from the models and computed point-wise normals via a least squares kernel within a RANSAC scheme to preserve
Geometry edges. In this way 9 ground truth normal maps have been created for evaluation. Each of them has then been distorted via noise of varying strength, i.e. a noisy normal is computed to \( \mathbf{n}' = \frac{\mathbf{n} + (r_x, r_y, r_z)}{\| \mathbf{d} \|_2} \), where \( \mathbf{d} = \mathbf{n} + (r_x, r_y, r_z)^2 \) and \( r_i \) are Gaussian random variables with \( \sigma \in [0.1, \ldots, 0.9] \).

We compare our results to cross bilateral filtering [21] (which also exhibits edge-preserving properties) adapted to orientation fields. There each normal is replaced by a weighted average of normals from nearby pixels. The weight itself is based on two Gaussian distributions over the Euclidean distance of pixels and the intensity difference of the underlying image, respectively. The corresponding standard deviations where set to \( \sigma_S = \frac{3}{2} w \), where \( w \) is half the window size of the filter, and \( \sigma_I = 0.1 \) for image intensities in the range \([0, 1]\). The weighted average of normals is computed as the extrinsic mean (i.e. the normalized Euclidean mean).

Quantitative performance is given in Fig. 6. We measure the deviation of our denoised results as mean squared error (MSE) between normals on the unit-sphere; as pointed out in Eq. (5) this coincides with the cosine distance. The reported peak-signal-to-noise-ratios are computed to \( PSRN = 10 \log_{10} (\frac{d_{max}^2}{MSE}) \), where \( d_{max} = 2 \) and denotes the maximal error between normals. As can be seen in Fig. 6 the performance of bilateral filtering is quite sensitive to the chosen filter size (reported are the top results among several tested kernel sizes). In contrast our approach achieves superior results with one set of parameters over the analyzed noise range.

B. Regularization and Fusion of Normal Estimates

In a second experiment we explore the task of regularization and fusion of predicted 3D surface normal estimates from a single image. For denoising we utilize state-of-the-art prediction results [10] on the NYU2 [22] dataset. Fig. 7 lists several examples for our denoising results. Initial predictions describe the observed scene geometry well, but are noisy and may contain visually infeasible estimates. The goal of the regularization is to obtain cleaner, more simplistic results — which is accomplished by our algorithm. It produces a consistent labeling within homogeneous regions and follows boundaries well. The parameter settings for this experiments have been \( \lambda = 0.1 \) and \( \epsilon = 0.05 \). The decreased weight on the data term compared to the previous experiment is explained by the weaker quality of observations, requesting stronger regularization. Typical examples for partial infeasible predictions are highlighted in the examples. Our method resolves those areas to more consistent estimates. We do not report a PSNR value, because there are no accurate ground truth normals available for the NYU2 dataset. The training data used in [9] or [10] is quite noisy as well, especially for distant structures. However, we evaluated our results equivalently to [10] to obtain an estimate about the (rough) correctness of normals. Table I lists the scores and indicates that we do well. To ensure fair comparison we also applied bilateral filtering with varying kernel sizes and only report the best performance for \( w = 7 \).

For fusion of normal fields we follow the idea to combine several, potentially complementary predictions to obtain a merged result with increased accuracy. In this regards, Fouhey et al. [9] have presented and interesting work. Their method aims at extracting both visually-discriminative and geometrically-informative shape primitives from RGB-D training data. During inference the learned detector will fire only at sparse positions and reason about the underlying scene geometry by means of the learned shape primitives. Obtained primitives are typically located at 3D corner- or edge-like structures and are a strong cue for the local normal direction. We fuse these sparse predictions with estimates obtained from a boosted classifier, which was trained in a similar way to [10] but without super-pixel features to obtain estimates which do not follow boundaries. Thereby we can demonstrate the applicability of our approach for fusion of normal maps; though, obtained results are likely to be worse than those in Fig 7, simply because initial estimates are less detailed. Still, exemplary fusion results in Fig 8 demonstrate the increased accuracy we obtain over the input estimates. In this regard, we are looking forward to see new, complementary estimation methods being published, which then can be merged by our approach to obtain even more accurate models.
VII. CONCLUSION

In this work we have presented a convex optimization scheme for variational, vectorial regularization and fusion of surface normal maps where the underlying scene depth is unknown. As normals have constant length, optimization on them is non-linear and non-convex. We have introduced a convex relaxation of the original objective function and proven guaranteed convergence. Performed experiments demonstrate that our algorithm is able to generate more con-
Figure 8: Exemplary results for fusion of different normal estimates. “Prediction” denotes pixel-wise predictions from a boosted classifier (similar to [10]) and “3DP” are the detected primitives from [9].

sistent normal maps. It is expected that with the introduction of complementary surface normal estimation methods results will improve further.

Future work will concentrate on adaptive fusion of normal estimates and the use of obtained normals for objected detection and scene understanding or as a prior in 3D modeling.

REFERENCES


608