# A Convex Discrete-Continuous Approach for Markov Random Fields

Christopher Zach and Pushmeet Kohli

Microsoft Research Cambridge, UK

**Abstract.** We propose an extension of the well-known LP relaxation for Markov random fields to explicitly allow continuous label spaces. Unlike conventional continuous formulations of labelling problems which assume that the unary and pairwise potentials are convex, our formulation allows them to be general piecewise convex functions with continuous domains. Furthermore, we present the extension of the widely used efficient scheme for handling $L^1$ smoothness priors over discrete ordered label sets to continuous label spaces. We provide a theoretical analysis of the proposed model, and empirically demonstrate that labelling problems with huge or continuous label spaces can benefit from our discrete-continuous representation.

## 1 Introduction

Energy minimization algorithms have been used to infer the Maximum a Posteriori (MAP) solutions of probabilistic formulations of many labeling problems encountered in Computer Vision. For instance, the Graph Cut algorithm, which enables the exact minimization of certain classes of pseudoboolean functions, has been used for building extremely efficient interactive tools for image segmentation [1]. Researchers, however, have not been able to replicate the success of discrete energy minimization algorithms, such as Graph Cuts and Message Passing, for building fast methods for solving labeling problems with very large or even continuous label sets such as image denoising and optical flow.

The above-mentioned problems, like many others encountered in Computer Vision, are defined over variables taking continuous labels. If the data term and the smoothness costs are convex functions, the underlying Markov random field (MRF) can be solved directly with convex optimization methods. However, this is not the case for many real world image labelling problems which have non-convex data terms and sometimes non-convex smoothness priors. For handling such problems, standard MRF algorithms discretize the label space, sample the energy costs associated with the chosen discrete label set, and then minimize the resulting energy which is defined over discrete variables.

The label discretization needed for handling continuous label problems can lead to arbitrarily large deviations from the true solution. The number of discrete labels needed to represent a continuous label set should to be large enough to provide a faithful solution to the original, continuous problem. Increasing the

number of labels implies substantial memory requirements, e.g. the memory consumption grows at least linearly with the number of labels, but can also grow quadratically (depending on the utilized formulation and optimization method). In particular, this often rules out e.g. leveraging graphics processing units to accelerate MAP estimation in such MRFs.

In this paper, we propose a convex discrete-continuous formulation of labeling problems. Our approach can compactly represent arbitrary data costs while still operating on a continuous label space. In this sense, it bring together the best of both discrete and continuous formulations. It is inspired from the observation that the data cost can be approximated as a lower envelope of convex basis functions. Fig. 1 illustrates the difference between (unary) potentials handled by well-established LP relaxations for MRFs and the potentials allowed in our approach. In this work the focus lies on how a more expressive potential function Fig. 1(b) can be used to efficiently approximate an arbitrary cost profile for a multi-label problem.

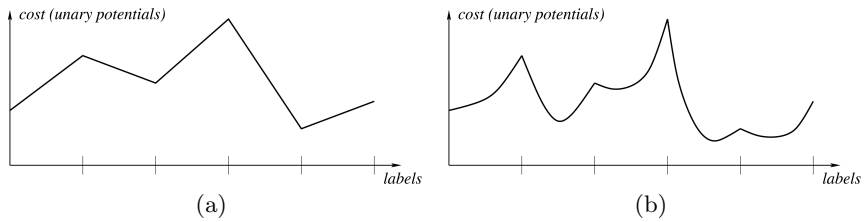



**Fig. 1.** Unary potentials as induced by standard LP relaxations (left) under a *special ordered set of type 2* interpretation [2], and one example that can be modeled using our discrete-continuous approach (right).

An outline of the paper follows. We discuss prior work on inference in MRFs with a focus on continuous label sets in section 2. Section 3 provides the mathematical notations and presents the main result enabling our convex discrete-continuous formulation for labeling problems, which is described in detail in section 4. We compare the performance of our method with traditional discrete optimization methods in section 5, and conclude in section 6 by listing some key conclusions and ideas for future work.

## 2 Related Work

There is a vast literature on different formulations and solution methods for Markov random fields. We refer to [3] for an extensive treatment of the respective statistical background, and e.g. [4] for an exposition of the corresponding LP relaxations. In certain cases the LP relaxation of discrete labeling problems provides globally optimal answers as discussed in [5, 6] for $L^1$ smoothness priors and in [7] for general pairwise potentials that are convex in the label difference.

The key observation that motivates our approach is, that unary and pairwise clique potentials for many continuous labeling problems encountered in the real world are not arbitrarily complex, but can be represented or at least well-approximated by piecewise convex functions[1]. Our proposed formulation generalizes the one described in [8], which completely convexifies the unary potentials in order to obtain (in combination with a convex smoothness prior) a convex problem over continuous labels. This method is a consequent continuation of the celebrated local linearization/convexification principle prominently utilized in the computation of optical flow (dating back to [9, 10]). The assumption of parametric or otherwise structured potentials in general also enables efficient inference methods as discussed in [11–13], where the focus lies on exploiting structured representation of higher-order clique potentials.

Continuous label spaces naturally raise the question of also using a continuous representation of the underlying domain (e.g. the 2D image domain). Recently, the analysis of the Mumford-Shah functional and its extensions in [14] proved to be a source of inspiration to address general labeling problems on image grids (e.g. [15]). [16, 17] propose to solve discretized versions of the underlying continuous model [14], thereby also handling continuous transitions of labels between neighboring pixels. Their formulation is based on a primal-dual saddlepoint problem and the underlying representation turns out to be substantially different to our proposed one.[2]

Peng et al [18] share their goal with the one of our work—inference in general MRFs with continuous label spaces—, but employ a completely different representation and inference procedure, which is based on convex belief propagation and particle sampling to represent beliefs over continuous labels. Finally, we want to mention that there are algorithmically very different approaches to tackle large or continuous label spaces, e.g. by adaptive refinement of the discrete label set as proposed in [19].

## 3 Notations and Preliminary Result

### 3.1 Notations

In the following we consider extended real-valued functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. The domain of $f$, $\mathrm{dom}(f)$, is $\{x \in \mathbb{R}^n : f(x) < \infty\}$, and we assume that $\mathrm{dom}(f) \neq \emptyset$. Hence, constraints on the feasible domain and infinite function values can be interchanged. The convex conjugate of $f$, denoted by $f^*$, is defined as $f^*(y) = \sup_x x^T y - f(x)$. The biconjugate $f^{**}$ is obtained by applying convex conjugation twice. It is known that for any function $f$ the convex conjugate $f^*$ is a lower-semicontinuous (l.s.c.) convex function, and $f^{**} = f$ iff $f$ is convex and l.s.c. Otherwise $f^{**}$ is the lower convex envelope of $f$, i.e. the supremum of all convex functions below the epigraph of $f$.

[1] This is also the main assumption of our framework.

[2] An in-depth discussion of the relation between the models is subject of an extended manuscript.

For a convex set $C$ we will use $\imath_C$ for its indicator function. i.e. $\imath_C(x) = 0$ for $x \in C$ and $\infty$ otherwise. We will also write e.g. $\imath\{f(x) \leq 0\}$ to convert a constraint (here $f(x) \leq 0$) into a function. For a convex function $f$ the corresponding perspective will be denoted by $f_\oslash$. It is defined indirectly as the convex conjugate $f_\oslash = (f_\oslash)^{**}$ of

$$(f_\oslash)^*(z, w) = \imath\{-z \geq f^*(w)\}.$$

For $x > 0$ we have $f_\oslash(x, y) = xf(y/x)$. We use the somewhat indirect definition above to handle also the case $x = 0$. Since $(f_\oslash)^*$ is an indicator function (i.e. a set of constraints), $f_\oslash$ is convex and positively 1-homogeneous, i.e. $f_\oslash(kx, ky) = kf_\oslash(x, y)$. Again, for $x > 0$ this can be verified directly from $kxf(ky/kx) = kxf(y/x)$, but this extends also to $x = 0$. If the domain of $f^*$ is $\mathbb{R}$ (or $\mathbb{R}^D$ if $f$ takes a vector argument), one obtains $f_\oslash(0, y) = \imath\{y = 0\}$ (we refer to the supplementary material).

In this work we focus on MRFs with pairwise interactions. Hence, the potentials will be defined on nodes $s \in \mathcal{V}$ and edges $(s, t) \in \mathcal{E}$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We will use $\sum_{s \sim t}$ as the easier-to-read replacement for $\sum_{s,t:(s,t)\in\mathcal{E}}$.

### 3.2 Minimizing a Family of Convex Functions

It is well known and has been extensively used in the literature, that the minimum of a finite number of values, e.g. $\{\theta^1, \ldots, \theta^n\}$ can be formulated as convex problem

$$\min_i \theta^i = \min_{\mathbf{x}} \sum_{i=1}^n \theta^i x^i = \min_{\mathbf{x}} \theta^T \mathbf{x} \tag{1}$$

subject to $x_i \geq 0$, $\sum_{i=1}^n x_i = 1$ (i.e. $\mathbf{x} \in \Delta_n$, the $n$-dimensional unit simplex). Note that the right hand side is a convex problem in $\mathbf{x}$. The equivalence above is essentially the basis of LP relaxations of Markov random fields, where $x$ is usually called the pseudo-marginal. Introducing the explicit weight vector $\mathbf{x}$ allows to add constraints linking unary and clique potentials. We extend this construction to a family of convex functions. Let $\{f^1, \ldots, f^n\}$ be a family of convex functions, $f^i : [0, 1] \to \mathbb{R}$ and we consider the minimization problem

$$\min_i \min_{\xi \in [0,1]} f^i(\xi). \tag{2}$$

The task is to reformulate this problem into a form explicitly incorporating a convex combination of weights, similar to Eq. 1. We will use the following reformulation in our construction:

**Observation 1** *We have that*

$$\min_i \min_{\xi \in [0,1]} f^i(\xi) = \min_{\mathbf{x} \in \Delta^n, \mathbf{y}: y_i \leq x_i} \sum_{i=1}^n f^i_\oslash(x_i, y_i) = \min_{\mathbf{x} \in \Delta^n, \mathbf{y}: y_i \leq x_i} \sum_{i=1}^n x_i f^i(y_i/x_i). \tag{3}$$

The proof is given in the supplementary material. At this point we want to add a few remarks:

- If $f^i(\xi) = \theta^i$ (constant functions), then Eq. 3 degenerates to Eq. 1. Surprisingly, the same is true if $f^i$ linearly interpolates between e.g. $\theta^i$ and $\theta^{i+1}$, i.e. $f^i(\xi) = \xi\theta^{i+1} + (1-\xi)\theta^i$ (like in Fig. 1(a)). This follows from the fact, that the minimum is (also) attained at one breakpoint for piecewise linear functions.
- If $p := \min_i \min_{\xi\in[0,1]} f^i(\xi)$ is not unique (with respect to $i$), then the minimizer $\mathbf{x}^*$ is not strictly binary, and for all $i$ with $x_i^* > 0$, one has $x_i^* f_i(y_i^*/x_i^*) = p$ for optimal $(\mathbf{x}^*, \mathbf{y}^*)$.

This construction also extends to families of functions $\{f^i\}_{i=1,\dots,n}$ with vector-valued arguments, i.e. $f^i : [0,1]^D \to \mathbb{R}$ with $D$ being the dimension of the domain. In this setting we have an analogy to Eq. 3,

$$\min_i \min_{\xi\in[0,1]^D} f^i(\xi) = \min_{\mathbf{x}\in\Delta^n, \mathbf{Y}:\mathbf{y}_i\le x_i} \sum_{i=1}^n f_\oslash^i(x_i, \mathbf{y}_i), \tag{4}$$

where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n) \in [0,1]^{D\times n}$ is now a $D \times n$ matrix instead of a $n$-vector, and $\mathbf{y}_i \le x_i$ is understood element-wise.

## 4 Convex Discrete-Continuous Formulation for MRFs

This section describes our combined discrete-continuous approach to represent labels for inference in MRFs. First, we review and reinterpret the standard local polytope relaxation for MRFs, and subsequently introduce our extensions.

Recall from section 1 that we are addressing continuous labeling problems, where a label value from the range $[0, L]$ is assigned to each node $s$, where $L$ is the number of segments to represent the piecewise convex potentials. We are going to represent the continuous label $l_s$ assigned at node $s$ as $l_s = \sum_i (i x_s^i + y_s^i)$, where $x_s^i \in \{0,1\}$ indicates the selected label range $[i, i+1]$ and $y_s^i \in [0,1]$ is the continuous offset within the subrange. Since exactly one subrange needs to be selected, we have $\sum_i x_s^i = 1$. Enforcing $x_s^i$ to be either 0 or 1 is hard, hence we relax $x_s^i \in \{0,1\}$ to $x_s^i \in [0,1]$. The restriction to label ranges $[0, L]$ is not limiting, since general bounded label domains can be scaled and translated without essentially changing the minimizer.

### 4.1 MAP Inference Using LP-MRF Relaxations Revisited

If one recalls the standard LP relaxation of discrete and labeling problem with at most pairwise potentials, assigning one out of $L$ states to each node,

$$E_{\text{LP-MRF}}(\mathbf{x}) = \sum_{s,i} \theta_s^i x_s^i + \sum_{s\sim t} \sum_{i,j} \theta_{st}^{ij} x_{st}^{ij} \tag{5}$$

subject to $x_s \in \Delta^L$, $x_{st} \in \Delta^{L^2}$, and the marginalization constraints,

$$\sum_j x_{st}^{ij} = x_s^i, \qquad \sum_i x_{st}^{ij} = x_t^j \qquad \forall s \sim t, \forall i, j, \tag{6}$$

then it is immediately clear that the objective is a sum of convex combinations of the clique potentials $\theta_s^i$ and $\theta_{st}^{ij}$, respectively. In the general higher-order MRF setting there is one such convex combination for each clique. Since the "barycentric" weights (or pseudo-marginals), $x_s^i$ and $x_{st}^{ij}$, are made explicit in the above formulation, they can be coupled via marginalization constraints, thereby connecting clique variables $x_{st}^{ij}$ with node marginals $x_s^i$. By recalling the extension described in Section 3.2 to find the global minimum of finite families of convex functions via explicit weights, it becomes imminent that $E_{\text{LP-MRF}}$ can be generalized in order to allow piecewise convex potentials.

### 4.2 Unary Potentials

The cost induced by unary potentials, $\sum_i \theta_s^i x_s^i$, is easy to generalize: we assume that real-valued convex functions $f_s^i : [0,1] \to \mathbb{R}$ are given, then the new unary cost reads as

$$\sum_s \sum_i (f_s^i)_\oslash(x_s^i, y_s^i) = \sum_s \sum_i x_s^i f_s^i(y_s^i / x_s^i),$$

where we have also introduced continuous variables $y_s^i \geq 0$ to represent the minimizer of $f_s^i$. Since $\operatorname{dom}(f_s^i) = [0,1]$ (i.e. $f_s^i(\xi) = \infty$ for $\xi \notin [0,1]$), we have additional implicit constraints $y_s^i \leq x_s^i$ (in addition to the unit simplex constraint on $x_s$). Since an assignment of an integral label value $i > 0$ at node $s$ is represented twice (as $x_s^i = 1$, $y_s^i = 0$, and alternatively as $x_s^{i-1} = 1$, $y_s^i = 1$), we require that $f_s^i(0) = f_s^{i-1}(1)$ (recall the graph in Fig. 1(b), which is continuous).

### 4.3 The General Discrete-Continuous Model

Generalizing the pairwise cost to continuous labels is far less trivial. In the following we present two models: the first formulation described in this section is the proper generalization of the LP relaxation Eq. 5 and introduces extended marginalization constraints. The second model discussed in Section 4.4 specifically addresses $L^1$ smoothness costs, and is much more efficient to optimize.

Assume that a family of convex functions $f_{st}^{ij} : [0,1]^2 \to \mathbb{R}$ is given. Like for the unary potentials we require that the induced graph of the smoothness cost is continuous, i.e. $f_{st}^{ij}(0,\cdot) = f_{st}^{i-1,j}(1,\cdot)$ and $f_{st}^{ij}(\cdot,0) = f_{st}^{i,j-1}(\cdot,1)$. We replace the standard pairwise cost for edge $(s,t)$,

$$\sum_{i,j} \theta_{st}^{ij} x_{st}^{ij}, \quad \text{by} \quad \sum_{i,j} (f_{st}^{ij})_\oslash \left( x_{st}^{ij}, y_{st\leftrightarrows}^{ij} \right) = \sum_{i,j} x_{st}^{ij} f_{st}^{ij} \left( \frac{y_{st\to s}^{ij}}{x_{st}^{ij}}, \frac{y_{st\to t}^{ij}}{x_{st}^{ij}} \right),$$

where we introduced local copies $y^{ij}_{st\to s}$ and $y^{ij}_{st\to t}$ of $y^i_s$ and $y^j_t$ only relevant in the scope of a label transition from $i$ to $j$. Further, we use $y^{ij}_{st\leftrightarrows}$ for the 2-vector $(y^{ij}_{st\to s}, y^{ij}_{st\to t})$. The values of $y^{ij}_{st\to s}$ and $y^{ij}_{st\to t}$, respectively, must be consistent with the node values, $y^i_s$ and $y^j_t$: we propose the following "decomposition" constraints,

$$y^i_s = \sum_j y^{ij}_{st\to s} \qquad \text{and} \qquad y^j_t = \sum_i y^{ij}_{st\to t}. \tag{7}$$

These constraints are not arbitrary, but define (together with the marginalization constraints on $x_{st}$ and the bounds constraints $0 \le y^{ij}_{st\to s} \le x^{ij}_{st}$, $0 \le y^{ij}_{st\to t} \le x^{ij}_{st}$) the convex hull of allowed configurations. We refer to the supplementary material for the derivation. By combining the unary costs from Section 4.2 and the extension for pairwise costs described above, one obtains the following discrete-continuous labeling model,

$$\begin{aligned} E_{\text{DC-MRF}}(\mathbf{x},\mathbf{y}) &= \sum_{s,i} (f^i_s)_\oslash(x^i_s, y^i_s) + \sum_{s\sim t}\sum_{i,j} (f^{ij}_{st})_\oslash(x^{ij}_{st}, y^{ij}_{st\leftrightarrows}) \\ \text{s.t. } & \sum_j x^{ij}_{st} = x^i_s,\ \sum_i x^{ij}_{st} = x^j_t,\ y^i_s = \sum_j y^{ij}_{st\to s},\ y^j_t = \sum_i y^{ij}_{st\to t} \\ & x_s \in \Delta^L,\ x_{st} \in \Delta^{L^2},\ 0 \le y^i_s \le x^i_s,\ 0 \le y^{ij}_{st\leftrightarrows} \le x^{ij}_{st}. \end{aligned} \tag{8}$$

The constraints $0 \le y^i_s \le x^i_s$ and $0 \le y^{ij}_{st\leftrightarrows} \le x^{ij}_{st}$ are actually redundant, since these constraints are implied by restricting $f^i_s$ and $f^{ij}_{st}$ to have domain $[0,1]$ and $[0,1]^2$, respectively. Since the underlying implicit domain constraints might be easily missed, we state these constraints on $\mathbf{y}$ here explicitly.

By comparing $E_{\text{LP-MRF}}$ (Eq. 5) with $E_{\text{DC-MRF}}$ it is clear that the latter one is a proper generalization of the LP relaxation and reduces to $E_{\text{LP-MRF}}$ for piecewise linear potentials $f^i_s$ and $f^{ij}_{st}$. The downside of this model is, that it is expensive to optimize. Unlike $E_{\text{LP-MRF}}$, which can be efficiently optimized by dual coordinate methods (usually referred as message passing [20, 4, 21, 22]), the dual of $E_{\text{DC-MRF}}$ is more involved and may have $O(L^2)$ variables per node (depending whether the $0 \le y^{ij}_{st\leftrightarrows} \le x^{ij}_{st}$ constraints are moved into $f^{ij}_{st}$ or not).

It is natural to ask whether the model in Eq. 8 returns globally optimal results for arbitrary unary potentials and convex pairwise priors as it is the case for the LP relaxation Eq. 5 [7]. Unfortunately, the answer seems to be negative as indicated by our experiments. Global optimality is currently guaranteed only for piecewise linear unary potentials and piecewise linear and convex pairwise priors, and to globally convex unary and pairwise priors, i.e. the original problem in terms of continuous labels assigned to nodes is convex. While in such cases it is not reasonable to artificially subdivide the label range into $L > 1$ segments and to optimize $E_{\text{DC-MRF}}$, it is at least encouraging that the minimizer of $E_{\text{DC-MRF}}$ returns the same solution as the original convex problem. Hence, the model in Eq. 8 does not weaken the optimal solution.

We will give a sketch of the proof: let $E_1(\mathbf{z}) = \sum_s F_s(\mathbf{z}) + \sum_{s\sim t} G_{st}(\mathbf{z})$ be the original problem, with $F_s$ and $G_{st}$ convex functions corresponding to the

unary and pairwise costs, respectively. For a given $\mathbf{z}$ we can determine feasible arguments for $E_{\text{DC-MRF}}$ via e.g.

$$x_s^i(\mathbf{z}) = \begin{cases} 1 & \text{if } z_s \in [i, i+1) \\ 0 & \text{otherwise} \end{cases} \qquad y_s^i(\mathbf{z}) = \begin{cases} z_s - i & \text{if } z_s \in [i, i+1) \\ 0 & \text{otherwise.} \end{cases}$$

Further, $f_s^i(\xi) = F_s(\xi + i)$ and $f_{st}^{ij}(\xi_s, \xi_t) = G_{st}(\xi_s + i, \xi_t + j)$. Let $\mathbf{z}^*$ a minimizer of $E_1$ and $(\mathbf{x}^*, \mathbf{y}^*)$ the minimizer of the corresponding discrete-continuous program $E_{\text{DC-MRF}}$ Eq. 8. Since $\mathbf{z}^*$ can be converted into a feasible argument $(\mathbf{x}(\mathbf{z}^*), \mathbf{y}(\mathbf{z}^*))$ of $E_{\text{DC-MRF}}$, we have $E_1(\mathbf{z}^*) \geq E_{\text{DC-MRF}}(\mathbf{x}^*, \mathbf{y}^*)$. On the other hand, $E_{\text{DC-MRF}}$ consists only of convex combinations (subject to marginalization constraints) of samples from convex functions, hence by Jensen's inequality we have (for the unary potentials)

$$\sum_i (f_s^i)_\oslash(x_s^i, y_s^i) = \sum_i x_s^i f_s^i(y_s^i / x_s^i) \geq F_s\left(\sum_i \left(i x_s^i + y_s^i\right)\right).$$

A similar inequality holds for the pairwise terms. Overall we obtain $E_1(\mathbf{z}^*) \leq E_{\text{DC-MRF}}(\mathbf{x}^*, \mathbf{y}^*)$ and together with $E_1(\mathbf{z}^*) \geq E_{\text{DC-MRF}}(\mathbf{x}^*, \mathbf{y}^*)$ equality holds. Experimentally we confirm this fact in Fig. 2, which shows the results of image denoising using the (anisotropic) Rudin-Osher-Fatemi (ROF) model [23],

$$E_{ROF}(u;\, f) = \frac{\lambda}{2} \sum_s (u_s - f_s)^2 + \sum_{s \sim t} |u_s - u_t|. \tag{9}$$

Since the data term in the ROF energy is quadratic, $f_s^i$ are quadratic and given by $f_s^i(\xi) = \lambda(i + \xi - f_s)^2/2$ with $\xi \in [0, 1]$. The $L^1$ smoothness term is slightly more involved:

$$f_{st}^{ij}(\xi_s, \xi_t) = \begin{cases} |\xi_s - \xi_t| & \text{if } i = j \\ \pm(i - j + \xi_s - \xi_t) & \text{if } i \gtrless j \end{cases} \tag{10}$$

If e.g. $i > j$ we have $z_s = i + \xi_s > z_t = j + \xi_t$, and the $L^1$ smoothness cost is therefore $|z_s - z_t| = i - j + \xi_s - \xi_t$. The other cases are similar.

### 4.4 Generalizing the "battleship" construction for $L^1$ smoothness

The LP relaxation of standard discrete labeling problems with general pairwise smoothness costs (recall Eq. 5) requires to maintain $O(L^2)$ variables per node (with respect to the label subrange count $L$) in the primal domain (the pairwise pseudo-marginals $x_{st}^{ij}$). The constructions presented in [5, 6][3] maintain only $O(L)$ unknowns per node. In order to understand the construction for $L^1$ smoothness in our discrete-continous approach, the properties of Eq. 5 must be analyzed for

[3] which is sometimes called the "battleship" construction due to the graph structure

(a) $E_{ROF} = 143992$ (b) $E_{ROF} = 144002$ (c) $E_{ROF} = 565809$ (d) $E_{ROF} = 565973$

**Fig. 2.** Visual and energy comparison between direct optimization [24] of the ROF image denoising model (a, c), and artificially subdividing the label space $[0, 255]$ into 4 even ranges and subsequent optimization of $E_{\text{DC-MRF}}$ (b, d). The small differences in the obtained energy are likely due to the ultimately slow convergence of minimizing $E_{\text{DC-MRF}}$ and due to numerical inaccuracies in require cube root computations.

this particular choice of pairwise potentials. The corresponding LP relaxation reads as

$$E_{\text{LP-MRF}-L^1}(\mathbf{x}) = \sum_{s,i} \theta_s^i x_s^i + \sum_{s \sim t} \sum_{i,j} |i-j| x_{st}^{ij}. \tag{11}$$

subject to $x_s \in \Delta^L$, $x_{st} \in \Delta^{L^2}$, and the marginalization constraints Eq. 6. The constraint $\sum_{i,j} x_{st}^{ij} = 1$ is redundant and can be dropped, thus $x_{st} \in \Delta^{L^2}$ can be replaced by $x_{st}^{ij} \geq 0$. The link between Eq. 11 and the battleship construction is, that $x_{st}^{ii}$ is eliminated from the marginalization constraints and the corresponding non-negativity constraint is dropped.[4]

Via Fenchel duality it can be shown that the capacity constraints in the dual of Eq. 11 are given by $p_{st}^i + p_{ts}^j \leq |i-j|$ and $p_{st}^i - p_{ts}^i = 0$, respectively, for dual variables $p_{st}^i$, $p_{ts}^i$, representing the directed flow over edge $(s,t)$. Since $p_{st}^i - p_{ts}^i = 0$, $p_{ts}^i$ can be eliminated, leaving capacity constraints $p_{st}^i - p_{st}^j \leq |i-j|$. Since $|\cdot|$ is a metric, all other capacity constraints follow from $p_{st}^i - p_{st}^{i+1} \leq 1$ and $p_{st}^i - p_{st}^{i-1} \leq 1$, i.e. only these constraints are necessary. By going back to the primal program, this translates to the following objective,

$$\begin{aligned} E_{\text{LP-Linear}}(\mathbf{x}) = \sum_{s,i} \theta_s^i x_s^i + \sum_{s \sim t} \sum_{i} \left( x_{st}^{i,i+1} + x_{st}^{i,i-1} \right) \qquad (12) \\ \text{s.t. } x_s^i = x_{st}^{ii} + x_{st}^{i,i+1} + x_{st}^{i,i-1} \qquad x_t^i = x_{st}^{ii} + x_{st}^{i-1,i} + x_{st}^{i+1,i} \\ x_s \in \Delta^L,\, x_{st}^{i,i+1} \geq 0, x_{st}^{i,i-1} \geq 0. \end{aligned}$$

It is possible to further simplify Eq. 12 to obtain the minimum-cut formulation presented in [5, 6], but the above formulation is the basis for our construction.

[4] Dropping $x_{st}^{ii} \geq 0$ converts general pairwise potentials $\theta_{st}^{ij}$ into a metric, hence nothing is lost whenever $\theta_{st}^{ij}$ is already a metric. A more detailed discussion of these relations can be found in [25].

Before we continue, it is instructive to see, how a jump e.g. from label $i$ to $i+3$ at edge $(s,t)$ is actually represented in the above program: one has $x_{st}^{i,i+1} = 1$, $x_{st}^{i+1,i+1} = -1$, $x_{st}^{i+1,i+2} = 1$,$x_{st}^{i+2,i+2} = -1$ and finally $x_{st}^{i+2,i+3} = 1$, leading the the correct smoothness cost of 3 and still satisfying all marginalization constraints. Thus, label discontinuities along an edge $(s,t)$ correspond to chains of $+1$ for $x_{st}^{i,i+1}$ (or $x_{st}^{i,i-1}$) and $-1$'s at respective $x_{st}^{ii}$ to fulfill the marginalization constraints for in-between labels.

In view of the energy Eq. 12 we only need to specify pairwise potentials for transitions $x_{st}^{ii}$, $x_{st}^{i,i+1}$, and $x_{st}^{i,i-1}$. These potentials are similar to the ones given in Eq. 10 and read as

$$f_{st}^{ii}(\xi_s,\xi_t) = |\xi_s - \xi_t| \qquad\qquad f_{st}^{i,i\pm1}(\xi_s,\xi_t) = 1 \pm (\xi_t - \xi_s). \tag{13}$$

Further, since $x_{st}^{ii}$ can be negative, we split $x_{st}^{ii}$ into non-negative and non-positive components, i.e. we replace $x_{st}^{ii}$ by $x_{st}^{ii} - \bar{x}_{st}^{ii}$ with $x_{st}^{ii} \geq 0$ and $\bar{x}_{st}^{ii} \geq 0$. After combining everything the full program reads as

$$E_{\text{DC-Linear}}(\mathbf{x},\mathbf{y}) = \sum_{s,i}(f_s^i)_\oslash(x_s^i,y_s^i) + \sum_{s\sim t}\sum_i \sum_{k\in\{0,\pm1\}} (f_{st}^{i,i+k})_\oslash(x_{st}^{i,i+k}, y_{st\leftrightarrows}^{i,i+k}) \tag{14}$$

$$\begin{aligned} \text{s.t. } x_s^i &= x_{st}^{ii} + x_{st}^{i,i+1} + x_{st}^{i,i-1} - \bar{x}_{st}^{ii}, \quad x_t^i = x_{st}^{ii} + x_{st}^{i-1,i} + x_{st}^{i+1,i} - \bar{x}_{st}^{ii}, \\ y_s^i &= y_{st\to s}^{ii} + y_{st\to s}^{i,i+1} + y_{st\to s}^{i,i-1}, \qquad y_t^i = y_{st\to t}^{ii} + y_{st\to t}^{i-1,i} + y_{st\to t}^{i+1,i}, \\ & x_s \in \Delta^L,\ y_s^i \in [0,x_s^i],\ x_{st}^{i,i+\{0,\pm1\}} \geq 0,\ y_{st\leftrightarrows}^{i,i+\{0,\pm1\}} \in \left[0, x_{st}^{i,i+\{0,\pm1\}}\right]. \end{aligned}$$

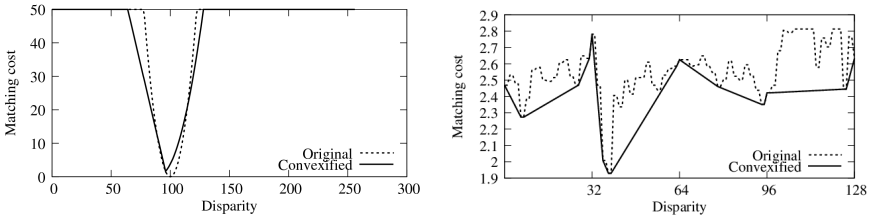
Unlike $E_{\text{LP-Linear}}$, which is equivalent to $E_{\text{LP-MRF}-L^1}$, we do not yet know whether $E_{\text{DC-Linear}}$ and $E_{\text{DC-MRF}}$ are equivalent for $L^1$ smoothness. We observed in our experiments, that minimizers of $E_{\text{DC-Linear}}$ have a slightly higher (true) energy than solutions of $E_{\text{DC-MRF}}$, but the numerical results are not fully conclusive.

## 5 Numerical Results

In our numerical experiments we compare several methods: the baseline method is the LP relaxation with a natural number of labels (256 for grayscale image processing, 64 for dense stereo computation). Since we have $L^1$ smoothness priors, we solve a respective minimum cut problem to find the minimizer. We call the result "baseline" rather than ground truth, since the continuous label space is still discretized. In order to judge the benefits of using convex potentials over constant ones, we also apply the minimum cut on compressed sets of labels, where the unary potentials are averaged ("GC-Avg") or subsampled versions of the original ones ("GC-Sampled"). Further, the impact of replacing the original unary potentials by convex segments is demonstrated by running the baseline method using convexified unary costs ("GC-Convex"). The full general model described in Section 4.3 is denoted as "DC-MRF", and the construction for $L^1$ smoothness priors is depicted as "DC-Linear" in the following. Comparing the

results of DC-MRF/DC-Linear and the ones of GC-Convex indicate the tightness of our proposed convex models.

We implemented prototype algorithms for finding minimizers of $E_{\text{DC-MRF}}$ for $L^1$ smoothness priors and $E_{\text{DC-Linear}}$ using the simultaneous-direction method of multipliers (SDMM, see e.g. [26] for a comprehensive overview of proximal methods for non-smooth convex optimization). General unary potentials are transformed into a convex surrogate in each label range by fitting a quadratic polynomial, such that the unary costs at the end-points are preserved. If the fitted polynomial is concave, a linear approximation is used. Fig. 3 depicts two cost profiles, a truncated quadratic data term of a robust ROF model and the matching costs used for stereo, and the corresponding piecewise convex surrogates.




(a) Robust ROF (b) Stereo

**Fig. 3.** Fitting piecewise convex quadratic unary potentials (a) and lower envelopes (b) to given data costs.

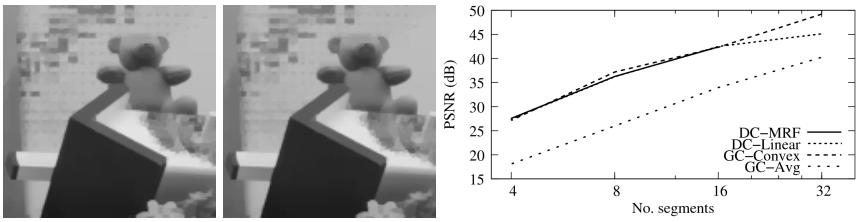
### 5.1 Robust ROF Energy

In order to determine the influence of reducing the number of ranges in the label space, we utilize a robust version of the ROF energy, that replaces the quadratic data term by the non-convex truncated one,

$$E_{\text{robust-ROF}}(u;\, f) = \frac{\lambda}{2} \sum_s \min\left\{(u_s - f_s)^2, T\right\} + \sum_{s\sim t} |u_s - u_t|. \qquad (15)$$

Our choice of $T$ is 500, hence deviations of about 22 grayscale units are considered as outliers. $\lambda$ is set to $1/5$. Figure 4(a) illustrates the baseline (reference) result, and the minimizer of $E_{\text{DC-MRF}}$ using 8 ranges for the label space $[0, 255]$ is shown in Fig. 4(b). Further, a comparison in terms of PSNR between the baseline and the results of the tested methods with varying numbers of ranges is displayed in Fig. 4(c). From the graph we can learn two things: first, the discrete-continuous models DC-MRF and DC-Linear are very close to the 256-label result with convexified data term. Second, DC-MRF and DC-Linear are far superior to naively reducing the number of discrete labels as in GC-Avg and GC-Sampled. A full visual illustration of the obtained results and corresponding

values of $E_{\text{robust-ROF}}$ are provided in the supplementary material. The supplementary material also illustrates the evolution of energies with respect to CPU time for the baseline approach and for the proposed discrete-continuous formulation. In a nutshell minizming $E_{\text{DC-Linear}}$ quickly reaches an energy close to the converged value for the robust ROF problem.




(a) Baseline (b) DC-MRF (c) PSNR

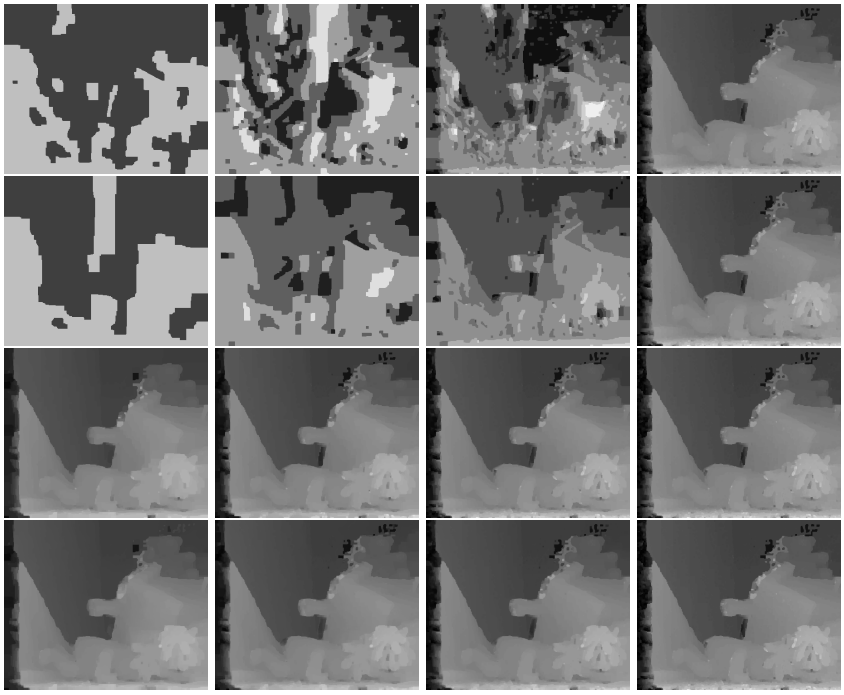
**Fig. 4.** Visual comparison between the baseline solution (a) and the DC-MRF approach (b) for the robust ROF denoising model. (c) illustrates the image differences to the baseline solution for different methods. We were not able to minimize the full $E_{\text{DC-MRF}}$ for 32 subranges due to memory consumption.

### 5.2 Dense Stereo

Since the unary potentials of the robust ROF energy are very structured, we illustrate our discrete-continuous MRF approach for the non-parametric data term induced by stereo matching costs. We use the truncated $L^1$ difference between $3 \times 3$ patches after Sobel-filtering the (grayscale) input images, since this is a realistic similarity measure for stereo. The truncation threshold is set to 20. Since the full energy $E_{\text{DC-MRF}}$ is too expensive to optimize, we report visual results for DC-Linear in Fig. 5. Following [8] we use a non-parametric convex envelope to obtain a piece-wise convex approximation to the original matching costs. The supplementary material also shows results for a parametric convex fit using piece-wise quadratic surrogates.

## 6 Summary and Future Work

We presented a novel, discrete-continuous formulation to solve Markov random fields over continuous label spaces. The proposed energy model is a proper extension of widely-used relaxations for labeling problems. While the proposed convex approach is not a perfectly tight relaxation in general, preliminary numerical results are promising. There are several directions for future work: most importantly, finding more efficient algorithms than our prototype implementation to minimize the DC-MRF energy will enable applications for larger-scale problems.



(a) 2 levels (b) 4 levels (c) 8 levels (d) Baseline

**Fig. 5.** Visual comparison for computational stereo using an $L^1$ smoothness prior. The columns correspond to numbers of used segment to represent the range $[0, 128]$. The rightmost column always shows the (globally optimal) result of a max-flow approach using 128 discrete labels for reference. 1st row: max-flow with sampled unary potentials. 2nd row: max-flow with averaged unary potentials. 3rd row: full 128-label maxflow, but with convexified unary potentials in each subrange. 4th row: minimizers of $E_{\text{DC-Linear}}$. The differences between the 3rd and the 4th row illustrate how much a convexified unary potential affects the minimizer and what is lost due to the compressed discrete-continuous representation in $E_{\text{DC-Linear}}$.

Further, an improved theoretical analysis of the construction described in Section 4.4 and respective generalizations are of high relevance. Finally, avoiding the grid bias in the solution and using a more isotropic regularization approach are beneficial in low-level vision applications and subject of future research.

## References

1. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proc. ICCV. (2001) 105–112
2. Beale, E.M.L., Forrest, J.J.H.: Global optimization using special ordered sets. Mathematical Programming **10**(1) (1976) 52–69

3. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. **1** (2008) 1–305
4. Werner, T.: A linear programming approach to max-sum problem: A review. IEEE Trans. Pattern Anal. Mach. Intell. **29**(7) (2007)
5. Ishikawa, H., Geiger, D.: Occlusions, discontinuities, and epipolar lines in stereo. In: Proc. ECCV. (1998) 232–248
6. Boykov, Y., Veksler, O., Zabih, R.: Markov random fields with efficient approximations. In: Proc. CVPR. (1998) 648–655
7. Ishikawa, H.: Exact optimization for Markov random fields with convex priors. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10) (2003) 1333–1336
8. Bhusnurmath, A., Taylor, C.J.: Stereo matching using interior point methods. In: Proc. 3DPVT. (2008)
9. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Int. Joint Conf. on Artificial Intell. (1981) 674–679
10. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence **17** (1981) 185–203
11. Rother, C., Kohli, P., Feng, W., Jia, J.: Minimizing sparse higher order energy functions of discrete variables. In: Proc. CVPR. (2009)
12. Kohli, P., Kumar, M.P.: Energy minimization for linear envelope MRFs. In: Proc. CVPR. (2010)
13. Komodakis, N., Paragios, N.: Beyond pairwise energies: Efficient optimization for higher-order MRFs. In: Proc. CVPR. (2009)
14. Alberti, G., Bouchitté, G., Maso, G.D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calc. Var. Partial Differential Equations **16**(3) (2003) 299–333
15. Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. Technical report, Ecole Polytechnique (2008)
16. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: Proc. ICCV. (2009)
17. Strekalovskiy, E., Goldluecke, B., Cremers, D.: Tight convex relaxations for vector-valued labeling problems. In: Proc. ICCV. (2011)
18. Peng, J., Hazan, T., McAllester, D., Urtasun, R.: Convex max-product algorithms for continuous MRFs with applications to protein folding. In: Proc. ICML. (2011)
19. Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., Navab, N.: Optical flow estimation with uncertainties through dynamic MRFs. In: Proc. CVPR. (2008)
20. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(10) (2006) 1568–1583
21. Globerson, A., Jaakkola, T.: Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In: NIPS. (2007)
22. Weiss, Y., Yanover, C., Meltzer, T.: MAP estimation, linear programming and belief propagation with convex free energies. In: UAI. (2007)
23. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
24. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imag. Vision **20**(1–2) (2004) 89–97
25. Zach, C., Häne, C., Pollefeys, M.: What is optimized in convex relaxations for multi-label problems: Connecting discrete and continuously-inspired MAP inference. Technical report, Microsoft Research (2012)
26. Combettes, P.L., Pesquet, J.C.: Proximal Splitting Methods in Signal Processing. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer (2011) 185–212

# A Convex Discrete-Continuous Approach for Markov Random Fields ECCV 2012 Supplementary Material

Christopher Zach and Pushmeet Kohli
Microsoft Research Cambridge

## Contents



## 1 Proof of Observation 1

One direct way to prove the observation is to introduce $\xi_i := y_i/x_i \in [0, 1]$ and restate

$$\min_{\mathbf{x}\in\Delta^n, \mathbf{y}: y_i \le x_i} \sum_{i=1}^n x_i f^i(y_i/x_i) = \min_{\mathbf{x}\in\Delta^n, \mathbf{t}: \xi_i \in [0,1]} \sum_{i=1}^n x_i f^i(\xi_i).$$

For each assignment of $\xi$, $\min_{\mathbf{x}\in\Delta^n} \sum_{i=1}^n x_i f^i(\xi_i)$ is just $\min_i f^i(\xi_i)$. By minimizing over all $\xi \in [0, 1]^n$, we have the claimed result.

It is instructive to derive the observation also via duality. If we augment the perspective of $f^i$ with the constraints $x_i \ge 0$ and $y_i \le x_i$, i.e. we define temporarily

$$f_i(x, y) := x f^i(x/y) + \iota_{\ge 0}(x) + \iota_{\ge 0}(x - y),$$

we obtain for the convex conjugate

$$\begin{aligned} f_i^*(z,w) &= \max_{x,y:x\geq 0, y\leq x} xz + yw - xf^i(x/y) && [\xi := y/x] \\ &= \max_{x\geq 0,\xi\in[0,1]} xz + x\xi w - xf^i(\xi) \\ &= \max_{x\geq 0} x\Big\{ z + \underbrace{\max_{\xi\in[0,1]} w - f^i(\xi)}_{=(f^i)^*(w)} \Big\} \\ &= \max_{x\geq 0} x(z + (f^i)^*(w)) = \imath\{-z \geq (f^i)^*(w)\}. \end{aligned}$$

Applying Fenchel duality on the primal problem

$$\underbrace{\sum_i x_i f^i(y_i/x_i) + \imath_{\geq 0}(x) + \imath_{\geq 0}(x-y)}_{=:F(\mathbf{x},\mathbf{y})} + \underbrace{\imath\{\sum_i x_i = 1\}}_{=:G(A[\mathbf{x},\mathbf{y}])}$$

(with $A = (\overbrace{1,\dots,1}^{n \text{ times}}, \overbrace{0,\dots 0}^{n \text{ times}}) \in \mathbb{R}^{1\times 2n}$) yields the corresponding dual $-F^*(A^T\mathbf{z}) - G^*(-\mathbf{z}) \to \max_{\mathbf{z}}$ as

$$\max z \text{ subject to } z \leq -(f^i)^*(0),$$

but $-(f^i)^*(0) = -\max_t\{0 - f^i(\xi)\} = \min_\xi f^i(\xi)$. Consequently, the constraints read as $z \leq \min_\xi f^i(\xi)$. Since the dual objective is attained at $\min_i \min_\xi f^i(\xi)$ and we have strong duality, the primal optimal value of the r.h.s. in Eq. 3 is as claimed in the observation.

# 2 Derivation of the Decomposition Constraints

In this section we are going to motivate the proposed constraints in Eq. 7. We are focussing on a node $s$ and an adjacent edge $(s,t)$, and consider the relation between $x_s^i$, $x_{st}^{ij}$, $y_s^i$, and $y_{st\leftarrow}^{ij}$ for a particular label $i$. We drop the subscripts in the following and will also use shorthand notations $\mathbf{x}^{i\cdot} := (x^{i1},\dots,x^{iL})$ and $\mathbf{y}^{i\cdot} := (y^{i1},\dots,y^{iL})$. We define $\mathcal{S}$ as the set of feasible configurations $(x^i,\mathbf{x}^{i\cdot},y^i,\mathbf{y}^{i\cdot}) \in [0,1]^{2L+2}$:

$$\mathcal{S} := \{\mathbf{0}_{2L+2}\} \cup \{(1,\mathbf{e}_j,y,y\mathbf{e}_j) : y \in [0,1]\}, \tag{1}$$

where $\mathbf{e}_j$ is the $L$-dimensional (row) vector with 1 at element $j$ and 0 otherwise. $\mathbf{0}_{2L+1}$ corresponds to the configuration, that label $i$ is not selected at node $s$ (i.e. $x^i$ and all $x^{ij}$ are 0). In this case we (somewhat arbitrarily, but this is consistent with the l.s.c. extension of the perspective) enforce $y^i = y^{ij} = 0$. If label $i$ is selected at $s$, and some label $j$ is picked at $t$, we have $x^{ij} = 1$ and $x^{ik} = 0$ for $k \neq j$, corresponding to $\mathbf{e}_j$. Again, we enforce $y^{ik} = 0$ whenever $x^{ik} = 0$. Since $x^{ij} = 1$, $y^{ij}$ is any value in $[0,1]$ and also determines $y^i = y^{ij}$. This provides the other feasible configurations in $\mathcal{S}$. Consider the convex set $\mathcal{C}$ with

$$\mathcal{C} := \Big\{(x^i,\mathbf{x}^{i\cdot},y^i,\mathbf{y}^{i\cdot}) : x^{ij} \geq 0, \textstyle\sum_j x^{ij} = x^i,\ 0 \leq y^{ij} \leq x^{ij},\ y^i = \sum_j y^{ij}\Big\}. \tag{2}$$

It can be easily verified that $\mathcal{S} \subseteq \mathcal{C}$, and consequently the convex hull of $\mathcal{S}$, $\mathrm{conv}(\mathcal{S})$, is contained in $\mathcal{C}$. It remains to show that $\mathcal{C} \subseteq \mathrm{conv}(\mathcal{S})$, which we demonstrate by writing any element of $\mathcal{C}$ as convex combination of elements in $\mathcal{S}$. Let $(x^i,\mathbf{x}^{i\cdot},y^i,\mathbf{y}^{i\cdot}) \in \mathcal{C}$. We define weights

$$\lambda_j = x^{ij} \qquad \lambda_{\mathbf{0}} = 1 - \sum_{j=1}^{L} \lambda_j \qquad \mu_j = \begin{cases} 1/x^{ij} & \text{if } x^{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

and write $(x^i, \mathbf{x}^{i\cdot}, y^i, \mathbf{y}^{i\cdot})$ as convex combination

$$(x^i, \mathbf{x}^{i\cdot}, y^i, \mathbf{y}^{i\cdot}) = \sum_{j=1}^{L} \lambda_j \underbrace{(1, \mathbf{e}_j, \mu_j y^{ij}, \mu_j y^{ij} \mathbf{e}_j)}_{\in \mathcal{S},\ \text{since } \mu_j y^{ij} \in [0,1]} + \lambda_{\mathbf{0}} \underbrace{\mathbf{0}}_{\in \mathcal{S}}.$$

This shows $\mathcal{C} \subseteq \operatorname{conv}(\mathcal{S})$ and overall we have $\mathcal{C} = \operatorname{conv}(\mathcal{S})$.

# 3 The Biconjugate of the Perspective

Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. The perspective of $f$ is defined as $f : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, $f(x, y) : (x, y) \mapsto xf(y/x)$ with $x > 0$. As shown in the proof of Observation 1, the convex conjugate of $f$ is given by $f^*(z, w) = \iota\{-z \geq f^*(w)\}$. The biconjugate of $f$, $f^{**}$, can be derived as follows:

$$f^{**}(x, y) = \max_{\substack{w \in \operatorname{dom}(f^*) \\ -z \leq f^*(w)}} \left\{xz + yw\right\}$$

by definition. We made the domain constraint on $w$ explicit, since it is important whenever $\operatorname{dom}(f^*) \subset \mathbb{R}$. We assume $\operatorname{dom}(f^*) \neq \emptyset$. If $x < 0$, we can make the term $xz$ arbitrarily large by letting $-z \to \infty$. Hence, the domain of $f^{**}$ will have the constraint $x \geq 0$. If $x > 0$, $-z$ should be as large as possible, i.e. $-z = f^*(w)$, and we have in this case

$$f^{**}(x, y) = \max_{w \in \operatorname{dom}(f^*)} -xf^*(w) + yw = x \max_{w \in \operatorname{dom}(f^*)} \left\{-f^*(w) + \frac{y}{z} w\right\} = xf(y/x).$$

If $x = 0$, one obtains

$$f^{**}(0, y) = \max_{w \in \operatorname{dom}(f^*)} yw = \sigma_{\operatorname{dom}(f^*)}(y),$$

where $\sigma_C(y) = \sup_{w \in C} yw$ is the support function of a set $C$. If $\operatorname{dom}(f^*) = \mathbb{R}$, then $f^{**}(0, y) = \iota\{y = 0\}$. Combining everything we have for the biconjugate of the perspective

$$f^{**}(x, y) = \begin{cases} xf(y/x) & \text{if } x > 0 \\ \sigma_{\operatorname{dom}(f^*)}(y) & \text{if } x = 0 \\ \infty & \text{if } x < 0. \end{cases}$$

Since we assume a bounded domain for $f$ in this manuscrips ($f : [0, 1] \to \mathbb{R}$), we always have $\operatorname{dom}(f^*) = \mathbb{R}$ and $\sigma_{\operatorname{dom}(f^*)} = \iota_{\{0\}}(y)$.

# 4 Proximity Operators for Some Perspectives

We use SDMM to find a minimizer of our proposed energies. We refer to [1] for an extensive overview of proximal methods for non-smooth convex optimization. In a nutshell SDMM applies proximity operators on each of the terms in the energy and updates respective Lagrange multipliers, which makes it a generalization of augmented Lagrangian methods. The core of the method, the proximity operator, is defined as

$$\operatorname{prox}_f(\bar{\mathbf{x}}) = \arg\min_{\mathbf{x}} \frac{1}{2} \left\|\mathbf{x} - \bar{\mathbf{x}}\right\|^2 + f(\mathbf{x}). \tag{3}$$

Since the minimizer can be shown to be always unique, $\operatorname{prox}_f$ is a proper function. Note that the proximity operator generalizes the projection operation into a convex set (i.e. finding the closest point to a given one in a convex set) to general convex functions.

In the following we derive some proximity operators for perspectives used in the main text. We will make extensive use of Moreau's decomposition, $x = \operatorname{prox}_f(x) + \operatorname{prox}_{f^*}(x)$ or $x = \operatorname{prox}_{\gamma f}(x) + \gamma \operatorname{prox}_{f^*/\gamma}(x/\gamma)$ for $\gamma > 0$, to derive the proximity operators. Thus, we can avoid the complications of perspectives at $x = 0$.

## 4.1 $f(y) = ay + b$

We have $f^*(w) = \imath\{w = a\} - b$ and

$$(f_\oslash)^*(z, w) = \imath\big\{-z \geq \imath\{w = a\} - b\big\} = \imath\{z \leq b,\, w = a\}.$$

Consequently,

$$\operatorname{prox}_{\gamma(f_\oslash)^*}(\bar{z}, \bar{w}) = \begin{pmatrix} \Pi_{[-\infty,b]}(\bar{z}) \\ a \end{pmatrix}$$

and therefore (by Moreau's decomposition)

$$\begin{aligned} \operatorname{prox}_{\gamma f_\oslash}(\bar{x}, \bar{y}) &= \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} - \gamma \begin{pmatrix} \Pi_{[-\infty,b]}(\bar{x}/\gamma) \\ a \end{pmatrix} = \begin{pmatrix} \bar{x} - \min\{\gamma b, \bar{x}\} \\ \bar{y} - \gamma a. \end{pmatrix} \\ &= \begin{pmatrix} \max\{0, \bar{x} - \gamma b\} \\ \bar{y} - \gamma a. \end{pmatrix} \end{aligned}$$

## 4.2 $f(y) = \frac{1}{2}ay^2 + by + c$

Since $f$ is assumed to be convex, we have $a > 0$ (the case $a = 0$ is addressed in the previous section). By elementary calculus we obtain

$$f^*(w) = \frac{1}{2a}(w - b)^2 - c,$$

and therefore

$$(f_\oslash)^*(z, w) = \imath\Big\{-z \geq \frac{1}{2a}(w - b)^2 - c\Big\} = \imath\Big\{z \leq c - \frac{1}{2a}(w - b)^2\Big\}.$$

For the proximity operator $\operatorname{prox}_{(f_\oslash)}(\bar{z}, \bar{w})$ we have to solve

$$\operatorname{prox}_{(f_\oslash)}(\bar{z}, \bar{w}) = \arg\min_{z,w} \frac{1}{2}(z - \bar{z})^2 + \frac{1}{2}(w - \bar{w})^2 \qquad \text{s.t. } z \leq c - \frac{1}{2a}(w - b)^2.$$

If $\bar{z}$ and $\bar{w}$ already satisfy the constraint, $\bar{z} \leq c - \frac{1}{2a}(\bar{w} - b)^2$, then $z = \bar{z}$ and $w = \bar{w}$. Otherwise the closest point $(z, w)$ satisfying the constraint will be exactly on the boundary, i.e. $z = c - \frac{1}{2a}(w - b)^2$, leaving a minimization problem solely in $w$,

$$w = \arg\min_{w} \frac{1}{2}\Big(c - \frac{1}{2a}(w - b)^2 - \bar{z}\Big)^2 + \frac{1}{2}(w - \bar{w})^2.$$

At this point we substitute $\tilde{w} = w - b$ (thus $w = \tilde{w} + b$), which will be beneficial later:

$$w = b + \arg\min_{\tilde{w}} \frac{1}{2}\Big(c - \frac{1}{2a}\tilde{w}^2 - \bar{z}\Big)^2 + \frac{1}{2}(\tilde{w} + b - \bar{w})^2.$$

First order conditions on $\tilde{w}$ are

$$\begin{aligned} 0 &\overset{!}{=} -\frac{\tilde{w}}{a}\Big(c - \frac{1}{2a}\tilde{w}^2 - \bar{z}\Big) + \tilde{w} + b - \bar{w} \\ &= \frac{1}{2a^2}\tilde{w}^3 + \Big(1 - \frac{c - \bar{z}}{a}\Big)\tilde{w} + b - \bar{w} \\ &\propto \tilde{w}^3 + 2a(a + \bar{z} - c)\tilde{w} + 2a^2(b - \bar{w}), \end{aligned}$$

which is a depressed cubic and can be solved more efficiently than general cubic equations. Further, we know that there is exactly one real root due to the fact, that the Moreau envelope has a unique minimum (making the proximity operator a function). Defining $p := 2a(a+\bar{z}-c)$ and $q := 2a^2(b-\bar{w})$, Cardano's method yields

$$\tilde{w} = \sqrt[3]{-\frac{q}{2} + \sqrt{D}} + \sqrt[3]{-\frac{q}{2} - \sqrt{D}}$$

with $D := q^2/4 + p^3/27 \geq 0$ by construction. $w$ and $z$ are obtained via $w = \tilde{w} + b$ and $z = c - \frac{1}{2a}(w-b)^2 = c - \frac{1}{2a}\tilde{w}^2$. Moreau's decomposition allows then to compute $\operatorname{prox}_{\gamma f_\oslash}$.

### 4.3 $f(y_1, y_2) = \lambda|y_1 - y_2|$

The conjugate of this function is $f^*(w_1, w_2) = \imath\{w_1 + w_2 = 0\} + \imath\{|w_1| \leq \lambda\}$. Thus,

$$(f_\oslash)^*(z, w_1, w_2) = \imath\big\{z \leq 0, w_1 + w_2 = 0, |w_1| \leq \lambda\big\}$$

and we have to determine

$$\begin{aligned}\operatorname{prox}_{\gamma(f_\oslash)^*}(\bar{z}, \bar{w}_1, \bar{w}_2) = \arg\min_{z,w_1,w_2} \frac{1}{2}\left\|\begin{matrix} z - \bar{z} \\ w_1 - \bar{w}_1 \\ w_2 - \bar{w}_2 \end{matrix}\right\|_2^2 \\ + \imath_{\mathbb{R}_0^-}(z) + \imath\{w_1 + w_2 = 0, |w_1| \leq \lambda\}.\end{aligned}$$

Since the constraints on $z$ and $w_1$, $w_2$ are independent, we can determine the proximity operators separately for $z$ and $w_1$, $w_2$, respectively. $z$ is simply given by $z = \Pi_{\mathbb{R}_0^-}(\bar{z}) = \min\{0, \bar{z}\}$. After elimination of $w_2 = -w_1$ we have to solve

$$\begin{aligned}w_1 &= \arg\min_{w_1:|w_1|\leq\lambda} \frac{1}{2}(w_1 - \bar{w}_1)^2 + \frac{1}{2}(w_1 + \bar{w}_2)^2 \\ &= \arg\min_{w_1:|w_1|\leq\lambda} \big(w_1 - (\bar{w}_1 - \bar{w}_2)\big)^2 = \Pi_{[-\lambda,\lambda]}\big((\bar{w}_1 - \bar{w}_2)/2\big).\end{aligned}$$

Using Moreau's decomposition we finally arrive at

$$\begin{aligned}\operatorname{prox}_{\gamma f_\oslash}(\bar{x}, \bar{y}_1, \bar{y}_2) &= \begin{pmatrix}\bar{x} \\ \bar{y}_1 \\ \bar{y}_2\end{pmatrix} - \gamma \begin{pmatrix}\min\{0, \bar{x}/\gamma\} \\ \Pi_{[-\lambda,\lambda]}\big((\bar{y}_1 - \bar{y}_2)/2\gamma\big) \\ \Pi_{[-\lambda,\lambda]}\big((\bar{y}_2 - \bar{y}_1)/2\gamma\big)\end{pmatrix} \\ &= \begin{pmatrix}\max\{0, \bar{x}\} \\ \bar{y}_1 - \Pi_{[-\gamma\lambda,\gamma\lambda]}\big((\bar{y}_1 - \bar{y}_2)/2\big) \\ \bar{y}_2 + \Pi_{[-\gamma\lambda,\gamma\lambda]}\big((\bar{y}_1 - \bar{y}_2)/2\big)\end{pmatrix}.\end{aligned}$$

## 5 Additional Numerical Results

### 5.1 Robust ROF Energy

Table 1 lists the final objective values for the tested methods to solve the robust ROF energy,

$$E_{\text{robust-ROF}}(u;\, f) = \frac{\lambda}{2}\sum_s \min\big\{(u_s - f_s)^2, T\big\} + \sum_{s\sim t}|u_s - u_t|. \tag{4}$$

Note that the baseline energy value (194134) is computed using 256 discrete labels and does not allow fractional grayscale values. The visual comparison of the respective minimizers is provided in Fig. 2. A quantitative analysis based on the PSNR between the baseline and the obtained results is presented in the main manuscript. Fig. 1 depicts the evolution of energies with respect to CPU time for the baseline max-flow formulation and the discrete-continuous one. The CPU is a quad-core Intel Xeon running at 2.53 GHz clock speed.

| Baseline | #segments | GC-Avg | GC-Sampled | GC-Convexified | DC-MRF | DC-Linear |
|---|---|---|---|---|---|---|
| 194134 | 32 | 202513 | 202279 | 194282 | N/A | 194994 |
| | 16 | 226210 | 225594 | 194929 | 197040 | 197041 |
| | 8 | 336446 | 332744 | 205074 | 208303 | 208660 |
| | 4 | 504475 | 494866 | 276298 | 277421 | 279722 |

Table 1: Robust ROF energies obtained for the different methods. See Fig. 2 for visual results.

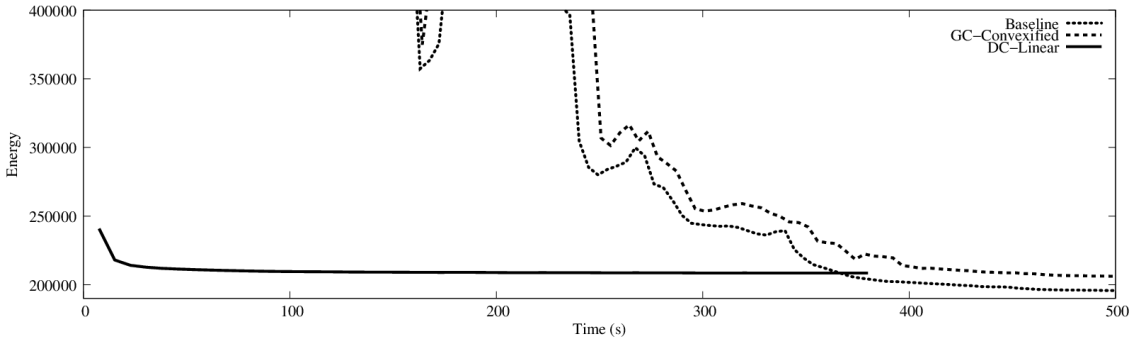


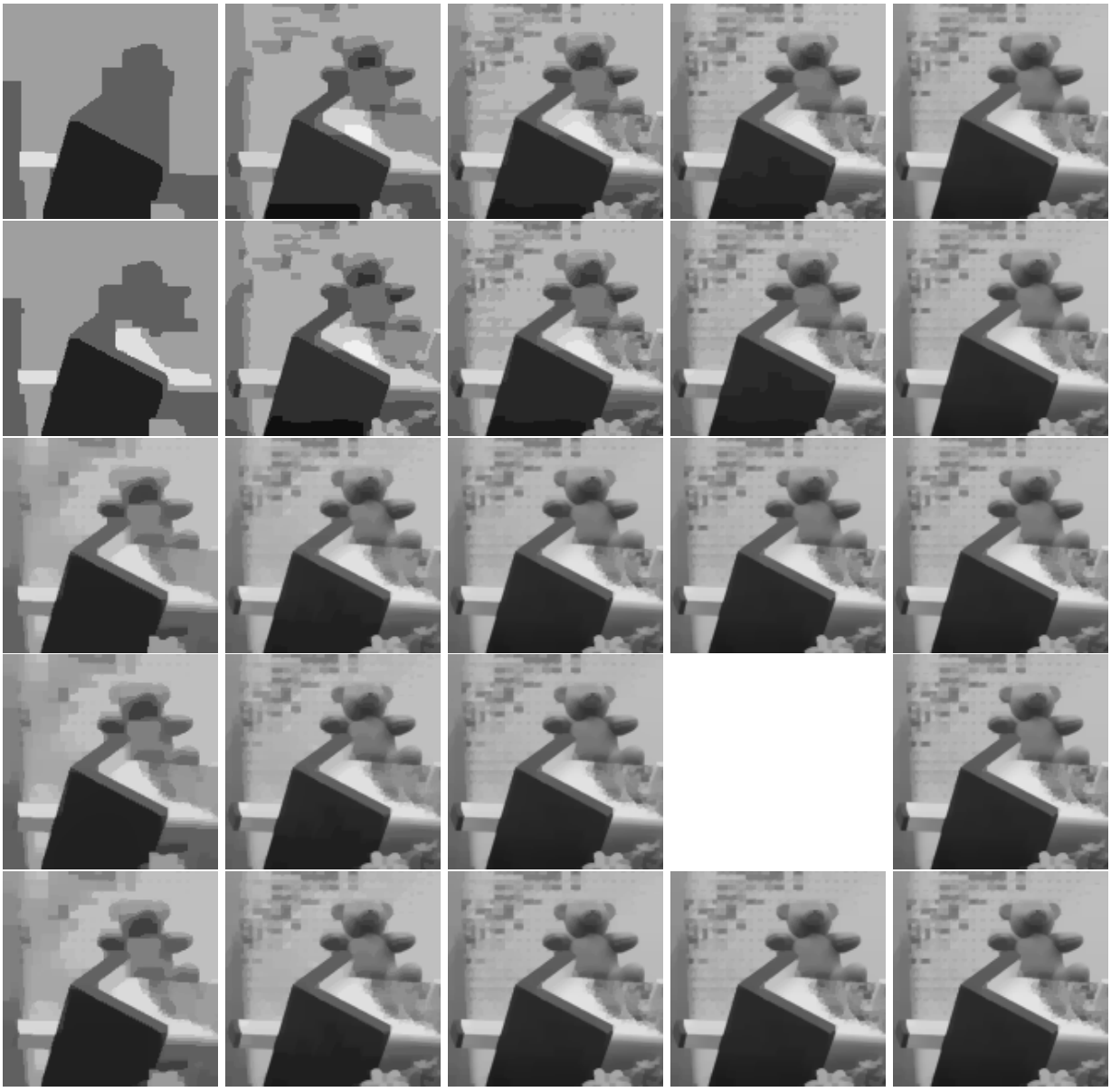

Figure 1: Evolution of energies for the baseline maximum flow method and the proposed discrete-continuous formulation with respect to CPU time (in $s$).

### 5.2 Dense Stereo Estimation

Fig. 3 is essentially the same as Fig. 5 in the main manuscript, but with parametric (i.e. piece-wise convex quadratic) unary potentials (similar to approximating the truncated quadratic potential in the robust ROF example).

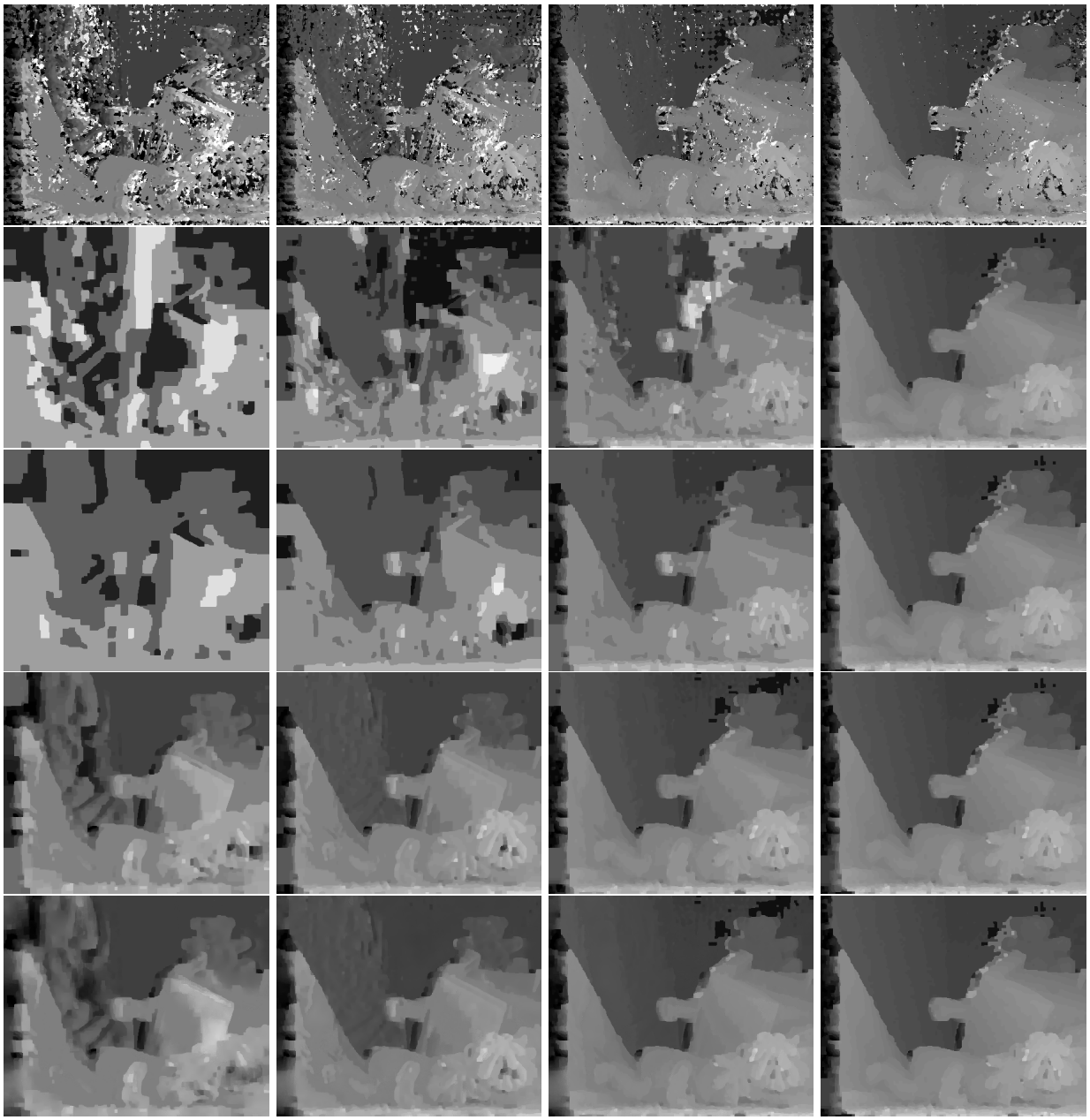
## References

[1] Combettes, P.L., Pesquet, J.C.: Proximal Splitting Methods in Signal Processing. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer (2011) 185–212



(a) 4 levels (b) 8 levels (c) 16 levels (d) 32 levels (e) Baseline

Figure 2: Visual comparison of denoising using a robust (truncated quadratic) ROF model. The columns correspond to numbers of used segment to represent the range $[0, 255]$. The rightmost column always shows the (globally optimal) result of a max-flow approach using 256 discrete labels for reference. 1st row: max-flow with averaged unary potentials. 2nd row: max-flow with sampled unary potentials. 3rd row: full 256-label maxflow, but with convexified unary potentials in each subrange. 4th row: result of optimizing $E_{\text{DC-MRF}}$. 5th row: minimizers of $E_{\text{DC-Linear}}$.



(a) 4 levels (b) 8 levels (c) 16 levels (d) Baseline

Figure 3: Visual comparison for computational stereo using an $L^1$ smoothness prior. The columns correspond to numbers of used segment to represent the range $[0, 63]$. The rightmost column always shows the (globally optimal) result of a max-flow approach using 64 discrete labels for reference. 1st row: winner-takes-all result of convexified unary potentials. 2nd row: max-flow with sampled unary potentials. 3rd row: max-flow with averaged unary potentials. 4th row: full 64-label maxflow, but with convexified unary potentials in each subrange. 5th row: minimizers of $E_{\text{DC-Linear}}$. The differences between the 4th and the 5th row illustrate how much a convexified unary potential affects the minimizer and what is lost due to the compressed discrete-continuous representation in $E_{\text{DC-Linear}}$.