Sublabel–Accurate Relaxation of Nonconvex Energies

Thomas Möllenhoff\(^*\)  
TU München  
moellenh@in.tum.de

Emanuel Laude\(^*\)  
TU München  
laudee@in.tum.de

Michael Moeller  
TU München  
moellerm@in.tum.de

Jan Lellmann  
University of Lübeck  
lellmann@mic.uni-luebeck.de

Daniel Cremers  
TU München  
cremers@tum.de

Abstract

We propose a novel spatially continuous framework for convex relaxations based on functional lifting. Our method can be interpreted as a sublabel–accurate solution to multilabel problems. We show that previously proposed functional lifting methods optimize an energy which is linear between two labels and hence require (often infinitely) many labels for a faithful approximation. In contrast, the proposed formulation is based on a piecewise convex approximation and therefore needs far fewer labels – see Fig. 1. In comparison to recent MRF-based approaches, our method is formulated in a spatially continuous setting and shows less grid bias. Moreover, in a local sense, our formulation is the tightest possible convex relaxation. It is easy to implement and allows an efficient primal-dual optimization on GPUs. We show the effectiveness of our approach on several computer vision problems.

1. Introduction

Energy minimization methods have become the central paradigm for solving practical problems in computer vision. The energy functional can often be written as the sum of a data fidelity and a regularization term. One of the most popular regularizers is the total variation (TV) due to its many favorable properties [4]. Hence, an important class of optimization problems is given as

\[
\min_{u: \Omega \to \Gamma} \int_{\Omega} \rho(x, u(x)) \, dx + \lambda \text{TV}(u),
\]  

(1)

\(^*\)Those authors contributed equally.

Pock et al. [17], 48 labels, 1.49 GB, 52 s.  Proposed, 8 labels, 0.49 GB, 30 s.

Figure 1. We propose a convex relaxation for the variational model (1), which opposed to existing functional lifting methods [17, 18] allows continuous label spaces even after discretization. Our method (here applied to stereo matching) avoids label space discretization artifacts, while saving on memory and runtime.

defined for functions \(u\) with finite total variation, arbitrary, possibly nonconvex dataterms \(\rho : \Omega \times \Gamma \to \mathbb{R}\), label spaces \(\Gamma\) which are closed intervals in \(\mathbb{R}\), \(\Omega \subset \mathbb{R}^d\), and \(\lambda \in \mathbb{R}^+\). The multilabel interpretation of the datatrm is that \(\rho(x, u(x))\) represents the costs of assigning label \(u(x)\) to point \(x\). For (weakly) differentiable functions \(TV(u)\) equals the integral over the norm of the derivative, and therefore favors a spatially coherent label configuration. The difficulty of minimizing the nonconvex energy (1) has motivated researchers to develop convex reformulations.

Convex representations of (1) and more general related energies have been studied in the context of the calibration method for the Mumford-Shah functional [1]. Based on these works, relaxations for the piecewise constant [15] and piecewise smooth Mumford-Shah functional [16] have been proposed. Inspired by Ishikawa’s graph-theoretic globally
optimal solution to discrete variants of (1), continuous analogues have been considered by Pock et al. in [17, 18]. Continuous relaxations for multilabeling problems with finite label spaces $\Gamma$ have also been studied in [11].

Interestingly, the discretization of the aforementioned continuous relaxations is very similar to the linear programming relaxations proposed for MAP inference in the Markov Random Field (MRF) community [10, 22, 24, 26]. Both approaches ultimately discretize the range $\Gamma$ into a finite set of labels. A closer analysis of these relaxations reveals, however, that they are not well-suited to represent the continuous valued range that we face in most computer vision problems such as stereo matching or optical flow. More specifically, the above relaxations are not designed to assign meaningful cost values to non-integral configurations. As a result, a large number of labels is required to achieve a faithful approximation. Solving real-world vision problems therefore entails large optimization problems with high memory and runtime requirement. To address this variation can be represented in a very simple way, introducing no overhead compared to [17, 18]. In contrast, the regularizer in [27] is much more involved.

- Since our method is derived in a spatially continuous setting, the proposed approach easily allows different gradient discretizations. In contrast to [25, 27] the regularizer is isotropic leading to noticeably less grid bias.

1.1. Contributions

We propose the first sublabel–accurate convex relaxation of nonconvex problems in a spatially continuous setting. It exhibits several favorable properties:

- In contrast to existing spatially continuous lifting approaches [17, 18], the proposed method provides substantially better solutions with far fewer labels – see Fig. 1. This provides savings in runtime and memory.

- In Sec. 3 we show that the functional lifting methods [17, 18] are a special case of the proposed framework.

- In Sec. 3 we show that, in a local sense, our formulation is the tightest convex relaxation which takes dataterm and regularizer into account separately. It is unknown whether this “local convex envelope” property also holds for the discrete approach [27].

- Our formulation is compact and requires only half the amount of variables for the dataterm than the formulation in [27]. We prove that the sublabel–accurate total variation can be represented in a very simple way, introducing no overhead compared to [17, 18]. In contrast, the regularizer in [27] is much more involved.

2. Notation and Mathematical Preliminaries

We make heavy use of the convex conjugate, which is given as $f^*(y) = \sup_{x \in \mathbb{R}^n} \langle y, x \rangle - f(x)$ for functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. The biconjugate $f^{**}$ denotes its convex envelope, i.e. the largest lower-semicontinuous convex under-approximation of $f$. For a set $C$ we denote by $\delta_C$ the function which maps any element from $C$ to 0 and is $\infty$ otherwise. For a comprehensive introduction to convex analysis, we refer the reader to [19]. Vector valued functions $u : \Omega \to \mathbb{R}^k$ are written in bold symbols. If it is clear from the context, we will drop the $x \in \Omega$ inside the functions, e.g., we write $\rho(u)$ for $\rho(x, u(x))$, or $\alpha$ for $\alpha(x)$.

3. Functional Lifting

To derive a convex representation of (1), we rely on the framework of functional lifting. The idea is to reformulate the optimization problem in a higher dimensional space. We numerically show in Sec. 5 that considering the convex envelope of the dataterm and regularizer in this higher dimensional space leads to a better approximation of the original nonconvex energy. We start by sampling the range...
\[ \Gamma \] at \( L = k + 1 \) labels \( \gamma_1 < \ldots < \gamma_L \in \Gamma \). This partitions the range into \( k \) intervals \( \Gamma_i = [\gamma_i, \gamma_{i+1}] \) so that \( \Gamma = \Gamma_1 \cup \ldots \cup \Gamma_k \). For any value in the range of \( u : \Omega \to \Gamma \) there exist a label index \( 1 \leq i \leq k \) and \( \alpha \in [0, 1] \) such that
\[
u(x) = \gamma_i^\alpha := \gamma_i + \alpha(\gamma_{i+1} - \gamma_i). \tag{2}\]
We represent a value in the range \( \Gamma \) by a vector in \( \mathbb{R}^k \)
\[
u(x) = 1_i^\alpha := \alpha 1_i + (1 - \alpha)1_{i-1}, \tag{3}\]
where \( 1_i \) denotes a vector starting with \( i \) ones followed by \( k - i \) zeros. We call \( u : \Omega \to \mathbb{R}^k \) the lifted representation of \( u \), representing the graph of \( u \). This notation is depicted in Fig. 2 for \( k = 4 \). Back-projecting the lifted \( \nu(x) \) to the range of \( u \) using the layer cake formula yields a one-to-one correspondence between \( \nu(x) = \gamma_i^\alpha \) and \( \nu(x) = 1_i^\alpha \) via
\[
u(x) = \gamma_1 + \sum_{i=1}^k u_i(x)(\gamma_{i+1} - \gamma_i). \tag{4}\]
We write problem (1) in terms of such graph functions, a technique that is used in the theory of Cartesian currents [8].

3.1. Convexification of the Dataterm

For now, we consider a fixed \( x \in \Omega \). Then the dataterm from (1) is a possibly nonconvex real-valued function (cf. Fig. 3) that we seek to minimize over a compact interval \( \Gamma \):
\[
\min_{u \in \Gamma} \rho(u). \tag{5}\]
Due to the one-to-one correspondence between \( \gamma_i^\alpha \) and \( 1_i^\alpha \) it is clear that solving problem (5) is equivalent to finding a minimizer of the lifted energy:
\[
\rho(u) = \min_{1 \leq i \leq k} \rho_i(u), \tag{6}\]
\[
\rho_i(u) = \begin{cases} 
\rho(\gamma_i^\alpha), & \text{if } u = 1_i^\alpha, \ \alpha \in [0, 1], \\
\infty, & \text{else.} 
\end{cases} \tag{7}\]
Note that the constraint in (7) is essentially the nonconvex special ordered set of type 2 (SOS2) constraint [3]. More precisely, we demand that the “derivative” in label direction \( \partial_i u \) is zero, except for two neighboring elements, which add up to one. In the following proposition, we derive the tightest convex relaxation of \( \rho \).

**Proposition 1.** The convex envelope of (6) is given as:
\[
\rho^{**}(u) = \sup_{v \in \mathbb{R}^k} \langle u, v \rangle - \max_{1 \leq i \leq k} \rho_i^*(v), \tag{8}\]
where the conjugate of the individual \( \rho_i \) is
\[
\rho_i^*(v) = c_i(v) + \rho_i^* \left( \frac{v_i}{\gamma_{i+1} - \gamma_i} \right), \tag{9}\]
with \( c_i(v) = \langle 1_{i-1}, v \rangle - \frac{\gamma_{i-1}}{\gamma_{i+1} - \gamma_i} v_i \) and \( \rho_i = \rho + \delta_{\Gamma_i} \).

**Proof.** See supplementary material. \( \square \)

The above proposition reveals that the convex relaxation implicitly convexifies the dataterm \( \rho \) on each interval \( \Gamma_i \). The equality \( \rho_i^* = \rho_i^{**} \) implies that starting with \( \rho_i \) yields exactly the same convex relaxation as starting with \( \rho_i^* \).

**Corollary 1.** If \( \rho \) is linear on each \( \Gamma_i \), then the convex envelopes of \( \rho(u) \) and \( \sigma(u) \) coincide, where the latter is:
\[
\sigma(u) = \begin{cases} 
\rho(\gamma_i^\alpha), & \text{if } \exists i : u = 1_i^\alpha, \ \alpha \in [0, 1], \\
\infty, & \text{else.} 
\end{cases} \tag{10}\]

**Proof.** Consider an additional constraint \( \delta_{\{\gamma_i, \gamma_{i+1}\}} \) for each \( \rho_i \), which corresponds to selecting \( \alpha \in [0, 1] \) in (7). The fact that our relaxation is independent of whether we choose \( \rho_i \) or \( \rho_i^* \), along with the fact that the convex hull of two points is a line, yields the assertion. \( \square \)

For the piecewise linear case, it is possible to find an explicit form of the biconjugate.
Proposition 2. Let us denote by \( r \in \mathbb{R}^k \) the vector with
\[
  r_i = \rho(\gamma_{i+1}) - \rho(\gamma_i), \quad 1 \leq i \leq k.
\]
Under the assumptions of Prop. 1, one obtains:
\[
  \sigma^{**}(u) = \begin{cases} 
    \rho(\gamma_1) + \langle u, r \rangle, & \text{if } u_i \geq u_{i+1}, u_i \in [0, 1], \\
    \infty, & \text{else}.
  \end{cases}
\]

Proof. See supplementary material.

Proposition 3. In the case of binary labeling, i.e., \( L = 2 \),
the convex envelope of (6) reduces to
\[
  \rho^{**}(u) = \rho^{**}(\gamma_1 + u(\gamma_2 - \gamma_1)), \quad \text{with } u \in [0, 1].
\]

Proof. See supplementary material.

3.2. A Lifted Representation of the Total Variation

We now want to find a lifted convex formulation that emulates the total variation
regularization in (1). We follow [5] and define an appropriate integrand of the functional
\[
  TV(u) = \int_\Omega \Phi(x, Du),
\]
where the distributional derivative \( Du \) is a finite \( \mathbb{R}^{k \times d} \)-valued Radon measure [2]. We define
\[
  \Phi(g) = \min_{1 \leq i \leq j \leq k} \Phi_{i,j}(g).
\]
The individual \( \Phi_{i,j} : \mathbb{R}^{k \times d} \rightarrow \mathbb{R} \cup \{ \infty \} \) are given by:
\[
  \Phi_{i,j}(g) = \begin{cases} 
    |\gamma_i - \gamma_j| \cdot |\nu|_2, & \text{if } g = (1_i^\alpha - 1_j^\beta) \nu_T, \\
    \infty, & \text{else}.
  \end{cases}
\]

Figure 4. Illustration of the epigraph projection. In the left subfigure the projection onto the epigraph of the conjugate of a convex quadratic \( \rho_i \) is shown. In the right subfigure the piecewise linear case is illustrated. In the both cases all points that lie in the green sets are orthogonally projected onto the respective linear parts whereas the points that lie in the red sets are projected onto the parabolic part (in the quadratic case) respectively the kinks (in the piecewise linear case). In the piecewise linear case the green sets are normal cones. The red dashed lines correspond to the boundary cases. \( \gamma_i, \gamma_{i+1}, \mu_1, \mu_2 \) are the slopes of the segments of \( \rho_i \) respectively the (sub-)label positions of \( \rho_i \) for some \( \alpha, \beta \in [0, 1] \) and \( \nu \in \mathbb{R}^d \). The intuition is that \( \Phi_{i,j} \) penalizes a jump from \( \gamma_i^\alpha \) to \( \gamma_j^\beta \) in the direction of \( \nu \). Since \( \Phi \) is nonconvex we compute the convex envelope.

Proposition 4. The convex envelope of (15) is
\[
  \Phi^{**}(g) = \sup_{p \in \mathcal{K}} \langle p, g \rangle,
\]
where \( \mathcal{K} \subset \mathbb{R}^{k \times d} \) is given as:
\[
  \mathcal{K} = \left\{ p \in \mathbb{R}^{k \times d} \mid \left| p^T(1_i^\alpha - 1_j^\beta) \right|_2 \leq |\gamma_i^\alpha - \gamma_j^\beta|, \quad \forall 1 \leq i \leq j \leq k, \quad \forall \alpha, \beta \in [0, 1] \right\}.
\]

Proof. See supplementary material.

The set \( \mathcal{K} \) from Eq. (18) involves infinitely many constraints which makes numerical optimization difficult. As the following proposition reveals, the infinite number of constraints can be reduced to only linearly many, allowing to enforce the constraint \( p \in \mathcal{K} \) exactly.

Proposition 5. If the labels are ordered \( \gamma_1 < \gamma_2 < \ldots < \gamma_L \) then the constraint set \( \mathcal{K} \) from Eq. (18) is equal to
\[
  \mathcal{K} = \left\{ p \in \mathbb{R}^{k \times d} \mid |p_i|_2 \leq \gamma_{i+1} - \gamma_i, \quad \forall i \right\}.
\]

Proof. See supplementary material.
where the relaxed energy minimization problem becomes

\[ \rho \rightarrow \max_{\gamma} \langle \rho, \gamma \rangle \]

by replacing the pointwise maximum over the conjugates \( \rho^* \) with a maximum over a real number \( q \in \mathbb{R} \) and obtain

\[ \min_{u : \Omega \rightarrow \mathbb{R}^d} \sum_{x \in \Omega} \rho^*(x, u(x)) + \Phi^*(x, \nabla u(x)), \]

(20)

where \( \nabla \) denotes a forward-difference operator with \( \nabla u : \Omega \rightarrow \mathbb{R}^{k \times d} \). We rewrite the dataterm given in equation (8) by replacing the pointwise maximum over the conjugates \( \rho^*_i \) with a maximum over a real number \( q \in \mathbb{R} \) and obtain

\[ \min_{u : \Omega \rightarrow \mathbb{R}^d} \sum_{x \in \Omega} \rho^*_i(x, u(x)) + \Phi^*(x, \nabla u(x)), \]

(22)

This shows that the proposed regularizer coincides with the total variation from [5], where it has been derived based on (16) for \( \alpha \) and \( \beta \) restricted to \( \{0, 1\} \). Prop. 5 together with Prop. 3 show that for \( k = 1 \) our formulation amounts to unlifted TV optimization with a convexified dataterm.

4. Numerical Optimization

Discretizing \( \Omega \subset \mathbb{R}^d \) as a \( d \)-dimensional Cartesian grid, the relaxed energy minimization problem becomes

\[ \min_{u : \Omega \rightarrow \mathbb{R}^d} \sum_{x \in \Omega} \rho^*_i(x, u(x)) + \Phi^*(x, \nabla u(x)), \]

(20)

where \( \nabla \) denotes a forward-difference operator with \( \nabla u : \Omega \rightarrow \mathbb{R}^{k \times d} \). We rewrite the dataterm given in equation (8) by replacing the pointwise maximum over the conjugates \( \rho^*_i \) with a maximum over a real number \( q \in \mathbb{R} \) and obtain

\[ \min_{u : \Omega \rightarrow \mathbb{R}^d} \sum_{x \in \Omega} \rho^*_i(x, u(x)) + \Phi^*(x, \nabla u(x)), \]

(22)

We numerically compute a minimizer of problem (21) using a first-order primal-dual method [6, 16] with diagonal preconditioning [14] and adaptive steps [9]. It alternates between a gradient descent step in the primal variable and a gradient ascent step in the dual variable. Subsequently the dual variables are orthogonally projected onto the sets \( C \) respectively \( K \). In the following we give some hints on the implementation of the individual steps. For a detailed discussion we refer to [9]. The projection onto the set \( K \) is a simple \( \ell_2 \)-ball projection. To simplify the projection onto \( C \), we transform the \( k \)-dimensional epigraph constraints in (22) into 1-dimensional scaled epigraph constraints by introducing an additional variable \( z : \Omega \rightarrow \mathbb{R}^k \) with:

\[ z_i(x) = [q(x) - c_i(v(x))] (\gamma_{i+1} - \gamma_i) \]

(23)
Figure 7. Comparison to the MRF approach presented in [27]. The first row shows DC-Linear, second row DC-MRF and third row our results for 4, 8, 16 and 32 convex pieces on the truncated quadratic energy (26). Below the figures we show the final non-convex energy. We achieve competitive results while using a more compact representation and generalizing to isotropic regularizers.

Using equation (9) we can write the constraints in (22) as

\[
\frac{z_i(x)}{\gamma_{i+1} - \gamma_i} \geq \rho_i^* \left( \frac{v_i(x)}{\gamma_{i+1} - \gamma_i} \right).
\] (24)

We implement the newly introduced equality constraints (23) introducing a Lagrange multiplier \(s : \Omega \rightarrow \mathbb{R}^k\). It remains to discuss the orthogonal projections onto the epigraphs of the conjugates \(\rho_i^*\). Currently we support quadratic and piecewise linear convex pieces \(\rho_i\). For the piecewise linear case, the conjugate \(\rho_i^*\) is a piecewise linear function with domain \(\mathbb{R}\). The slopes correspond to the \(x\)-positions of the sublabels and the intercepts correspond to the function values at the sublabel positions.

The conjugates as well as the epigraph projections of both, a quadratic and a piecewise linear piece are depicted in Fig. 4. For the quadratic case, the projection onto the epigraph of a parabola is computed using [23, Appendix B.2].

5. Experiments

We implemented the primal-dual algorithm in CUDA to run on GPUs.\(^1\) For \(d = 2\), our implementation of the functional lifting framework [17], which will serve as a baseline method, requires \(4N(L-1)\) optimization variables, while the proposed method requires \(6N(L-1) + N\) variables, where \(N\) is the number of points used to discretize the domain \(\Omega \subset \mathbb{R}^d\). As we will show, our method requires much fewer labels to yield comparable results, thus, leading to an improvement in accuracy, memory usage, and speed.

5.1. Rudin-Osher-Fatemi Model

As a proof of concept, we first evaluate the novel relaxation on the well-known Rudin-Osher-Fatemi (ROF) model [20]. It corresponds to (1) with the following dataterm:

\[
\rho(x, u(x)) = (u(x) - f(x))^2,
\] (25)

where \(f : \Omega \rightarrow \mathbb{R}\) denotes the input data. While there is no practical use in applying convex relaxation methods to an already convex problem such as the ROF model, the purpose of this is two-fold. Firstly, it allows us to measure the overhead introduced by our method by comparing it to standard convex optimization methods which do not rely on functional lifting. Secondly, we can experimentally verify that the relaxation is tight for a convex dataterm.

In Fig. 5 we solve (25) directly using the primal-dual algorithm [9], using the baseline functional lifting method [17] and using our proposed algorithm. First, the globally optimal energy was computed using the direct method with a very high number of iterations. Then we measure how long each method took to reach this global optimum to a fixed tolerance.

The baseline method fails to reach the global optimum even for 256 labels. While the lifting framework introduces a certain overhead, the proposed method finds the same globally optimal energy as the direct unlifted optimization approach and generalizes to nonconvex energies.

---

\(^1\)https://github.com/tum-vision/sublabel_relax

Figure 8. We compare the proposed relaxation with anisotropic regularizer to isotropic regularization on the stereo matching example. Using an anisotropic formulation as in [27] leads to grid bias.
5.2. Robust Truncated Quadratic Dataterm

The quadratic dataterm in (25) is often not well suited for real-world data as it comes from a pure Gaussian noise assumption and does not model outliers. We now consider a robust truncated quadratic dataterm:

$$\rho(x,u(x)) = \min \left\{ (u(x) - f(x))^2, \nu \right\}. \quad (26)$$

To implement (26), we use a piecewise polynomial approximation of the dataterm. In Fig. 6 we degraded the input image with additive Gaussian and salt and pepper noise. The parameters in (26) were chosen as $\alpha = 25$, $\nu = 0.025$ and $\lambda = 1$. It can be seen that the proposed method requires fewer labels to find lower energies than the baseline.

5.3. Comparison to the Method of Zach and Kohli

We remark that Prop. 4 and Prop. 5 hold for arbitrary convex one-homogeneous functionals $\phi(\nu)$ instead of $|\nu|^2$ in equation (16). In particular, they hold for the anisotropic total variation $\phi(\nu) = |\nu|_1$. This generalization allows us to directly compare our convex relaxation to the MRF approach of Zach and Kohli [27].

In Fig. 7 we show the results of optimizing the two models entitled “DC-Linear” and “DC-MRF” proposed in [27], and of our proposed method with anisotropic regularization on the robust truncated denoising energy (26). We picked the parameters as $\alpha = 0.2$, $\nu = 500$, and $\lambda = 1$. The label space is also chosen as $\Gamma = [0,256]$ as described in [27]. Note that overall, all the energies are better than the ones reported in [27]. It can be seen from Fig. 7 that the proposed relaxation is competitive to the one proposed by Zach and Kohli. In addition, the proposed relaxation uses a more compact representation and extends to isotropic and convex one-homogeneous regularizers. To illustrate the advantages of isotropic regularizations, Fig. 8a and Fig. 8b show a comparison of our proposed method for isotropic and anisotropic regularization for the example of stereo matching discussed in the next section.

5.4. Stereo Matching

Given a pair of rectified images, the task of finding a correspondence between the two images can be formulated as an optimization problem over a scalar field $u : \Omega \rightarrow \Gamma$ where each point $u(x) \in \Gamma$ denotes the displacement along the epipolar line associated with each $x \in \Omega$. The overall cost functional fits Eq. (1). In our experiments, we computed $\rho(x,u(x))$ for 270 disparities on the Middlebury stereo benchmark [21] in a $4 \times 4$ patch using a truncated sum of absolute gradient differences. We convexify the matching cost $\rho$ in each range $\Gamma_i$ by numerically computing the convex envelope using the gift wrapping algorithm.

The first row in Fig. 9 shows the result of the proposed relaxation using the convexified energy between two labels. The second row shows the baseline approach using the same amount of labels. Even for $L = 2$, the proposed method produces a reasonable depth map while the baseline approach basically corresponds to a two region segmentation.

5.5. Phase Unwrapping

Many sensors such as time-of-flight cameras or interferometric synthetic aperture radar (SAR) yield cyclic data ly-
ing on the circle $S^1$. Here we consider the task of total variation regularized unwrapping. As is shown on the left in Fig. 11, the dataterm is a nonconvex function where each minimum corresponds to a phase shift by $2\pi$:

$$\rho(x, u(x)) = d_{S^1}(u(x), f(x))^2.$$  

(27)

For the experiments, we approximated the nonconvex energy by quadratic pieces as depicted in Fig. 11. The label space is chosen as $\Gamma = [0, 4\pi]$ and the regularization parameter was set to $\lambda = 0.005$. Again, it is visible in Fig. 11 that the baseline method shows label space discretization and fails to unwrap the depth map correctly if the number of labels is chosen too low. The proposed method yields a smooth unwrapped result using only 8 labels.

### 5.6. Depth From Focus

In depth from focus the task is to recover the depth of a scene, given a stack of images each taken from a constant position but in a different focal setting, so that in each image only the objects of a certain depth are sharp. images. We compute the dataterm cost $\rho$ by using the modified Laplacian function [13] as a contrast measure.

Similar to the stereo experiments, we convexify the cost on each label range by computing the convex hull. The results are shown in Fig. 10. While the baseline method clearly shows the label space discretization, the proposed approach yields a smooth depth map. Since the proposed method uses a convex lower bound of the lifted energy, the regularizer has slightly more influence on the final result. This explains why the resulting depth maps in Fig. 10 and Fig. 9 look overall less noisy.

### 6. Conclusion

In this work we proposed a tight convex relaxation that can be interpreted as a sublabel–accurate formulation of classical multilabel problems. The final formulation is a simple saddle-point problem that admits fast primal-dual optimization. Our method maintains sublabel accuracy even after discretization and for that reason outperforms existing spatially continuous methods. Interesting directions for future work include higher dimensional label spaces, manifold valued data and more general regularizers.

**Acknowledgement** This work was supported by the ERC Starting Grant “ConvexVision”. We greatly acknowledge NVIDIA for providing us a Titan X GPU.
References


