Sublabel-Accurate Convex Relaxation of Vectorial Multilabel Energies

Emanuel Laude*1, Thomas Möllenhoff*1, Michael Moeller1, Jan Lellmann2, and Daniel Cremers1

1Technical University of Munich  2University of Lübeck
{laudee, moellenh, moellerm}@in.tum.de, lellmann@mic.uni-luebeck.de, cremers@tum.de

Abstract. Convex relaxations of nonconvex multilabel problems have been demonstrated to produce superior (provably optimal or near-optimal) solutions to a variety of classical computer vision problems. Yet, they are of limited practical use as they require a fine discretization of the label space, entailing a huge demand in memory and runtime. In this work, we propose the first sublabel accurate convex relaxation for vectorial multilabel problems. The key idea is that we approximate the dataterm of the vectorial labeling problem in a piecewise convex (rather than piecewise linear) manner. As a result we have a more faithful approximation of the original cost function that provides a meaningful interpretation for the fractional solutions of the relaxed convex problem. In numerous experiments on large-displacement optical flow estimation and on color image denoising we demonstrate that the computed solutions have superior quality while requiring much lower memory and runtime.

Keywords: Convex Relaxation, Optimization, Variational Methods

1 Introduction

1.1 Nonconvex Vectorial Problems

In this paper, we derive a sublabel-accurate convex relaxation for vectorial optimization problems of the form

$$\min_{u: \Omega \rightarrow \Gamma} \int_{\Omega} \rho(x, u(x)) \, dx + \lambda TV(u),$$

* Those authors contributed equally.
Fig. 1: In (a) we show a nonconvex optical flow dataterm energy on a \( 45 \times 45 \) matching window. Convexifying the dataterm \( \rho \) without lifting would optimize the energy shown in (b). Classical lifting methods (c), such as [2], approximate the energy piecewise linearly between the labels. In (d) we show the energy our proposed method optimizes. Due to the separate convexification on each triangle, we are able to capture the structure of the original nonconvex dataterm energy much more accurately than the other approaches (b) and (c). This leads to superior results when combined with a regularizer.

where \( \Omega \subset \mathbb{R}^d \), \( \Gamma \subset \mathbb{R}^n \) and \( \rho : \Omega \times \Gamma \to \mathbb{R} \) denotes a generally nonconvex pointwise dataterm. As regularization we focus on the total variation defined via duality as:

\[
TV(u) = \sup_{q \in C_\infty^\infty(\Omega, \mathbb{R}^{n \times d}), \|q(x)\|_{S\infty} \leq 1} \int_{\Omega} \langle u, \text{Div} \ q \rangle \ dx,
\]

where \( \| \cdot \|_{S\infty} \) is the Schatten-\( \infty \) norm on \( \mathbb{R}^{n \times d} \), i.e., the largest singular value. For differentiable functions \( u \) we can integrate (2) by parts to find

\[
TV(u) = \int_{\Omega} \| \nabla u(x) \|_{S1} \ dx,
\]

where the dual norm \( \| \cdot \|_{S1} \) essentially penalizes Jacobians \( \nabla u \) which have high rank, i.e., the individual components of \( u \) jump in a different direction. This type of regularization is part of the framework of Sapiro and Ringach [1].

1.2 Related Work

Due to its nonconvexity the optimization of (1) is challenging. For the scalar case \( (n = 1) \), Ishikawa [3] proposed a pioneering technique to obtain globally optimal solutions in a spatially discrete setting, given by the minimum s-t-cut of a graph representing the space \( \Omega \times \Gamma \). A continuous formulation was introduced by Pock et al. [4] exhibiting several advantages such as less grid bias and parallelizability.
In a series of papers [5,6], connections of the above approaches were made to the mathematical theory of cartesian currents [7] and the calibration method for the Mumford-Shah functional [8], leading to a generalization of the convex relaxation framework [4] to more general (in particular nonconvex) regularizers.

In the following, researchers have strived to generalize the concept of functional lifting and convex relaxation to the vectorial setting \((n > 1)\). If the dataterm and the regularizer are both separable in the label dimension, one can simply apply the above convex relaxation approach in a channel-wise manner to each component separately. However, when either the dataterm or the regularizer couple the label components, the situation becomes more complex [9,10].

The approach which is most closely related to our work, and which we consider as a baseline method, is the one by Lellmann et al. [2]. They consider coupled dataterms with coupled total variation regularization of the form (2).

A drawback shared by all mentioned papers is that ultimately one has to discretize the label space. While Lellmann et al. [2] propose a sublabel-accurate regularizer, we show that their dataterm leads to solutions which still have a strong bias towards the label grid. For the scalar-valued setting, continuous label spaces have been considered in the MRF community by Zach et al. [11] and Fix et al. [12]. The paper [13] proposes a method for mixed continuous and discrete vectorial label spaces, where everything is derived in the spatially discrete MRF setting. Möllenhoff et al. [14] recently proposed a novel formulation of the scalar-valued case which retains fully continuous label spaces even after discretization. The contribution of this work is to extend [14] to vectorial label spaces, thereby complementing [2] with a sublabel-accurate dataterm.

Many other methods for optimizing nonconvex energies of the form (1) exist, such as coarse-to-fine continuation methods [15,16,17] or quadratic relaxation approaches [18,19]. While these algorithms often provide excellent results in practice, in contrast to convex relaxation methods they are dependent on the initialization and have many tuning parameters without a clear interpretation.

### 1.3 Contribution

In this work we propose the first sublabel-accurate convex formulation of vectorial labeling problems. It generalizes the formulation for scalar-valued labeling problems [14] and thus includes important applications such as optical flow estimation or color image denoising. We show that our method, derived in a spatially continuous setting, has a variety of interesting theoretical properties as well as practical advantages over the existing labeling approaches:
– We generalize existing functional lifting approaches (see Sec. 2.2).
– We show that our method is the best convex under-approximation (in a local sense), see Prop. 1 and Prop. 2.
– Due to its sublabel-accuracy our method requires only a small amount of labels to produce good results which leads to a drastic reduction in memory. We believe that this is a vital step towards the real-time capability of lifting and convex relaxation methods. Moreover, our method eliminates the label bias, that previous lifting methods suffer from, even for many labels.
– In Sec. 2.3 we propose a regularizer that couples the different label components by enforcing a joint jump normal. This is in contrast to [9], where the components are regularized separately.
– For convex dataterms, our method is equivalent to the unlifted problem – see Prop. 4. Therefore, it allows a seamless transition between direct optimization and convex relaxation approaches.

1.4 Notation

We write $\langle x, y \rangle = \sum_i x_i y_i$ for the standard inner product on $\mathbb{R}^n$ or the Frobenius product if $x, y$ are matrices. Similarly $\| \cdot \|$ without any subscript denotes the usual Euclidean norm, respectively the Frobenius norm for matrices.

We denote the convex conjugate of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by $f^*(y) = \sup_{x \in \mathbb{R}^n} \langle y, x \rangle - f(x)$. It is an important tool for devising convex relaxations, as the biconjugate $f^{**}$ is the largest lower-semicontinuous (lsc.) convex function below $f$. For the indicator function of a set $C$ we write $\delta_C$, i.e., $\delta_C(x) = 0$ if $x \in C$ and $\infty$ otherwise. $\Delta_n^U \subset \mathbb{R}^n$ stands for the unit $n$-simplex.

2 Convex Formulation

2.1 Lifted Representation

Motivated by Fig. 1, we construct an equivalent representation of (1) in a higher dimensional space, before taking the convex envelope.

Let $\Gamma \subset \mathbb{R}^n$ be a compact and convex set. We partition $\Gamma$ into a set $\mathcal{T}$ of $n$-simplices $\Delta_i$ so that $\Gamma$ is a disjoint union of $\Delta_i$ up to a set of measure zero. Let $t^i_j$ be the $j$-th vertex of $\Delta_i$ and denote by $\mathcal{V} = \{t^1, \ldots, t^{\|\mathcal{V}\|}\}$ the union of all vertices, referred to as labels, with $1 \leq i \leq |\mathcal{T}|$, $1 \leq j \leq n + 1$ and $1 \leq i_j \leq |\mathcal{V}|$. For $u : \Omega \to \Gamma$, we refer to $u(x)$ as a sublabel. Any sublabel can be written as a convex combination of the vertices of a simplex $\Delta_i$ with $1 \leq i \leq |\mathcal{T}|$ for
appropriate barycentric coordinates $\alpha \in \Delta_n^U$:

$$u(x) = T_i \alpha := \sum_{j=1}^{n+1} \alpha_j t^{ij}, \quad T_i := (t^{i1}, t^{i2}, \ldots, t^{i_{n+1}}) \in \mathbb{R}^{n \times n+1}. \quad (4)$$

By encoding the vertices $t^k \in \mathcal{V}$ using a one-of-$|\mathcal{V}|$ representation $e^k$ we can identify any $u(x) \in \Gamma$ with a sparse vector $u(x)$ containing at least $|\mathcal{V}| - n$ many zeros and vice versa:

$$u(x) = E_i \alpha := \sum_{j=1}^{n+1} \alpha_j e^{ij}, \quad E_i := (e^{i1}, e^{i2}, \ldots, e^{i_{n+1}}) \in \mathbb{R}^{|\mathcal{V}| \times n+1},$$

$$u(x) = \sum_{k=1}^{|\mathcal{V}|} t^k u_k(x), \quad \alpha \in \Delta_n^U, \quad 1 \leq i \leq |\mathcal{T}|. \quad (5)$$

The entries of the vector $e^{ij}$ are zero except for the $(i_j)$-th entry, which is equal to one. We refer to $u : \Omega \rightarrow \mathbb{R}^{|\mathcal{V}|}$ as the lifted representation of $u$. This one-to-one-correspondence between $u(x) = T_i \alpha$ and $u(x) = E_i \alpha$ is shown in Fig. 2.
2.2 Convexifying the Dataterm

Let for now the weight of the regularizer in (1) be zero. Then, at each point \( x \in \Omega \) we minimize a generally nonconvex energy over a compact set \( \Gamma \subset \mathbb{R}^n \):

\[
\min_{u \in \Gamma} \rho(u).
\] (6)

We set up the lifted energy so that it attains finite values if and only if the argument \( u \) corresponds to a valid encoding \( u = E_i \alpha \) of a sublabel \( u \in \Gamma \):

\[
\rho(u) = \min_{1 \leq i \leq |T|} \rho_i(u), \quad \rho_i(u) = \begin{cases} 
\rho(T_i \alpha), & \text{if } u = E_i \alpha, \alpha \in \Delta^n_U, \\
\infty, & \text{else.}
\end{cases}
\] (7)

Problems (6) and (7) are equivalent due to the one-to-one correspondence of \( u = T_i \alpha \) and \( u = E_i \alpha \). However, energy (7) is finite on a nonconvex set only. In order to make optimization tractable, we minimize its convex envelope.

**Proposition 1** The convex envelope of (7) is given as:

\[
\rho^{**}(u) = \sup_{v \in \mathbb{R}^{|V|}} \langle u, v \rangle - \max_{1 \leq i \leq |T|} \rho_i^*(v), \quad \rho_i^*(v) = \langle E_i b_i, v \rangle + \rho_i^*(A_i^T E_i^T v), \quad \rho_i := \rho + \delta \Delta_i.
\] (8)

\( b_i \) and \( A_i \) are given as \( b_i := M_i^{n+1}, A_i := (M_i^1, M_i^2, \ldots, M_i^n) \), where \( M_i^j \) are the columns of the matrix \( M_i := (T_i^T, 1)^{-T} \in \mathbb{R}^{n+1 \times n+1} \).

**Proof.** Follows from a calculation starting at the definition of \( \rho^{**} \). See Appendix A for a detailed derivation.

Note that if one prescribes the value of \( \rho_i \) in (7) only on the vertices of the unit simplices \( \Delta^n_U \), i.e., \( \rho(u) = \rho(t^k) \) if \( u = e^k \) and \( +\infty \) otherwise, one obtains the linear biconjugate \( \rho^{**}(u) = \langle u, s \rangle, \ s = (\rho(t^1), \ldots, \rho(t^L)) \) on the feasible set. This coincides with the standard relaxation of the dataterm used in [5,20,21,2].

In that sense, our approach can be seen as a relaxing the dataterm in a more precise way, by incorporating the true value of \( \rho \) not only on the finite set of labels \( V \), but also everywhere in between, i.e., on every sublabel.

2.3 Lifting the Vectorial Total Variation

We define the lifted vectorial total variation as

\[
TV(u) = \int_{\Omega} \Psi(Du),
\] (9)
where $Du$ denotes the distributional derivative of $u$ and $\Psi$ is positively one-homogeneous, i.e., $\Psi(cu) = c\Psi(u), c \geq 0$. For such functions, the meaning of $(9)$ can be made fully precise using the polar decomposition of the Radon measure $Du$ [22] Cor. 1.29, Thm. 2.38]. However, in the following we restrict ourselves to an intuitive motivation for the derivation of $\Psi$ for smooth functions.

Our goal is to find $\Psi$ so that $TV(u) = TV(u)$ whenever $u : \Omega \rightarrow \mathbb{R}^{|V|}$ corresponds to some $u : \Omega \rightarrow \Gamma$, in the sense that $u(x) = E_i \alpha$ whenever $u(x) = T_i \alpha$. In order for the equality to hold, it must in particular hold for all $u$ that are classically differentiable, i.e., $Du = \nabla u$, and whose Jacobian $\nabla u(x)$ is of rank 1, i.e., $\nabla u(x) = (T_i \alpha - T_j \beta) \otimes \nu$ for some $\nu \in \mathbb{R}^d$. In that case, we observe

$$TV(u) = \int_\Omega \| T_i \alpha - T_j \beta \| \cdot \| \nu \| \, dx. \quad (10)$$

For the corresponding lifted representation $u$, we have $\nabla u(x) = (E_i \alpha - E_j \beta) \otimes \nu$. Therefore it is natural to require $\Psi(\nabla u(x)) = \Psi((E_i \alpha - E_j \beta) \otimes \nu) := \| T_i \alpha - T_j \beta \| \cdot \| \nu \|$ in order to achieve the goal $TV(u) = TV(u)$. Motivated by these observations, we define

$$\Psi(p) := \begin{cases} \| T_i \alpha - T_j \beta \| \cdot \| \nu \| & \text{if } p = (E_i \alpha - E_j \beta) \otimes \nu, \\ \infty & \text{otherwise}, \end{cases} \quad (11)$$

where $\alpha, \beta \in \Delta_{U_{n+1}}$, $\nu \in \mathbb{R}^d$ and $1 \leq i, j \leq |T|$. A “locally” tight convex relaxation of $(9)$ is given as

$$R(u) = \sup_{q : \Omega \rightarrow \mathbb{R}^d \times |V|} \int_\Omega \langle u, \text{Div } q \rangle - \Psi^*(q) \, dx. \quad (12)$$

**Proposition 2** The convex conjugate of $\Psi$ is

$$\Psi^*(q) = \delta_K(q) \quad (13)$$

with convex set

$$K = \bigcap_{1 \leq i, j \leq |T|} \left\{ q \in \mathbb{R}^{d \times |V|} \mid \| Q_i \alpha - Q_j \beta \| \leq \| T_i \alpha - T_j \beta \|, \alpha, \beta \in \Delta_{U_{n+1}} \right\}, \quad (14)$$

and $Q_i = (q^{i1}, q^{i2}, \ldots, q^{in+1}) \in \mathbb{R}^{d \times n+1}$. $q^i \in \mathbb{R}^d$ are the columns of $q$.

**Proof.** Follows from a calculation starting at the definition of the convex conjugate $\Psi^*$. See Appendix A.
Interestingly, although in its original formulation the set $K$ has infinitely many constraints, one can equivalently represent $K$ by finitely many.

**Proposition 3** The set $K$ in equation is the same as

$$K = \left\{ q \in \mathbb{R}^{d \times |V|} \mid \|D_q^i\|_{S^\infty} \leq 1, 1 \leq i \leq |T| \right\}, \quad D_q^i = Q_i D (T_i D)^{-1}, \quad (15)$$

where the matrices $Q_i D \in \mathbb{R}^{d \times n}$ and $T_i D \in \mathbb{R}^{n \times n}$ are given as

$$Q_i D := (q_{i1} - q_{in+1}, \ldots, q_{in} - q_{in+1}), \quad T_i D := (t_{i1} - t_{in+1}, \ldots, t_{in} - t_{in+1}).$$

**Proof.** Similar to the analysis in [2] in the continuous case, equation basically states the Lipschitz continuity of a piecewise linear function defined by the matrices $q \in \mathbb{R}^{d \times |V|}$. Therefore, one can expect that the Lipschitz constraint is equivalent to a bound on the derivative. For the complete proof, see Appendix A.

2.4 Lifting the Overall Optimization Problem

Combining dataterm and regularizer, the overall optimization problem is given

$$\min_{u: \Omega \to \mathbb{R}^{|V|}} \sup_{q: \Omega \to K} \int_{\Omega} \rho^{**}(u) + \langle u, \text{Div} q \rangle \, dx. \quad (16)$$

A highly desirable property is that, opposed to any other vectorial lifting approach from the literature, our method with just one simplex applied to a convex problem yields the same solution as the unlifted problem.

**Proposition 4** If the triangulation contains only 1 simplex, $T = \{\Delta\}$, i.e., $|V| = n + 1$, then the proposed optimization problem is equivalent to

$$\min_{u: \Omega \to \Delta} \int_{\Omega} \rho \Delta^{**}(x, u(x)) \, dx + \lambda TV(u), \quad (17)$$

which is [1] with a globally convexified dataterm on $\Delta$.

**Proof.** For $u = t^{n+1} + TD\bar{u}$ the substitution $u = (\bar{u}_1, \ldots, \bar{u}_n, 1 - \sum_{j=1}^n \bar{u}_j)$ into $\rho^{**}$ and $R$ yields the result. For a complete proof, see Appendix A.

3 Numerical Optimization

3.1 Discretization

For now assume that $\Omega \subset \mathbb{R}^d$ is a $d$-dimensional Cartesian grid and let $\text{Div}$ denote a finite-difference divergence operator with $\text{Div} q : \Omega \to \mathbb{R}^{|V|}$. Then the relaxed energy minimization problem becomes

$$\min_{u: \Omega \to \mathbb{R}^{|V|}} \max_{q: \Omega \to K} \sum_{x \in \Omega} \rho^{**}(x, u(x)) + \langle \text{Div} q, u \rangle. \quad (18)$$
In order to get rid of the pointwise maximum over $\rho^*_i(v)$ in Eq. (8), we introduce additional variables $w(x) \in \mathbb{R}$ and additional constraints $(v(x), w(x)) \in C$, $x \in \Omega$ so that $w(x)$ attains the value of the pointwise maximum:

$$
\min_{u: \Omega \to |V|} \max_{q: \Omega \to K} \sum_{x \in \Omega} \langle u(x), v(x) \rangle - w(x) + \langle \text{Div } q, u \rangle,
$$

(19)

where the set $C$ is given as

$$
C = \bigcap_{1 \leq i \leq |T|} C_i, \quad C_i := \left\{ (x, y) \in \mathbb{R}^{|V|+1} \mid \rho^*_i(x) \leq y \right\}.
$$

(20)

For numerical optimization we use a GPU-based implementation of a first-order primal-dual method [6]. The algorithm requires the orthogonal projections of the dual variables onto the sets $C$ respectively $K$ in every iteration. However, the projection onto an epigraph of dimension $|V|+1$ is difficult for large values of $|V|$. Hence we rewrite the constraints $(v(x), w(x)) \in C_i, 1 \leq i \leq |T|, x \in \Omega$ as $(n+1)$-dimensional epigraph constraints introducing additional variables $r^i(x) \in \mathbb{R}^n$ and $s^i(x) \in \mathbb{R}$:

$$
\rho_i^*(r^i(x)) \leq s_i(x), \quad r^i(x) = A_i^T E_i^T v(x), \quad s_i(x) = w(x) - \langle E_i b_i, v(x) \rangle.
$$

(21)

These equality constraints can be implemented using Lagrange multipliers. For the projection onto the set $K$ we use an approach similar to [23, Figure 7].

### 3.2 Epigraphical Projections

Computing the Euclidean projection onto the epigraph of $\rho^*_i$ is a central part of the numerical implementation of the presented method. However, for $n > 1$ this is nontrivial. Therefore we provide a detailed explanation of the projection methods used for different classes of $\rho_i$. We will consider quadratic, truncated quadratic and piecewise linear $\rho$.

**Quadratic case:** Let $\rho$ be of the form $\rho(u) = \frac{a}{2} u^T u + b^T u + c$. A direct projection onto the epigraph of $\rho^*_i = (\rho + \delta_{\Delta_i})^*$ for $n > 1$ is difficult. However, the epigraph can be decomposed into separate epigraphs for which it is easier to project onto: For proper, convex, lsc. functions $f, g$ the epigraph of $(f + g)^*$ is the Minkowski sum of the epigraphs of $f^*$ and $g^*$ (cf. [24, Exercise 1.28]). This means that it suffices to compute the projections onto the epigraphs of a quadratic function $f^* = \rho^*$ and a convex, piecewise linear function $g^*(v) = \max_{1 \leq j \leq n+1} \langle t^i_j, v \rangle$ by rewriting constraint (21) as

$$
\rho^*(r_f) \leq s_f, \quad \delta_{\Delta_i}^*(c_g) \leq d_g \quad \text{s.t. } (r, s) = (r_f, s_f) + (c_g, d_g).
$$

(22)
Fig. 3: ROF denoising of a vector-valued signal $f : [0, 1] \rightarrow [-1, 1]^2$, discretized on 50 points (shown in red). We compare the proposed approach (right) with two alternative techniques introduced in [2] (left and middle). The labels are visualized by the gray grid. While the naive (standard) multilabel approach from [2] (left) provides solutions that are constrained to the chosen set of labels, the sublabel accurate regularizer from [2] (middle) does allow sublabel solutions, yet – due to the dataterm bias – these still exhibit a strong preference for the grid points. In contrast, the proposed approach does not exhibit any visible grid bias providing fully sublabel-accurate solutions: With only 4 labels, the computed solutions (shown in blue) coincide with the “unlifted” problem (green).

For the projection onto the epigraph of a $n$-dimensional quadratic function we use the method described in [10, Appendix B.2]. The projection onto a piecewise linear function is described in the last paragraph of this section.

**Truncated quadratic case:** Let $\rho$ be of the form $\rho(u) = \min \{ \nu, \frac{\nu}{2} u^\top u + b^\top u + c \}$ as it is the case for the nonconvex robust ROF with a truncated quadratic dataterm in Sec. 4.2. Again, a direct projection onto the epigraph of $\rho_i^*$ is difficult. However, a decomposition of the epigraph into simpler epigraphs is possible as the epigraph of $\min\{f, g\}^*$ is the intersection of the epigraphs of $f^*$ and $g^*$. Hence, one can separately project onto the epigraphs of $(\nu + \delta_{\Delta_i})^*$ and $(\frac{\nu}{2} u^\top u + b^\top u + c + \delta_{\Delta_i})^*$. Both of these projections can be handled using the methods from the other paragraphs.

**Piecewise linear case:** In case $\rho$ is piecewise linear on each $\Delta_i$, i.e., $\rho$ attains finite values at a discrete set of sampled sublabels $\mathcal{V}_i \subset \Delta_i$ and interpolates linearly between them, we have that

$$ (\rho + \delta_{\Delta_i})^*(v) = \max_{\tau \in \mathcal{V}_i} \langle \tau, v \rangle - \rho(\tau). \quad (23) $$
Fig. 4: Convex ROF with vectorial TV. Direct optimization and proposed method yield the same result. In contrast to the baseline method \[2\] the proposed approach has no discretization artefacts and yields a lower energy. The regularization parameter is chosen as \(\lambda = 0.3\).

Again this is a convex, piecewise linear function. For the projection onto the epigraph of such a function, a quadratic program of the form

\[
\min_{(x,y) \in \mathbb{R}^{n+1}} \frac{1}{2} \|x - c\|^2 + \frac{1}{2} \|y - d\|^2 \quad \text{s.t.} \quad \langle \tau, x \rangle - \rho(\tau) \leq y, \forall \tau \in \mathcal{V}_i \quad (24)
\]

needs to be solved. We implemented the primal active-set method described in \[25\] Algorithm 16.3, and found it solves the program in a few (usually 2 – 10) iterations for a moderate number of constraints.

4 Experiments

4.1 Vectorial ROF Denoising

In order to validate experimentally, that our model is exact for convex dataterms, we evaluate it on the Rudin-Osher-Fatemi \[26\] (ROF) model with vectorial TV \[2\]. In our model this corresponds to defining \(\rho(x, u(x)) = \frac{1}{2} \|u(x) - I(x)\|^2\).

As expected based on Prop. \[4\] the energy of the solution of the unlifted problem is equal to the energy of the projected solution of our method for \(|V| = 4\) up to machine precision, as can be seen in Fig. \[3\] and Fig. \[4\]. We point out, that the sole purpose of this experiment is a proof of concept as our method introduces an overhead and convex problems can be solved via direct optimization. It can be seen in Fig. \[3\] and Fig. \[4\] that the baseline method \[2\] has a strong label bias.
Fig. 5: ROF with a truncated quadratic dataterm ($\lambda = 0.03$ and $\nu = 0.025$). Compared to the baseline method [2] the proposed approach yields much better results, already with a very small number of 4 labels.

4.2 Denoising with Truncated Quadratic Dataterm

For images degraded with both, Gaussian and salt-and-pepper noise we define the dataterm as $\rho(x, u(x)) = \min \left\{ \frac{1}{2} \|u(x) - I(x)\|^2, \nu \right\}$. We solve the problem using the epigraph decomposition described in the second paragraph of Sec. 3.2.

It can be seen, that increasing the number of labels $|\mathcal{V}|$ leads to lower energies and at the same time to a reduced effect of the TV. This occurs as we always compute a piecewise convex underapproximation of the original nonconvex dataterm, that gets tighter the larger the number of labels. The baseline method [2] again produces strong discretization artefacts even for a large number of labels $|\mathcal{V}| = 4 \times 4 \times 4 = 64$. 
Fig. 7: We compute the optical flow using our method, the product space approach [9] and the baseline method [2] for a varying amount of labels and compare the average endpoint error (aep). The product space method clearly outperforms the baseline, but our approach finds the overall best result already with 2 × 2 labels. To achieve a similarly precise result as the product space method, we require 150 times fewer labels, 10 times less memory and 3 times less time. The run times and memory requirements are to be taken as rough estimates, as they depend on the used stopping criteria and implementation.

4.3 Optical Flow

We compute the optical flow $v : \Omega \rightarrow \mathbb{R}^2$ between two input images $I_1, I_2$. The label space $\Gamma = [-d, d]^2$ is chosen according to the estimated maximum displacement $d \in \mathbb{R}$ between the images. The dataterm is $\rho(x, v(x)) = \|I_2(x) - I_1(x + v(x))\|$, and $\lambda(x)$ is based on the norm of the image gradient $\nabla I_1(x)$.

In Fig. 7 we compare the proposed method to the product space approach [9]. Note that we implemented the product space dataterm using Lagrange multipliers, also referred to as the global approach in [9]. While this increases the
Fig. 8: Joint estimation of mean and variance. Our formulation can optimize difficult nonconvex joint optimization problems with continuous label spaces.

memory consumption, it comes with lower computation time and guaranteed convergence. For our method, we sample the label space \( \Gamma = [-15, 15]^2 \) on \( 150 \times 150 \) sublabels and subsequently convexify the energy on each triangle using the quickhull algorithm \cite{27}. For the product space approach we sample the label space at equidistant labels, from \( 5 \times 5 \) to \( 27 \times 27 \). As the regularizer from the product space approach is different from the proposed one, we chose \( \mu \) differently for each method. For the proposed method, we set \( \mu = 0.5 \) and for the product space and baseline approach \( \mu = 3 \). We can see in Fig. 7 our method outperforms the product space approach w.r.t. the average end-point error. While our method outperforms previous lifting approaches, it does not produce competitive results on the Middlebury benchmark. For that, one would need to engineer a better dataterm. In Fig. 6 we compare our method on large displacement optical flow to the baseline \cite{2}.

4.4 Adaptive Denoising

In this experiment we jointly estimate the mean \( \mu \) and variance \( \sigma \) of an image \( I : \Omega \to \mathbb{R} \) according to a Gaussian model. The label space is chosen as \( \Gamma = [0, 255] \times [1, 10] \) and the dataterm as proposed in \cite{9}:

\[
\rho(x, \mu(x), \sigma(x)) = \frac{(\mu(x) - I(x))^2}{2\sigma(x)^2} + \frac{1}{2} \log(2\pi\sigma(x)^2).
\]  

(25)

As the projection onto the epigraph of \((\rho + \delta_\Delta)^*\) seems difficult to compute, we approximate \( \rho \) by a piecewise linear function using \( 29 \times 29 \) sublabels and convexify it using the quickhull algorithm \cite{27}. In Fig. 8 we show the result of minimizing (25) with total variation regularization.
5 Conclusions

We proposed the first sublabel-accurate convex relaxation of vectorial multilabel problems. To this end, we approximate the generally nonconvex dataterm in a piecewise convex manner as opposed to the piecewise linear approximation done in the traditional functional lifting approaches. This assures a more faithful approximation of the original cost function and provides a meaningful interpretation for the non-integral solutions of the relaxed convex problem. In experimental validations on large-displacement optical flow estimation and color image denoising, we show that the computed solutions have superior quality to the traditional convex relaxation methods while requiring substantially less memory and runtime.
A Appendix

Proof (Proof of Proposition 1). By definition the biconjugate of $\rho$ is again given as

$$\rho^{**}(u) = \sup_{v \in \mathbb{R}^{|V|}} \langle u, v \rangle - \left( \min_{1 \leq i \leq |T|} \rho^*_i(v) \right)^*$$

(26)

We proceed computing the conjugate of $\rho_i^*$:

$$\rho_i^*(v) = \sup_{u \in \mathbb{R}^{|V|}} \langle u, v \rangle - \rho_i(u)$$

$$= \sup_{\alpha \in \Delta_{n+1}} \langle E_i \alpha, v \rangle - \rho(T_i \alpha),$$

(27)

We introduce the substitution $r := T_i \alpha \in \Delta_i$ and obtain

$$\alpha = K_i^{-1} \begin{pmatrix} r \\ 1 \end{pmatrix}, \quad K_i := \begin{pmatrix} T_i \\ 1^\top \end{pmatrix} \in \mathbb{R}^{n+1 \times n+1},$$

(28)

since $K_i$ is invertible for $(V, T)$ being a non-degenerate triangulation and $\sum_{j=1}^{n+1} \alpha_j = 1$. With this we can further rewrite the conjugate as

$$\ldots = \sup_{r \in \Delta_i} \langle A_i r + b_i, E_i^\top v \rangle - \rho(r)$$

$$= \langle E_i b_i, v \rangle + \sup_{r \in \mathbb{R}^n} \langle r, A_i^\top E_i^\top v \rangle - \rho(r) - \delta_{\Delta_i}(r)$$

$$= \langle E_i b_i, v \rangle + \rho_i^*(A_i^\top E_i^\top v).$$

Proof (Proof of Proposition 2). Define $\Psi_{i,j}$ as

$$\Psi_{i,j}(p) := \begin{cases} \|T_i \alpha - T_j \beta\| \cdot \|\nu\| & \text{if } p = (E_i \alpha - E_j \beta) \nu^\top, \alpha, \beta \in \Delta_{n+1}^U, \nu \in \mathbb{R}^d, \\ \infty & \text{otherwise}. \end{cases}$$

(30)

Then, $\Psi$ can be rewritten as a pointwise minimum over the individual $\Psi_{i,j}$

$$\Psi(p) = \min_{1 \leq i, j \leq |T|} \Psi_{i,j}(p).$$

(31)

We begin computing the conjugate of $\Psi_{i,j}$

$$\Psi_{i,j}^*(q) = \sup_{p \in \mathbb{R}^d \times |V|} \langle p, q \rangle - \Psi_{i,j}(p)$$

$$= \sup_{\alpha, \beta \in \Delta_{n+1}^U} \sup_{\nu \in \mathbb{R}^d} \langle Q_i \alpha - Q_j \beta, \nu \rangle - \|T_i \alpha - T_j \beta\| \cdot \|\nu\|$$

$$= \sup_{\alpha, \beta \in \Delta_{n+1}^U} \left( \|T_i \alpha - T_j \beta\| \cdot \|\nu\| \right)^* (Q_i \alpha - Q_j \beta)$$

$$= \delta_{K_{i,j}}(q),$$

(32)
Fig. 9: Figure illustrating the second direction of the proof of Proposition 4. The gray dots and lines visualize the triangulation $(V, T)$. The line segment between $T_i\alpha$ and $T_j\beta$ is composed of shorter line segments which are fully contained in one of the triangles. On each of the triangles the inequality \cite{40} holds, which allows to conclude that it holds for the whole line segment.

with the set $K_{i,j}$ being defined as

$$K_{i,j} := \left\{ q \in \mathbb{R}^{d \times |V|} \mid \|Q_i\alpha - Q_j\beta\| \leq \|T_i\alpha - T_j\beta\|, \alpha, \beta \in \Delta_{n+1}^U \right\}. \quad (33)$$

Since the maximum over indicator functions of sets is equal to the indicator function of the intersection of the sets we obtain for $\Psi^*$

$$\Psi^*(q) = \max_{1 \leq i, j \leq |T|} \Psi^*_{i,j}(q) = \delta_{K}(q). \quad (34)$$

Proof (Proof of Proposition 3). Let $q \in \mathbb{R}^{d \times |V|}$ s.t. $\|Q_i\alpha - Q_j\beta\| \leq \|T_i\alpha - T_j\beta\|$ for all $\alpha, \beta \in \Delta_{n+1}^U$ and $1 \leq i, j \leq |T|$. For any $1 \leq i \leq |T|$ define

$$f_i : \mathbb{R}^n \to \mathbb{R}^n,$$

$$(\alpha_1, \ldots, \alpha_n) \mapsto \sum_{l=1}^{n} \alpha_l t_i^i + (1 - \sum_{l=1}^{n} \alpha_l) t_i^{i+1} = T_i\alpha, \quad (35)$$

and analogously

$$g_i : \mathbb{R}^n \to \mathbb{R}^{|V|},$$

$$(\alpha_1, \ldots, \alpha_n) \mapsto \sum_{l=1}^{n} \alpha_l q_i^i + (1 - \sum_{l=1}^{n} \alpha_l) q_i^{i+1} = Q_i\alpha. \quad (36)$$
Let us choose an \( \alpha \in \mathbb{R}^n \) such that \( \alpha_i > 0, \sum_l \alpha_l < 1 \). Then \( \|Q_i \alpha - Q_j \beta\| \leq \|T_i \alpha - T_j \beta\| \) for all \( \alpha, \beta \in \Delta^U_{n+1} \) and \( 1 \leq i, j \leq |T| \) implies that

\[
\|g_i(\alpha) - g_i(\alpha - h)\| \leq \|f_i(\alpha) - f_i(\alpha - h)\|,
\]

(37)

holds for all vectors \( h \) with sufficiently small entries. Inserting the definitions of \( g_i \) and \( f_i \) we find that

\[
\|Q_i D h\| \leq \|T_i D h\|
\]

(38)

holds for all \( h \) with sufficiently small entries. For a non-degenerate triangle, \( T_i D \) is invertible and a simple substitution yields that

\[
\|Q_i D(T_i D)^{-1} \tilde{h}\|_2 \leq \|\tilde{h}\|,
\]

(39)

holds for all \( \tilde{h} \) with sufficiently small entries. This means that the operator norm of \( D^i_q \) induced by the \( \ell^2 \) norm, i.e. the \( S^\infty \) norm, is bounded by one.

Let us now show the other direction. For \( q \in \mathbb{R}^{d \times |V|} \) s.t. \( \|D^i_q\|_{S^\infty} \leq 1, \ 1 \leq i \leq |T| \), note that inverting the above computation immediately yields that

\[
\|Q_k \alpha - Q_k \beta\| \leq \|T_k \alpha - T_k \beta\|
\]

(40)

holds for all \( 1 \leq k \leq |T|, \ \alpha, \beta \in \Delta^U_{n+1} \). Our goal is to show that having this inequality on each simplex is sufficient to extend it to arbitrary pairs of simplices. The overall idea of this part of the proof is illustrated in Fig. 9.

Let \( 1 \leq i, j \leq |T| \) and \( \alpha, \beta \in \mathbb{R}^n \) with \( \alpha_l, \beta_l \geq 0, \sum_l \alpha_l \leq \sum_l \beta_l \leq 1 \) be given. Consider the line segment

\[
c(\gamma) : [0, 1] \rightarrow \mathbb{R}^d
\]

\[
\gamma \mapsto \gamma T_j \beta + (1 - \gamma) T_i \alpha.
\]

(41)

Since the triangulated domain is convex, there exist \( 0 = a_0 < a_1 < \ldots < a_r = 1 \) and functions \( \alpha_l(\gamma) \) such that for \( \gamma \in [a_l, a_{l+1}], \ 0 \leq l \leq r - 1 \) one can write \( c(\gamma) = \gamma T_j \beta + (1 - \gamma) T_i \alpha = T_{k_l} \alpha_l(\gamma) \) for some \( 1 \leq k_l \leq T \). The continuity of \( c(\gamma) \) implies that \( T_{k_l} \alpha_l(a_{l+1}) = T_{k_{l+1}} \alpha_l(a_{l+1}) \), i.e. these points correspond to both simplices, \( k_l \) and \( k_{l+1} \). Note that this also means that \( Q_{k_l} \alpha_l(a_{l+1}) = Q_{k_{l+1}} \alpha_l(a_{l+1}) \). The intuition of this construction is that the \( c(a_{l+1}) \) are located
on the boundaries of adjacent simplices on the line segment. We find

\[ \| T_i \alpha - T_j \beta \| = \sum_{l=0}^{r-1} (a_{l+1} - a_l) \| T_i \alpha - T_j \beta \| \]

\[ = \sum_{l=0}^{r-1} \| (a_{l+1} - a_l)(T_i \alpha - T_j \beta) \| \]

\[ = \sum_{l=0}^{r-1} \| a_{l+1}T_i \alpha - a_lT_i \alpha - a_{l+1}T_j \beta + a_lT_j \beta \| \]

\[ = \sum_{l=0}^{r-1} \| a_lT_j \beta + (1 - a_l)T_i \alpha - (a_{l+1}T_j \beta + (1 - a_{l+1})T_i \alpha) \| \]

\[ = \sum_{l=0}^{r-1} \| T_k \alpha_l(a_l) - T_k \alpha_l(a_{l+1}) \| \]  \hfill (42)

\[ \| Q_k \alpha_0(a_0) - Q_k \alpha_r(a_r) \| \]

which yields the assertion.

**Proof (Proof of Proposition 4).** Let \( \Delta = \text{conv}\{t^1, \ldots, t^{n+1}\} \) be given by affinely independent vertices \( t^i \in \mathbb{R}^n \). We show that our lifting approach applied to the label space \( \Delta \) solves the convexified unlifted problem, where the dataterm was replaced by its convex hull on \( \Delta \). Let the matrices \( T \in \mathbb{R}^{n \times (n+1)} \) and \( D \in \mathbb{R}^{(n+1) \times n} \) be defined through

\[ T = \begin{pmatrix} t^1 \\ \vdots \\ t^{n+1} \end{pmatrix}, \quad D = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ -1 \ldots -1 \end{pmatrix}, \quad TD = \begin{pmatrix} t^1 - t^{n+1}, \ldots, t^n - t^{n+1} \end{pmatrix} \]  \hfill (43)

The transformation \( x \mapsto t^{n+1} + TDx \) maps \( \Delta_e = \text{conv}\{0, e^1, \ldots, e^n\} \subset \mathbb{R}^n \) to \( \Delta \). Now consider the following lifted function \( u : \Omega \rightarrow \mathbb{R}^{n+1} \) parametrized through
\[\tilde{u} : \Omega \to \Delta_c:\]

\[u(x) = \left(\tilde{u}_1(x), \ldots, \tilde{u}_n(x), 1 - \sum_{j=1}^n \tilde{u}_j(x)\right). \quad (44)\]

Consider a fixed \(x \in \Omega\). Plugging this lifted representation into the biconjugate of the lifted dataterm \(\rho\) yields:

\[
\rho^{**}(u) = \sup_{v \in \mathbb{R}^{n+1}} \langle u, v \rangle - \sup_{\alpha \in \Delta_{n+1}^V} \langle \alpha, v \rangle - \rho(T\alpha)
\]

\[
= \sup_{v \in \mathbb{R}^{n+1}} \left\langle \left(\tilde{u}_1(x), \ldots, \tilde{u}_n(x), 1 - \sum_{j=1}^n \tilde{u}_j(x)\right), v \right\rangle - \sup_{\alpha \in \Delta_{n+1}^V} \langle \alpha, v \rangle - \rho(T\alpha)
\]

\[
= \sup_{v \in \mathbb{R}^{n+1}} \langle \tilde{u}, D^T v \rangle + v_{n+1} - \sup_{\alpha \in \Delta_{n+1}^V} \langle \alpha, v \rangle - \rho(T\alpha)
\]

\[
= \sup_{v \in \mathbb{R}^{n+1}} \langle \tilde{u}, D^T v \rangle + v_{n+1} - \sup_{\alpha \in \Delta_{n+1}^V} \langle \alpha, D^T v \rangle - \rho(t^{n+1} + TD\alpha)
\]

(45)

Since \(D^T\) is surjective, we can apply the substitution \(\tilde{v} = D^T v\):

\[\ldots = \sup_{\tilde{v} \in \mathbb{R}^n} \langle \tilde{u}, \tilde{v} \rangle - \sup_{\alpha \in \Delta_{n+1}^V} \langle \alpha, \tilde{v} \rangle - \rho(t^{n+1} + TD\alpha)
\]

\[= \sup_{\tilde{v} \in \mathbb{R}^n} \langle \tilde{u}, \tilde{v} \rangle - \sup_{w \in \Delta} \langle (TD)^{-1}(w - t^{n+1}), \tilde{v} \rangle - \rho(w). \quad (46)\]

In the last step the substitution \(w = t^{n+1} + TD\alpha \Leftrightarrow \alpha = (TD)^{-1}(w - t^{n+1})\) was performed. This can be further simplified to

\[\ldots = \sup_{\tilde{v} \in \mathbb{R}^n} \langle \tilde{u}, \tilde{v} \rangle + \langle (TD)^{-1}t^{n+1}, \tilde{v} \rangle - (\rho + \delta\Delta)^*((TD)^{-T}\tilde{v})
\]

\[= \sup_{\tilde{v} \in \mathbb{R}^n} \langle \tilde{u} + (TD)^{-1}t^{n+1}, \tilde{v} \rangle - (\rho + \delta\Delta)^*((TD)^{-T}\tilde{v})
\]

\[= \sup_{\tilde{v} \in \mathbb{R}^n} \langle TD\tilde{u} + t^{n+1}, (TD)^{-T}\tilde{v} \rangle - (\rho + \delta\Delta)^*((TD)^{-T}\tilde{v}). \quad (47)\]

Since \(TD\) is invertible we can perform another substitution \(v' = (TD)^{-T}\tilde{v}\).

\[\ldots = \sup_{v' \in \mathbb{R}^n} \langle TD\tilde{u} + t^{n+1}, v' \rangle - (\rho + \delta\Delta)^*(v')
\]

\[= (\rho + \delta\Delta)^**(t^{n+1} + TD\tilde{u}). \quad (48)\]
The lifted regularizer is given as:

\[ R(u) = \sup_{q : \Omega \to \mathbb{R}^{d \times n + 1}} \int_{\Omega} \langle u, \text{Div } q \rangle - \Psi^*(q) \, dx \]  
(49)

Using the parametrization by \( \tilde{u} \), this can be equivalently written as

\[ \sup_{\tilde{q}(x) \in \tilde{K}} \int_{\Omega} \sum_{j=1}^{n} \tilde{u}_j \text{Div}(q_j - q_{n+1}) + \text{Div } q_{n+1} \, dx, \]  
(50)

where the set \( \tilde{K} \subset \mathbb{R}^{d \times n + 1} \) can be written as

\[ \tilde{K} = \{ q \in \mathbb{R}^{d \times n + 1} \mid \| D^\top q^\top (TD)^{-1} \|_S \leq 1 \}. \]  
(51)

Note that since \( q_{n+1} \in C_c^\infty(\Omega, \mathbb{R}^d) \), the last term \( \text{Div } q_{n+1} \) in (50) vanishes by partial integration. With the substitution \( \tilde{q}(x) = D^\top q(x)^\top \) we have

\[ \sup_{\tilde{q} \in \tilde{K}} \int_{\Omega} \langle \tilde{u}, \text{Div } \tilde{q} \rangle \, dx, \]  
(52)

with set \( \tilde{K} \subset \mathbb{R}^{d \times n} \):

\[ \tilde{K} = \{ q \in \mathbb{R}^{d \times n} \mid \| q(TD)^{-1} \|_S \leq 1 \}. \]  
(53)

Note that since \( q_i \in C_c^\infty(\Omega, \mathbb{R}^d) \), the same holds for the linearly transformed \( \tilde{q} \).

With another substitution \( q'(x) = \tilde{q}(x)(TD)^{-1} \) we have

\[ \cdots = \sup_{q' \in \mathcal{K}'} \int_{\Omega} \langle \tilde{u}, \text{Div } q'TD \rangle \, dx \]  
(54)

\[ = \sup_{q' \in \mathcal{K}'} \int_{\Omega} \langle TD\tilde{u}, \text{Div } q' \rangle \, dx \]

where the set \( \mathcal{K}' \subset \mathbb{R}^{d \times n + 1} \) is given as

\[ \mathcal{K}' = \{ q \in \mathbb{R}^{d \times n} \mid \| q \|_S \leq 1 \}, \]  
(55)

which is the usual unlifted definition of the total variation \( TV(t^{n+1} + TD\tilde{u}) \).

This shows that the lifting method solves

\[ \min_{\tilde{u} : \Omega \to \Delta_n} \int_{\Omega} (\rho(x, \cdot) + \delta \Delta)^{**}(t^{n+1} + TD\tilde{u}(x)) \, dx + \lambda TV(t^{n+1} + TD\tilde{u}), \]  
(56)

which is equivalent to the original problem but with a convexified data term.
References

1. Sapiro, G., Ringach, D.: Anisotropic diffusion of multivalued images with applications to color filtering. IEEE Trans. Img. Proc. 5(11) (1996) 1582–1586
2. Lellmann, J., Strekalovskiy, E., Koetter, S., Cremers, D.: Total variation regularization for functions with values in a manifold. In: ICCV. (December 2013)
3. Ishikawa, H.: Exact optimization for Markov random fields with convex priors. IEEE Trans. Pattern Analysis and Machine Intelligence 25(10) (2003) 1333–1336
4. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: European Conference on Computer Vision (ECCV), Marseille, France (October 2008)
5. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM J. Imaging Sci. 3(4) (2010) 1122–1145
6. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: ICCV. (2009)
7. Giaquinta, M., Modica, G., Souček, J.: Cartesian currents in the calculus of variations I, II. Volume 37-38 of Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Springer-Verlag, Berlin (1998)
8. Alberti, G., Bouchitté, G., Maso, G.D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calc. Var. Partial Dif. 3(16) (2003) 299–333
9. Goldluecke, B., Strekalovskiy, E., Cremers, D.: Tight convex relaxations for vector-valued labeling. SIAM Journal on Imaging Sciences 6(3) (2013) 1626–1664
10. Strekalovskiy, E., Chambolle, A., Cremers, D.: Convex relaxation of vectorial problems with coupled regularization. SIAM Journal on Imaging Sciences 7(1) (2014) 294–336
11. Zach, C., Kohli, P.: A convex discrete-continuous approach for markov random fields. In: ECCV. Volume 7577. Springer Berlin Heidelberg (2012) 386–399
12. Fix, A., Agarwal, S.: Duality and the continuous graphical model. In: Computer Vision ECCV 2014. Volume 8691 of Lecture Notes in Computer Science. Springer International Publishing (2014) 266–281
13. Zach, C.: Dual decomposition for joint discrete-continuous optimization. In: AISTATS. (2013) 632–640
14. Möllenhoff, T., Laude, E., Moeller, M., Lellmann, J., Cremers, D.: Sublabel-accurate relaxation of nonconvex energies. In: CVPR. (2016)
15. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: ECCV. (2004) 25–36
16. Mobahi, H., Fisher, III, J.W.: On the link between gaussian homotopy continuation and convex envelopes. In: Lecture Notes in Computer Science (EMMCVPR 2015), Springer (2015)
17. Mobahi, H., Fisher, III, J.W.: A theoretical analysis of optimization by gaussian continuation. In: AAAI-15: 29th Conference on Artificial Intelligence. (2015)
18. Aujol, J.F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition, algorithms, and parameter selection. International Journal of Computer Vision 67(1) (2006) 111–136
19. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l 1 optical flow. In: Pattern Recognition. Springer (2007) 214–223
20. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. SIAM Journal on Imaging Sciences 4(4) (2011) 1049–1096
21. Chambolle, A., Cremers, D., Pock, T.: A convex approach to minimal partitions. SIAM Journal on Imaging Sciences 5(4) (2012) 1113–1158
22. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York (2000)
23. Goldluecke, B., Strekalovskiy, E., Cremers, D.: The natural total variation which arises from geometric measure theory. SIAM Journal on Imaging Sciences 5(2) (2012) 537–563
24. Rockafellar, R., Wets, R.B.: Variational Analysis. Springer (1998)
25. Nocedal, J., Wright, S.J.: Numerical Optimization. 2nd edn. Springer, New York (2006)
26. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1) (1992) 259–268
27. Barber, C.B., Dobkin, D.P., Huhdanpaa, H.: The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS) 22(4) (1996) 469–483