Entropy Minimization for Convex Relaxation Approaches

Mohamed Souiai[1] Martin R. Oswald[2]
Youngwook Kee[3] Junmo Kim[3] Marc Pollefeys[2] Daniel Cremers[1]
[1]TU Munich, Germany [2]ETH Zürich, Switzerland [3]KAIST, South Korea

Abstract

Despite their enormous success in solving hard combinatorial problems, convex relaxation approaches often suffer from the fact that the computed solutions are far from binary and that subsequent heuristic binarization may substantially degrade the quality of computed solutions. In this paper, we propose a novel relaxation technique which incorporates the entropy of the objective variable as a measure of relaxation tightness. We show both theoretically and experimentally that augmenting the objective function with an entropy term gives rise to more binary solutions and consequently solutions with a substantially lower optimality gap. We use difference of convex function (DC) programming as an efficient and provably convergent solver for the arising convex-concave minimization problem. We evaluate this approach on three prominent non-convex computer vision challenges: multi-label inpainting, image segmentation and spatio-temporal multi-view reconstruction. These experiments show that our approach consistently yields better solutions with respect to the original integral optimization problem.

1. Introduction

Numerous problems in vision – including two-region and multi-region image segmentation, stereo- and multiview reconstruction or optical flow estimation – can be cast as variational multi-labeling problems of the form

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} E(\mathbf{u}), \qquad (1.1)$$

with a labeling function $\mathbf{u} : \Omega \to \Gamma$ from a domain Ω to a label space Γ. The domain and the label space can be both discrete or continuous. Numerous works have proposed methods for efficiently computing solutions to the above integer problem. A common approach [21] is to rephrase the above integer problem as a binary labeling problem with an indicator variable $u : \Omega \to \{0,1\}^{|\Gamma|}$:

$$u_{\text{bin}}^* = \arg\min_{u:\Omega \to \{0,1\}^{|\Gamma|}} E(u) \ , \qquad (1.2)$$

possibly subject to additional constraints. While this binary formulation often leads to a convex energy E, the binary constraint makes the optimization domain non-convex and generally yields a hard combinatorial problem.

Relaxation. The central idea of many convex relaxation techniques is to drop the integrality constraint and consider the relaxed convex problem:

$$u_{\text{rel}}^* = \arg\min_{u:\Omega \to [0,1]^{|\Gamma|}} E(u). \qquad (1.3)$$

Solving this problem is usually much easier, but its solution u_{rel}^* can be non-binary and consequently a rounding scheme needs to be applied to obtain a binary solution.

Rounding. In general, the rounded solution can be very different from the globally optimal solution u_{bin}^* and may not even be a local minimum of the binary objective (1.2). The simplest rounding scheme is to select the label with the highest likelihood. We will detail the rounding schemes later when looking at certain problem instances of (1.3).

For certain 2-label problems, a thresholding theorem [17] may assure provably optimal binary solutions upon simple thresholding of the relaxed solution u_{rel}^*. For more general multi-label problems the computed solutions are often far from binary and rounding may drastically increase the energy and the corresponding a-posteriori optimality bound (the energetic difference between rounded and relaxed solution). Moreover, a number of works incorporate additional constraints on the solution in order to constraint the volume/area [22, 27], or higher order moments [8] of the resulting segmentation, or to enforce size proportions of respective segments [15]. While these constraints are meaningful for the binary problem (1.2), they often change their physical meaning in the relaxed setting (1.3).

Novel relaxation scheme. In order to cope with the above mentioned problems we propose to augment problem (1.1) with an additional term which promotes the integrality of the solution during optimization and thereby leads to better solutions with substantially smaller optimality gaps. The key idea is to control the integrality of the objective variable by means of Shannon's entropy. Entropy minimization has been used in several computer vision applications including shadow removal [4] and image segmentation [23]

where it is used as a general color consistency criterion for separating histograms. To the best of our knowledge this work is the first to apply entropy minimization on the labeling function u. The augmented relaxed problem composes of the original labeling problem and the additional entropy term H weighted by $\theta \in \mathbb{R}_{\geq 0}$ and reads as follows:

$$u^* = \underset{u:\Omega \to [0,1]^{|\Gamma|}}{\arg\min} \; E(u) + \theta H(u). \tag{1.4}$$

This generalizes the convex relaxation problem (1.3), obtained by setting $\theta = 0$. Likewise, by taking $\theta \to \infty$ we also recover (1.2) since the entropy term becomes an integral constraint. Unfortunately problem (1.4) is not convex anymore for $\theta > 0$ and obtaining the global minimum of its relaxed version is generally not possible anymore. Luckily, problem (1.4) exhibits a special structure: it decomposes into a convex function $E(\cdot)$ and a concave function $\theta H(\cdot)$ being equivalent to a difference of convex functions (DC). This makes formulation (1.4) amenable to the DC programming approach [25] which guarantees to find a stationary point of the objective. Although there is no guarantee for finding a globally optimal solution, we show that the obtained non-rounded results are more binary than the results of state-of-the-art relaxation methods. In this paper, we focus on spatially continuous variational approaches as they are easily parallelizable and do not suffer from metrication errors in contrast to their discrete counterparts [7]. Though, our approach might also be applicable in a discrete setting.

1.1. Contributions

Our contributions can be summarized as follows:

- We propose to use the entropy of the objective variable to empirically measure the tightness of a relaxed solution. We then enhance integrality of solutions by jointly optimizing the objective function and the entropy.

- We propose to use a provably convergent algorithm for solving the arising convex-concave problem which combines DC programming and a state-of-the-art first order solver.

- We show theoretically and experimentally that our method provides solutions which are more integral and exhibit a tighter energy bound than state-of-the-art convex relaxation methods. Our approach thus promotes simple rounding schemes.

- The proposed entropy augmentation does not change the algorithm complexity. In all experiments we observed an improved convergence behavior and runtime speed-ups up to a factor of two.

- We demonstrate the effectiveness of our approach on several computer vision applications including multi-label image inpainting, image segmentation and spatio-temporal multi-view reconstruction.

2. Shannon's Entropy

Originally formulated for discrete random variables Shannon's information entropy is a measure of uncertainty. Suppose that $u : \Omega \subset \mathbb{R}^d \to [0,1]^{|\Gamma|}$ is a soft labeling function such that $\forall x \in \Omega : \sum_{\ell=1}^{|\Gamma|} u_\ell(x) = 1$. Interpreting $u(x)$ as a probability distribution on each $x \in \Omega$ we can apply information entropy on a labeling problem by directly imposing it on the indicator variable u. The total entropy of a labeling function can be written as follows:

$$H(u) = \int_\Omega -\sum_{\ell=1}^{|\Gamma|} u_\ell \log u_\ell \; dx \; , \tag{2.1}$$

which is an integral of a point-wise concave entropy measure in each $x \in \Omega$. For brevity we will refer to the total entropy as entropy for the rest of the paper.

3. Solving the Convex-Concave Program

In the field of variational convex relaxation approaches, non-convex optimization has gained tremendous popularity in recent years. However most of these works focus on realizing non-convex regularizers [18, 13]. We make use of DC programming, dating back to a seminal work by Tao *et al.* [25] which generalizes subgradient algorithms for convex maximization. The principle of minimizing the difference of convex functions heavily relies on concepts from convex optimization and especially DC duality [26]. Closely related to DC programming is the so called convex-concave procedure (CCCP) described later in [29] though it assumes differentiability of the objective function. In [6] DC programming is applied to a QP relaxation of MAP inference in order to cope with the non-convex objective.

3.1. DC Programming

DC programming deals with solving a non-convex problem of the following form:

$$\inf_u \{g(u) - h(u)\} \; , \tag{3.1}$$

where $g(u)$ and $h(u)$ are convex functions. In order to solve (3.1) we make use of a simplified form of the DC algorithm [24]. Based on DC duality and the KKT conditions for DC programs, the algorithm generates the following sequences $v^k \in \partial h(u^k)$ and $u^{k+1} \in \partial g^*(v^k)$ which guarantee to converge to a critical point. The overall DC algorithm is illustrated in algorithm 1, where g^* denotes the Legendre-Fenchel conjugate of g.

Algorithm 1 The DC Algorithm

Initialize $u^0 \in \text{dom } g$.
while not converged **do**
    Choose $v^k \in \partial h(u^k)$
    Choose $u^{k+1} \in \partial g^*(v^k)$
    $k \leftarrow k+1$
end while

3.2. Solving monotone inclusions

Note that since $u^{k+1} \in \partial g^*(v^k) \Leftrightarrow v^k \in \partial g(u^{k+1}) \Leftrightarrow 0 \in \partial g(u^{k+1}) - v^k$ and since g a convex, we solve a monotone inclusion problem in each iteration. Hence choosing $u^{k+1} \in \partial g^*(v^k)$ in algorithm 1 amounts to solving the following convex optimization problem:

$$u^{k+1} = \arg\min_u g(u) - \langle v^k, u \rangle \qquad (3.2)$$

In order to be able to solve problem (1.4) using DC programming we make the following identifications:

$$g(u) = E(u) + \delta_{U_{\text{rel}}}(u) \qquad h(u) = \theta H(u) , \qquad (3.3)$$

where $\delta_{U_{\text{rel}}}(u)$ denotes the characteristic function:

$$\delta_{U_{rel}}(u) = \begin{cases} 0 & \text{if } u(x) \in [0,1] \ \forall x \in \Omega, \\ \infty & \text{otherwise.} \end{cases} \qquad (3.4)$$

Most variational problems $g(u)$ are large scale and nonsmooth and therefore not easily solvable using standard solvers. To this end, we use the state-of-the-art primal-dual algorithm of Pock *et al.* [2] which solves a saddle point formulation of problem (3.2) (further details in the supplementary material). Using Algorithm 1 we can solve any problem of the form (1.4). For binary tomographic reconstruction, [28] proposed a similar DC programming framework for implicit rounding. In contrast to the entropy term which is well-grounded on information theory and naturally generalizes to higher dimensions, they use a negative quadratic term as a heuristic to impose integrality of the solution.

4. Experiments

We consider three problem instances which exhibit non-tight relaxations. For all experiments we initialized function u in our algorithm with a zero or random function while choosing the entropy parameter as $\theta \in [0.01, 0.5]$.

4.1. Multi-label Image Segmentation

In variational multi-label image segmentation one assumes a continuous domain Ω and a discrete label space Γ with $|\Gamma| \geq 2$. The image domain $\Omega \subset \mathbb{R}^2$ is to be segmented into $|\Gamma|$ pairwise disjoint regions Ω_ℓ which are encoded by the label indicator function $u : \Omega \to \{0,1\}^{|\Gamma|}$,

$u_\ell(x) = \mathbf{1}_{x \in \Omega_l}$. The overall problem can be stated as a minimal partition problem. To find a solution to such a problem augmented with an entropy term we solve the following optimization problem:

$$\min_u \sum_{\ell=1}^{|\Gamma|} \int_\Omega \left[\lambda u_\ell(x) \varrho_\ell(x) + \frac{1}{2}|Du_\ell| \right] dx + \theta H(u)$$

$$\text{s.t.} \sum_{\ell=1}^{|\Gamma|} u_\ell(x) = 1, \quad u(x) \geq 0 \qquad \forall x \in \Omega \qquad (4.1)$$

The data fidelity term $\varrho_l(x) : \Omega \to \mathbb{R}$ assigns a color-based pixel-wise cost to each pixel x for belonging to region ℓ. Expression Du in the smoothness term denotes the distributional derivative of u throughout the paper. It encourages regularity of the obtained partitions in the solution by minimizing its boundary length and is chosen as proposed by Zach *et al.* [30]. Although the relaxation of Chambolle, Cremers and Pock (CCP) [1] of the boundary length is tighter compared to Zach *et al.* [30], its complexity grows quadratically with the number of labels, making it impractical for large scale problems with many labels. Note that the energy functional (4.1) is a convex-concave program which can be solved by applying the DC algorithm 1. Figure 1 illustrates that despite the less tight relaxation (4.1) we obtain a drastic improvement of the solution by augmenting the optimization problem by an entropy term. We observe that for the inner optimization problem in Algorithm 1, only a few iterations (1-5) are necessary for convergence. As a result, we obtain tight solutions even with the simple relaxation of Zach *et al.* [30] without the need to drastically increase the runtime by using a tighter CCP relaxation.

Rounding. To compute a binary solution from the relaxed one, we select the most likely label point-wise, i.e. $\forall x : u_{\text{bin}}(x) = \hat{\mathbf{e}}_j$ with $j = \min\left\{ \arg\max_\ell (u_{\text{rel}}^*(x))_\ell \right\}$ and $\hat{\mathbf{e}}_j$ being the j-th unit vector in the space $\{0,1\}^{|\Gamma|}$. Note that Lellmann *et al.* [10] proposed a probabilistic rounding scheme which provides an a-priori bound on the energy. However it is slower and yields slightly degraded results in practice (see [9] for a comparison on the triple-junction problem).

4.1.1 Non-Uniqueness of the Solution

The energy (4.1) without $H(u)$ is not strictly convex and thus admits several binary solutions. If the relaxation is not tight then convex combinations of distinct binary solutions may get assigned lower energies which in turn promotes non-binary solutions. This is especially visible for larger numbers of labels. Figure 2 shows that already with 4 labels the CCP relaxation tends to produce a convex combination of binary solutions which is a valid minimizer in the case of the relaxed problem as we show in Proposition 4.1.
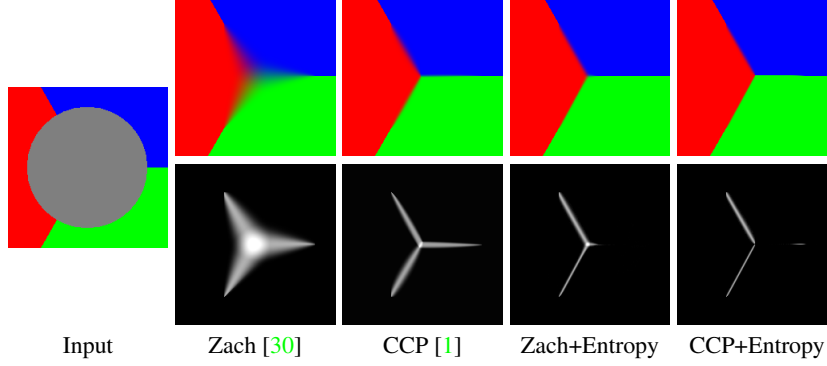
Figure 1. Inpainting using different relaxations of the Potts model with ($\theta = 0.02$) and without ($\theta = 0$) entropy. The visualized entropy shows that the joint minimization of the entropy yields more binary solutions.
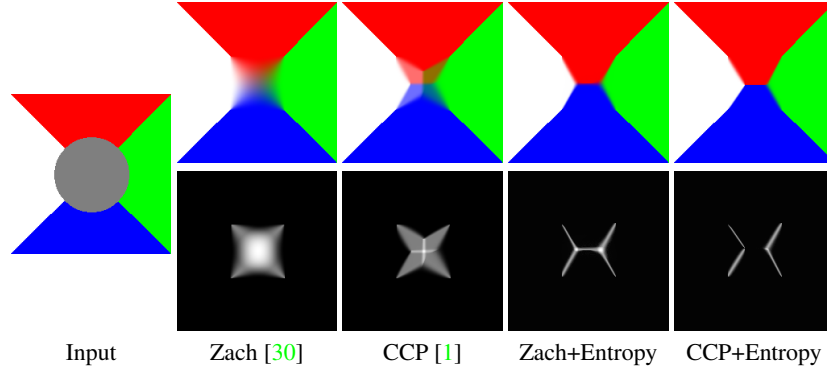

Figure 2. Inpainting in the case of 4 regions. In case of non uniqueness our approach (with $\theta = 0.02$) picks an almost integral solution with lower energy (see table 1).

Proposition 4.1 (Convex combinations of binary solutions have lower energies). *The convex combination $u_\alpha = \alpha u_1^* + (1-\alpha)u_2^*$ of two binary solutions $u_1^*, u_2^* : \Omega \to \{0,1\}^{|\Gamma|}$ of problem (1.2) has lower or equal energy (Eq. (1.3)) than the energies of the binary solutions, i.e. $E(u_\alpha) \leq E(u_1^*)$.*

Proof. By convexity of the objective $E(u)$ we have:
$E(\alpha u_1^* + (1-\alpha)u_2^*) \leq \alpha E(u_1^*) + (1-\alpha)E(u_2^*)$
and since $E(u_1^*) = E(u_2^*)$ we prove the claim. The generalization for more than two solutions is straightfoward. □

We observe that adding an entropy term tends to constrain the solution space to more integral solutions, hence our approach helps picking a solution which is more binary and hence tighter than the convex combination. In addition to the entropy, we measure the tightness of the solution by also evaluating the established posterior optimality bound [16] $G(u_{\text{bin}}, u_{\text{rel}}^*) = \frac{E(u_{\text{bin}}) - E(u_{\text{rel}}^*)}{E(u_{\text{rel}}^*)}$, where u_{rel}^* is the solution of the relaxed problem and u_{bin} the corresponding rounded solution. The exact runtimes and comparisons of different relaxations combined with entropy are presented in Table 1. Note that in addition to obtaining tighter solutions our algorithm outperforms the original formulation even in the runtime. The following Proposition 4.2 shows

that additionally minimizing the entropy promotes binary solutions over convex combinations.

Proposition 4.2 (Energy (1.4) favors binary solutions over convex combinations). *Lets denote the objective (1.4) by $E_H(u) = E(u) + \theta H(u)$ and let $u_\alpha = \sum_i \alpha_i u_i^*$ be a convex combination of different optimal binary labelings $u_i^* : \Omega \to \{0,1\}^{|\Gamma|}$ with normalized weights $\alpha_i \in [0,1]$, $\sum_i \alpha_i = 1$. Then, for sufficiently large θ, binary solutions u_i^* have a strictly lower entropy-augmented energy E_H than convex combinations of binary solutions u_α, that is, $E_H(u_\alpha) > E_H(u_i^*)$.*

Proof. Using the definition of E_H, the convexity of $E(u)$ and the property that $H(u)$ vanishes for binary u, we derive the following equalities and inequalities for any binary solution u_i^*:

$$E_H(u_\alpha) - E_H(u_i^*) = \underbrace{E(u_\alpha) - E(u_i^*)}_{\leq 0} + \theta \underbrace{\left(H(u_\alpha) - H(u_i^*)\right)}_{\geq 0}$$

(4.2)

By choosing $\theta > \frac{E(u_i^*) - E(u_\alpha)}{(H(u_\alpha) - H(u_i^*))}$ the following inequality holds:

$$E_H(u_\alpha) - E_H(u_i^*) > 0 \Leftrightarrow E_H(u_\alpha) > E_H(u_i^*) \quad (4.3)$$

□

	Zach	CCP	Zach+Entropy	CCP+Entropy
E_{rel}	630.1	634.3	639.6	**637.7**
E_{bin}	673.9	691.6	**666.8**	668.1
$G(u_{\text{bin}}, u^*_{\text{rel}})$	0.069	0.090	**0.042**	0.047
Entropy H	2769	1739	379	**306**
Runtime [s]	153	188	**68**	125

Table 1. Relaxed and binary energies for different relaxations with and without entropy term, optimality gaps as well as entropy values and runtimes in seconds for the 4-region inpainting in Fig. 2.

Table 1 shows the corresponding energies E_{bin} and E_{rel} to the results in Figure 2 for the binary and relaxed solutions for the Zach and CCP relaxation respectively. While the energies obtained using the additional entropy term are higher than the relaxed energies, the binarized energies of both relaxations combined with an entropy penalization are clearly lower than the mere convex relaxations.

4.2. Binary Image Segmentation with a Fixed Volume Constraint

By considering only two labels the relaxation of problem (1.3) becomes tight and optimal binary solutions can be computed via simple thresholding of the relaxed solution [17]. Unfortunately, this changes easily by adding further constraints to the optimization problem. We consider a fixed volume constraint on the solution of the segmentation problem. Volume constraints have been used with convex relaxation methods for image segmentation [22, 14], image-based modeling [27, 20] and they have also been generalized to higher order moment constraints [8]. In [14] they also addressed the problem of the non-tight relaxation, but their suggested algorithm is less general as user-provided seed points are required. Discrete approaches to this NP-hard problem have been suggested in [11, 3]. Both of them are restricted to equality constraints and the former only provides approximate solutions and its runtime is exponential in the number of labels. Our approach is more general than previous works, it provides the desired results for the fixed volume segmentation problem and it improves the convergence of a state-of-the-art solver at the same time. Thus, we consider the following minimization problem:

$$\min_u \int_\Omega \left[g|Du| + \lambda f u \right] dx + \theta H(u)$$

$$\text{s.t.} \int_\Omega u \, dx - V_t = 0 \;, \tag{4.4}$$

where λ steers the smoothness of the solution by changing the impact of the data fidelity term being defined by function $f : \Omega \to \mathbb{R}$, $f(x) = (c_1 - I(x))^2 - (c_2 - I(x))^2$ in which $c_1, c_2 \in \mathbb{R}$ are gray values for foreground and background, $I : \Omega \to \mathbb{R}$ is the input image and V_t denotes the pre-defined target volume. The total variation weight is defined as $g = \exp(-|\nabla f|)$. Note that one easily adapts the

approach to bound the volume with inequality constraints [8, 14], but the problems of the relaxation can be better demonstrated with an equality constraint.

Because of the volume constraint there is no free choice of thresholds to obtain a binary solution and the thresholding theorem [17], which directly relates solutions of the relaxed problem to binary one, does not apply anymore. For rounding, simple thresholding as in [3] might violate the volume constraint. The following rounding scheme from [27] addresses both the volume constraint and the binary constraint. It even guarantees fulfillment of the volume constraint, if several non-binary $u(x)$ have identical values, i.e. no threshold exists to binarize the solution without violating the volume constraint.

Proposition 4.3 (Rounding scheme for fixed volume constraint [27, Prop.2]). *The relaxed solution of* (4.4) *can be projected to the set of binary functions in such a way that the resulting binary function preserves the target volume V_t.*

Proof. It suffices to order the voxels $x_i \in V$ by decreasing values $u(x_1) \geq u(x_2) \geq \ldots \geq u(x_{|V|})$. Subsequently, one sets the value of the first V_t voxels to 1 and the value of the remaining voxels to 0. $\qquad\square$
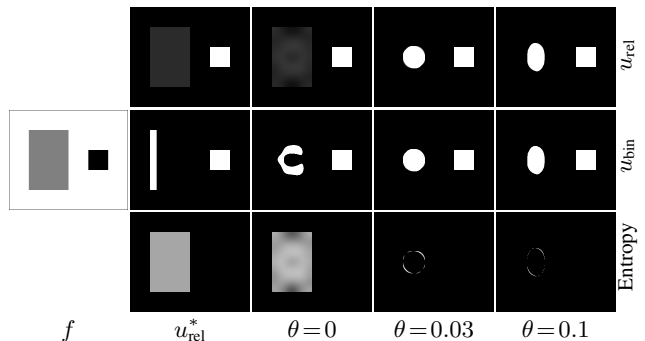

Figure 3. Segmenting a fraction of an homogeneous rectangle (gray) with the fixed volume constraint. **Left to right columns:** data term f, analytic optimal solution u^*_{rel} (top row) of the relaxed problem without entropy term. This solution is never perfectly reached due to slow convergence and oscillations. $\theta = 0$ depicts a sample solution. The corresponding rounded solution is far from the optimal solution: a disc with the area of the square next to the square. This optimal solution u^*_{bin} is found by adding the entropy term with $\theta = 0.03$. Increasing the impact of the entropy term to $\theta = 0.1$ increases the non-convexity and adds local optima.

The experiment in Fig. 3 demonstrates the relaxation problem with the volume constraint. The goal is to segment a fraction of a homogeneously colored rectangle using the volume constraint. The first column of Fig. 3 depicts the data term containing a square with strong foreground preference (black) and a rectangle with equal cost for foreground and background (gray). The white area strongly

5

prefers a background label. The area of the gray rectangle is six times larger than the square. The target volume V_t is chosen to be twice the size of the square, i.e. a sixth of the rectangle shall be filled. Due to the relaxation, the optimal solution is not necessarily compact anymore (u_{rel}^* top row). If no preference is given by the data term or by the boundary conditions, an equal distribution of the volume yields the lowest energy. For the corresponding rounded solution (middle row) we set the first V_t voxels to one according to Prop. 4.3, but since all pixels in the rectangle have equal value, their ordering is arbitrary and can lead to many different non-optimal solutions. Since the optimal relaxed solution contains the same information as the data term, optimal rounding is as hard as solving the original problem again in this case. Due to very slow convergence and oscillating behavior of the solver, the optimal relaxed solution is not reached because little volume portions are permanently shifted around and the induced pixel orderings are not related to the optimal binary solution ($\theta = 0$). An optimal binary solution u_{bin}^* - a compact disc with the area of the square - can be optained with the proposed entropy augmentation($\theta = 0.03$ middle row). The entropy term favors binary and thus compact solutions, and ensures that the relaxed solution is close to the binary one which ensures the applicability of simple rounding schemes.

Choice of the entropy weight θ. As a general rule for all experiments in the paper, we found that θ should be chosen as small as possible, but large enough to favor binary solutions. This is because larger θ increase the non-convexity of the problem and thus also the potential number of local optima, as illustrated in the last two columns of Fig. 3.
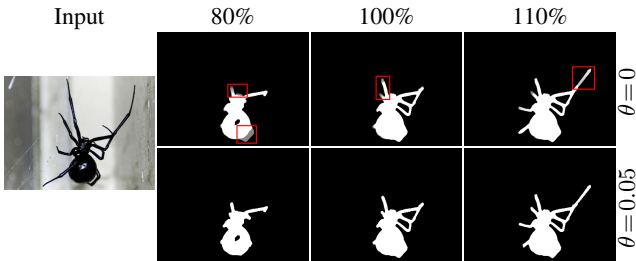

Figure 4. Changing the target volume V_t on a real image reveals that relaxation problems occur frequently. The figure compares relaxed solutions u_{rel} with and without entropy term. $V_t \widehat{=} 100\%$ corresponds to the segmentation result without volume constraint.

As shown in Fig. 4 the relaxation problems occur frequently and especially if fractions of regions with approximately homogeneous data costs need to be filled. Fig. 5 shows two more such cases for which the rounding yields solutions that heavily violate the expected minimal boundary length and how the entropy term avoids these problems.

The right plot in Fig. 6 illustrates the rounding scheme

and shows non-binary homogeneous regions as plateaus in the sorted label graph which are effectively eliminated by the proposed entropy augmentation. If the target volume seeks into a homogeneous region the rounding might result in shapes not having a minimal contour length (as shown in Fig. 5). The left plot of Fig. 6 shows that the entropy augmentation consistently yields lower binary energies and leads to smaller energy differences between relaxed and binary solutions which avoids the need for more complex rounding schemes. The relaxed solution of the original problem has always the lowest energy. Conversely, the corresponding rounded energy was always the largest in all our experiments. In all experiments the entropy augmentation stabilized the oscillating behavior of the numerical solver in the presence of non-binary homogeneous image regions and thus lead to better and faster convergence.

4.3. Spatio-temporal Multi-View Reconstruction with a Fixed Volume Constraint

The binary 2D image segmentation model from the previous section can be lifted to higher dimensions for spatio-temporal multi-view reconstruction [19]. Then, the fixed volume constraint can be applied separately to each time frame to express the prior that the overall scene volume should not change over time which is, for instance, approximately true when capturing humans with tight clothing. In this section, we demonstrate 1) that the findings of the previous section are practically even more relevant in a 3D reconstruction setting, and 2) that the proposed entropy augmentation gives consistently better results over entire sequences.

The surface Σ of the reconstructed model is represented as a binary labeling $u : V \times T \to \{0, 1\}$ of interior or exterior defined by the indicator function $u = \mathbf{1}_\Sigma$. It is observed by N cameras with known projections $\{\pi_i\}_{i=1}^N$ and approximate silhouettes $\{S_i(t)\}_{i=1}^N$. The volume-constrained and entropy-augmented reconstruction problem reads

$$\min_u \int_{V \times T} \left[\rho |D_x u| + g_t |D_t u| + \lambda f u \right] dx + \theta H(u)$$

$$\text{s.t.} \int_V u \, dx - V_t = 0 \qquad \forall t \in T \ . \qquad (4.5)$$

Note that we also change the domain from Ω to $V \times T$ in the entropy term $H(u)$ in Eq. (2.1). The regularization term in Eq. (4.5) is split into a spatial and a temporal part. The temporal term is weighted by function $g_t(x, t) = \exp\left(-|\nabla f(x, t)|\right)$ which reduces the temporal smoothing in the presence of motion. The spatial term contains the photoconsistency measure $\rho(x) : V \times T \to \mathbb{R}_{\geq 0}$ which locally attracts the surface to locations with high photometric consistency, which, in turn, is estimated by means of truncated normalized cross-correlation matching scores of image patches from neighboring cameras. Similarly to the 2D
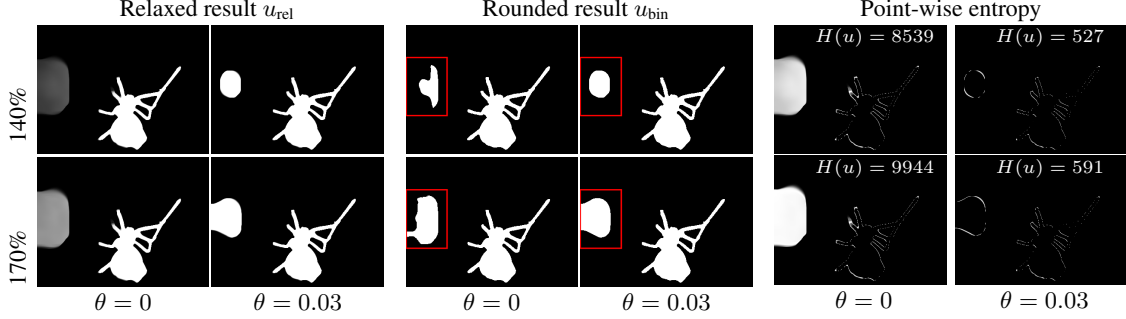
Figure 5. Effect of the entropy term on the rounded solution. This figure continues Fig. 4 with volume percentages 140% and 170% and shows the strong difference between relaxed and corresponding rounded solutions in regions with approximately homogeneous data costs. The level-sets of u_{rel} in these homogeneous regions do not necessarily obey a minimal boundary length for the enclosed volume (red boxes). The proposed entropy augmentation tackles the problem and ensures that relaxed and binary solutions are more similar.
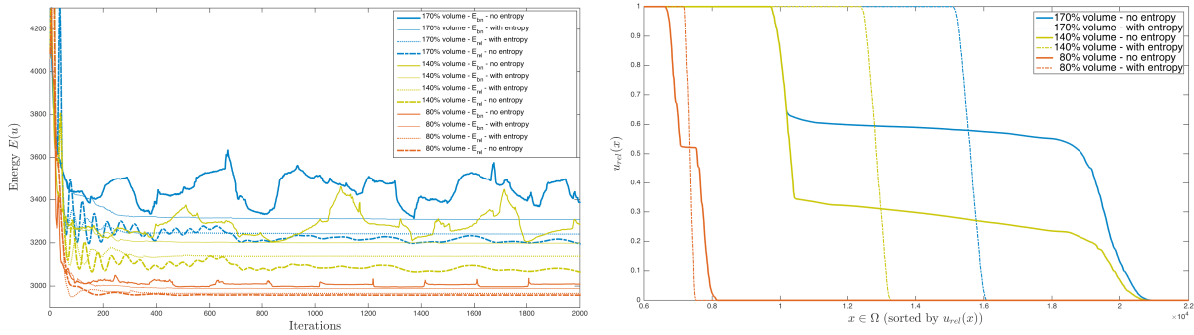

Figure 6. This figure further studies the target volumes 80%, 140% and 170% from Figs. 4, 5. These plots should be read color-wise. Each color represents a different target volume. **Left:** Energy plots during numerical optimization. Binary energies (solid lines) are always larger than relaxed ones (dashed lines). Apart from the first iterations, the energies without entropy term (thick lines) are almost always sandwiching the ones with entropy augmentation (thin lines). That is, the entropy term reduces the gap between relaxed and binary energies. **Right:** Visualization of the rounding scheme (Prop. 4.3). All pixels are ordered with respect to their relaxed label $u_{\text{rel}}(x)$. The plot shows non-binary pixels 6K to 22K (of 120K). The target volume represents a single point on the x-axis defining the transition between 0 and 1. The non-binary, almost homogeneous image regions form plateaus in this plot and lead to ambiguous selections in the rounding process.

segmentation case, the data fidelity function $f : V \times T \to \mathbb{R}$ gives local preferences for the label of u and is defined as the log-likelihood ratio of the probabilities of being either in the surface interior or exterior. For brevity and readability, we refer to [19] for the exact definitions of the data term and the regularizer weight, also because their influence on the solution is similar to the 2D case. For temporal consistency, three consecutive time frames are jointly solved and longer sequences are processed with a temporal sliding window approach. An iso-surface is extracted from the center frame using the rounding scheme in Prop. 4.3 and the Marching Cubes algorithm [12].

Fig. 7 shows a slice of a single reconstruction together with one of 16 input images, the data term f next to several solutions for different target volumes. The figure compares the impact of the entropy term and shows significant artifacts in the solutions without entropy augmentation. Besides the fact that homogenous cost regions occur frequently in the 3D setting, we observed that noisy cost regions cause the same problems as long they do not infer monotonicity

on a larger scale. Especially the Neumann boundary conditions attract the distribution of volume in the entire scene, because it minimizes the regularizer.

Figs. 8 and 9 show the evaluation of our method on the INRIA dataset [5]. Fig. 8 shows energy and volume plots over time. In many frames the volume constraint compensates for low photometric matching scores and distributes the volume according to their score. The energy plot demonstrates the robustness of the entropy augmentation as we consistently obtained lower binary energies for the entire sequence. In Fig. 9 we illustrate the benefit of the volume constraint in conjunction with the entropy term. While the energy without entropy distributes the volume to generate smooth transitions between opposing labels the entropy term concentrates the volume to the locations with the best likelihood score. In sum, the fixed-volume relaxation problem (4.4) remains tight as long as the data term or the boundary conditions enforce a monotonicity of local costs to avoid fractional homogeneous labelings. The supplementary material provides further details on the experiments.
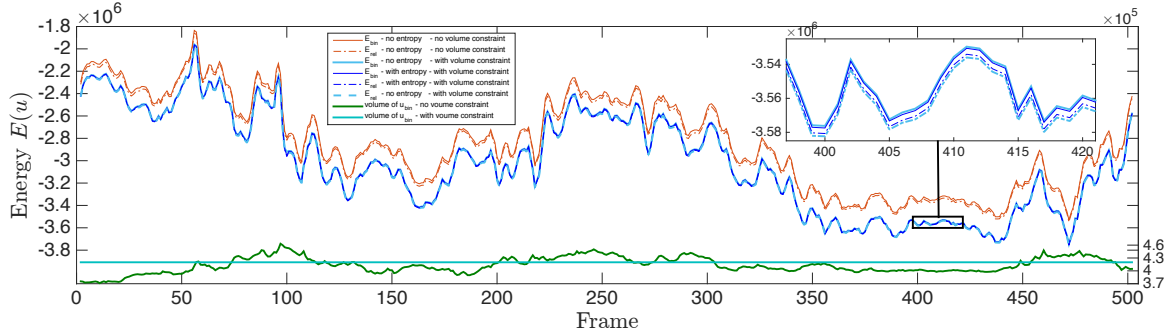
Figure 8. Plots of the volume (lower graphs) and several energies (upper graphs) in comparison for 500 frames of a multi-view video sequence (children playing sequence from [5]). Without volume constraint the volume changes over time due to insufficient matching information and occlusions which appear mostly when the volume is above our chosen target volume (cyan line). The upper graphs depict relaxed (dashed) and binary (solid) energies with (blue) and without (red) the volume constraint. Interestingly, the solution without volume constraint always has a higher energy. The blue graphs show binary and relaxed solutions with (light blue) and without (dark blue) the entropy term. For the entire sequence we verified that the solutions without entropy enclose the ones with entropy term, i.e. $E_{\text{rel}}^{\theta=0} < E_{\text{rel}}^{\theta>0} < E_{\text{bin}}^{\theta>0} < E_{\text{bin}}^{\theta=0}$ as shown in the magnified area. That is, our approach gives consistently better binary solutions.
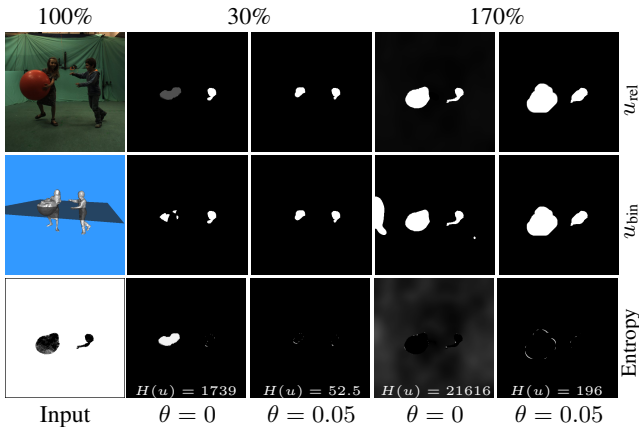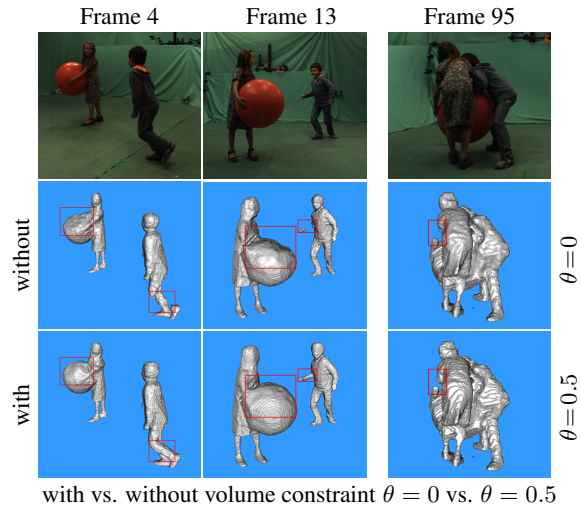

Figure 7. Changing the volume of a 3D reconstruction cross section. Homogeneous regions in the cost function occur often in a 3D reconstruction setup which in turn causes non-tight relaxations. The volume adaption can generate strong artifacts which are effectively suppressed with the proposed entropy augmentation.

with vs. without volume constraint $\theta = 0$ vs. $\theta = 0.5$

Figure 9. Reconstruction results with and without volume constraint (for $\theta = 0.5$) as well as with and without entropy term in comparison (for fixed V_t). For the first 50 frames the volume is too low (see Fig. 8 bottom) and the constraint improves the reconstruction (first two columns). Column 3 compares the impact of the entropy term. With entropy term, the regularizer concentrates the volume at locations with a higher data term rather than distributing volume around regions with low data term for attaining smoothness. Therefore, the girls arm is better recovered for the same target volume.

5. Conclusion

We proposed a relaxation technique for general multi-label problems which assures that the computed solutions of the relaxed problem are more binary and consequently have lower optimality gaps. The key idea is to combine the traditional convex relaxations with a concave entropy-term which favors binary solutions. We showed that the arising non-convex problem can be optimized with a provably convergent DC programming method. We demonstrated both theoretically and experimentally that binary solutions are energetically favored and that optimality gaps are smaller. Experiments on multi-region inpainting, image segmentation and spatio-temporal multi-view reconstruction demonstrate that the proposed entropy-based relaxation method is faster and consistently yields solutions of better visual quality and lower energy with respect to the original binary optimization problems.

6. Acknowledgement

References

[1] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences*, 5(4):1113–1158, 2012. 3, 4

[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 40(1):120–145, 2011. 3

[3] A. P. Eriksson, C. Olsson, and F. Kahl. Normalized cuts revisited: A reformulation for segmentation with linear grouping constraints. *Journal of Mathematical Imaging and Vision*, 39(1):45–61, 2011. 5

[4] G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *Proc. European Conference on Computer Vision (ECCV)*, pages 582–595. Springer, 2004. 1

[5] Institut national de recherche en informatique et en automatique (INRIA) Rhône Alpes. 4d repository. http://4drepository.inrialpes.fr/. 7, 8

[6] J. Kappes and C. Schnörr. Map-inference for highly-connected graphs with dc-programming. In *Pattern Recognition*, pages 1–10. Springer, 2008. 2

[7] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV*, Marseille, France, October 2008. 2

[8] M. Klodt, F. Steinbruecker, and D. Cremers. Moment constraints in convex optimization for segmentation and tracking. In *Advanced Topics in Computer Vision*. Springer, 2013. 1, 5

[9] J. Lellmann, F. Lenzen, and C. Schnörr. Optimality bounds for a variational relaxation of the image partitioning problem. In *Proceedings of the International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 132–146, 2011. 3

[10] J. Lellmann, F. Lenzen, and C. Schnörr. Optimality bounds for a variational relaxation of the image partitioning problem. *Journal of Mathematical Imaging and Vision*, 47(3):239–257, 2013. 3

[11] Y. Lim, K. Jung, and P. Kohli. Energy minimization under constraints on label counts. In *Proc. European Conference on Computer Vision (ECCV)*, pages 535–551, 2010. 5

[12] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21:163–169, August 1987. 7

[13] T. Möllenhoff, E. Strekalovskiy, M. Möller, and D. Cremers. Low rank priors for color image regularization. In *Proceedings of the International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2015. 2

[14] M. Niethammer and C. Zach. Segmentation with area constraints. *Medical Image Analysis*, 17(1):101–112, 2013. 5

[15] C. Nieuwenhuis, E. Strekalovskiy, and D. Cremers. Proportion priors for image sequence segmentation. In *Proc. International Conference on Computer Vision (ICCV)*, Sydney, Australia, December 2013. 1

[16] C. Nieuwenhuis, E. Toeppe, and D. Cremers. A survey and comparison of discrete and continuous multilabel segmentation approaches. *International Journal of Computer Vision*, 2013. 4

[17] M. Nikolova, S. Esedoglu, and T. F. Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648, 2006. 1, 5

[18] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock. An iterated l1 algorithm for non-smooth non-convex optimization in computer vision. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1759–1766. IEEE, 2013. 2

[19] M. R. Oswald and D. Cremers. A convex relaxation approach to space time multi-view 3D reconstruction. In *ICCV Workshop on Dynamic Shape Capture and Analysis (4DMOD)*, 2013. 6, 7

[20] M. R. Oswald, E. Töppe, and D. Cremers. Fast and globally optimal single view reconstruction of curved objects. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 534–541, June 2012. 5

[21] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *Proc. European Conference on Computer Vision (ECCV)*, pages 792–805, 2008. 1

[22] C. Reinbacher, T. Pock, C. Bauer, and H. Bischof. Variational segmentation of elongated volumetric structures. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 1, 5

[23] M. Tang, I. Ben Ayed, and Y. Boykov. Pseudo-bound optimization for binary energies. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision, ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*, pages 691–707. Springer International Publishing, 2014. 1

[24] P. D. Tao. Convex analysis approach to dc programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997. 2

[25] P. D. Tao and E. B. Souad. Algorithms for solving a class of nonconvex optimization problems. methods of subgradients. *North-Holland Mathematics Studies*, 129:249–271, 1986. 2

[26] J. Toland. Duality in nonconvex optimization. *Journal of Mathematical Analysis and Applications*, 66(2):399–415, 1978. 2

[27] E. Töppe, M. R. Oswald, D. Cremers, and C. Rother. Image-based 3D modeling via cheeger sets. In *Proc. Asian Conference on Computer Vision (ACCV)*, Nov. 2010. 1, 5

[28] S. Weber. *Discrete Tomography by Convex-Concave Regularization using Linear and Quadratic Optimization*. PhD thesis, Ruprecht-Karls-Universität Heidelberg, Germany, 2009. 3

[29] A. L. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Computation*, 15(4):915–936, 2003. 2

[30] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization*, pages 243–252, October 2008. 3, 4

Entropy Minimization for Convex Relaxation Approaches
Supplementary Material

Mohamed Souiai[1]   Martin R. Oswald[2]

Youngwook Kee[3]   Junmo Kim[3]   Marc Pollefeys[2]   Daniel Cremers[1]

[1]TU Munich, Germany   [2]ETH Zürich, Switzerland   [3]KAIST, South Korea

This supplementary document provides further details on the optimization of the problem instances discussed in the main paper. Please find more details on the experiments in the supplementary video.

1. Solving monotone inclusions via Primal-Dual Optimization

Remember that since $u^{k+1} \in \partial g^*(v^k) \Leftrightarrow v^k \in \partial g(u^{k+1}) \Leftrightarrow 0 \in \partial g(u^{k+1}) - v^k$ and since g is a convex functional, we solve a monotone inclusion problem in each iteration. Hence choosing $u^{k+1} \in \partial g^*(v^k)$ in algorithm 1 (in the main manuscript) amounts to solving the following convex optimization problem:

$$u^{k+1} = \arg\min_u g(u) - \langle v^k, u \rangle \qquad (1.1)$$

Remember, in order to be able to solve problem (1.4) in the main manuscript using DC programming we needed to make the following identifications:

$$g(u) = E(u) + \delta_{U_{\text{rel}}}(u) \qquad h(u) = H(u) \qquad (1.2)$$

Where $\delta_{U_{\text{rel}}}(u)$ denotes the indicator function:

$$\delta_{U_{rel}}(u) = \begin{cases} 0 & \text{if } u(x) \in [0,1] \;\; \forall x \in \Omega, \\ \infty & \text{otherwise.} \end{cases} \qquad (1.3)$$

Most variational problems $g(u)$ are large scale and non-smooth and are thus not easily solvable using standard solvers. To this end, we resort to state-of-the-art first-order algorithms [3, 1] which can deal with such problems. In our experiments we use the primal-dual algorithm of Pock *et al.* [3] which solves a saddle point formulation of problem (1.1) as illustrated in the supplementary materials. Trying to solve problem (1.1) directly using the e.g. the proximal point method [5] turns out to be a hard task for many optimization problems in vision since evaluating the resolvent operator of $\partial g(u^{k+1}) - v^k$ is an optimization problem itself. Therefore we use a splitting of $g(u)$ as follows:

$$g(u) = F(Ku) + G(u) \qquad (1.4)$$

where K denotes a continuous linear operator and F and G are convex functions. For solving such a splitting we use the saddle point form of (1.4):

$$g(u) = \sup_w < w, Ku > -F^*(w) + G(u) \qquad (1.5)$$

Now, algorithm 1 in the main paper can be reformulated by replacing $u^{k+1} \in \partial g^*(v^k)$ by the primal-dual procedure [3] as illustrated in algorithm 1.

Algorithm 1 The DC-Primal-Dual Algorithm

Initialize $u^0, \bar{u}^0 \in \text{dom } G$ and $w \in \text{dom } F^*$:
while not converged **do**
   Choose $v^k \in \partial h(u^k)$
   while not converged **do**
      $w^{n+1} = (I + \sigma \partial F^*)^{-1}(w^n + \sigma K \bar{u}^n)$
      $u^{n+1} = (I + \sigma \partial G)^{-1}(u^n - \tau K^* w^{n+1} + v^k)$
      $\bar{u}^{n+1} = 2u^{n+1} - u^n$
      $n \leftarrow n + 1$
   end while
   $k \leftarrow k + 1$
end while

We experienced that full convergence of the inner primal-dual iterations is not necessary for the overall convergence. Considering an existing primal-dual implementation, the proposed entropy augmentation extends the original code by just a few lines and is therefore very simple to apply.

2. Optimization for Multi-label Image Segmentation

In the case of multi-labeling we write down problem 4.1 in the main paper as the following saddle point problem:

$$\min_u \max_p \sum_{\ell=1}^{|\Gamma|} \int_\Omega u_\ell(x) \varrho_\ell(x) dx + \int_\Omega \langle \nabla u_\ell, p_\ell \rangle dx +$$
$$\delta_{\mathcal{P}}(p) + \delta_{\mathcal{U}}(u) \qquad (2.1)$$

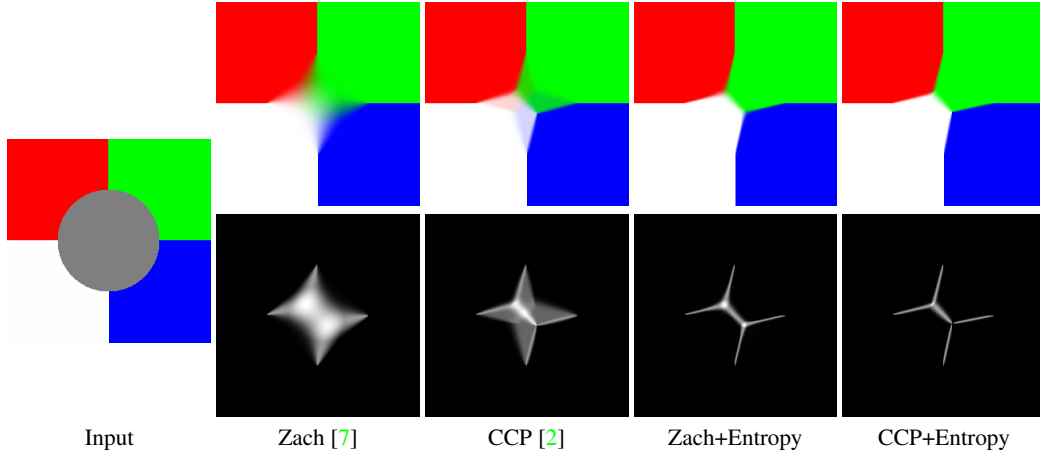| Input | Zach [7] | CCP [2] | Zach+Entropy | CCP+Entropy |

Figure 1. Inpainting in the case of 4 regions. In contrast to CCP [2] which yields a different solution on the rotated problem setting (see Fig. 2 in the main paper), our approach yields the same almost integral solutions for both the original and the rotated problem setting.

where the indicator functions $\delta_{\mathcal{P}}(p)$ and $\delta_{\mathcal{U}}(u)$ force the dual variable p and primal variable u to stay in the following constraint sets:

$$\mathcal{P}(p) = \left\{ p : \Omega \to (\mathbb{R}^2)^\Gamma \,\big|\, |p_\ell(x)| \le 1 \,\forall\, x \in \Omega \right\}$$

$$\mathcal{U}(u) = \left\{ u : \Omega \to [0,1]^\Gamma \,\Big|\, \sum_{\ell=1}^{n} u_\ell(x) = 1 \quad \forall\, x \in \Omega \right\}. \tag{2.2}$$

Note that the choice of the set P determines the type of the chosen relaxation [2]. Hence, the difference of the CCP and Zach relaxation lies in the choice of P. In order to solve problem (2.1) using algorithm 1, we make the identifications $G(u) = \sum_{\ell=1}^{|\Gamma|} \int_\Omega u_\ell(x)\varrho_\ell(x)dx + \delta_{\mathcal{U}}(u)$, $K = \nabla$ and $F^* = \delta_{\mathcal{P}}(p)$. Note that since $F^*(p)$ is an indicator function, its proximal operator $(I + \sigma \partial F^*)^{-1}$ is just the orthogonal projection onto the simplex constraint $\mathcal{U}$ which, in turn, can be performed using the algorithm in [4]. For more details on how to compute proximal operators see e.g. [3]. Since G is linear in the data term and includes an indicator function it resembles just a gradient descent in the term $\sum_{\ell}^{\Gamma} \int_\Omega u_\ell(x)\varrho_\ell(x)dx$ and a subsequent orthogonal projection onto the set $\mathcal{U}$.

Figure 1 provides a further experiment for the rotated case of the 4-junction problem in the main paper. Interestingly, the original CCP relaxation [2] yields different solutions for these two similar, but rotated cases while our approach yields consistently the globally optimal solution for both cases and random initialization.

3. Optimization for Binary Image Segmentation

Similar to the multi-label case we reformulate equation (5.1) in the main manuscript as the following saddle-point problem:

$$\min_u \max_{p,\xi} \; \lambda \int_\Omega u f \, dx + \int_\Omega \langle \nabla u, p \rangle dx + \xi \left(\int_\Omega u \, dx - V_t \right) + \delta_P(p) + \delta_U(u) \tag{3.1}$$

where the indicator functions $\delta_{\mathcal{P}}(p)$ and $\delta_{\mathcal{U}}(u)$ force the dual variable p and primal variable u to stay in the following constraint sets:

$$P(p) = \left\{ p : \Omega \to (\mathbb{R}^2) \,\big|\, |p(x)| \le 1 \,\forall\, x \in \Omega \right\}$$

$$U(u) = \left\{ u : \Omega \to \mathbb{R} \,\big|\, u(x) \in [0,1] \,\forall\, x \in \Omega \right\}. \tag{3.2}$$

Note that we additionally handle the volume constraint $\int_\Omega u \, dx = V_t$ by adding a Lagrange multiplier $\xi \in \mathbb{R}$ to the optimization problem.

Note that the case of spatio-temporal multi-view reconstruction is similar to the 2-label case except for the projection onto the dual space because of the higher dimensional regularizer (6.1, in the main paper). For more details on the optimization of spatio-temporal multi-view reconstruction without the entropy term we refer to [6].

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011. 1

[2] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences*, 5(4):1113–1158, 2012. 2

[3] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 40(1):120–145, 2011. 1, 2

[4] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of r^n. *Journal of Optimization Theory and Applications*, 50(1), 1986. 2

[5] G. J. Minty et al. Monotone (nonlinear) operators in hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962. 1

[6] M. R. Oswald and D. Cremers. A convex relaxation approach to space time multi-view 3D reconstruction. In *ICCV Workshop on Dynamic Shape Capture and Analysis (4DMOD)*, 2013. 2

[7] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization*, pages 243–252, October 2008. 2