

Continuous Energy Minimization Via Repeated Binary Fusion*

Werner Trobin¹, Thomas Pock^{1,2}, Daniel Cremers², and Horst Bischof¹

¹ Institute for Computer Graphics and Vision, Graz University of Technology

² Department of Computer Science, University of Bonn

Abstract. Variational problems, which are commonly used to solve low-level vision tasks, are typically minimized via a local, iterative optimization strategy, e.g. gradient descent. Since every iteration is restricted to a small, local improvement, the overall convergence can be slow and the algorithm may get stuck in an undesirable local minimum. In this paper, we propose to approximate the minimization by solving a series of binary subproblems to facilitate large optimization moves. The proposed method can be interpreted as an extension of discrete graph-cut based methods such as α -expansion or LogCut to a spatially continuous setting. In order to demonstrate the viability of the approach, we evaluated the novel optimization strategy in the context of optical flow estimation, yielding excellent results on the Middlebury optical flow datasets.

1 Introduction

Several fundamental problems in computer vision can be classified as inverse, ill-posed problems, where a direct solution is not possible (e.g. deblurring, stereo, optical flow). In such cases, a prior model of the forward process can help to infer physically meaningful solutions via a maximum a posteriori (MAP) estimation. Such MAP formulations naturally lead to energy minimization problems [1], where an energy term E_{prior} , representing the prior model, penalizes unlikely solutions and a data consistency term E_{data} enforces a close fit to the observed data:

$$\min_u \{E_{prior}(u) + \lambda E_{data}(u)\} . \quad (1)$$

Since we are dealing with spatially (and radiometrically) discrete images, at some point any optimization approach for (1) has to take the spatial discretization into account – there are two predominant strategies to do that. One currently very popular approach is to state the problem as a discrete, combinatorial optimization problem on a Markov Random Field (MRF). Since MRFs are a powerful tool for solving most low level vision tasks, a considerable research effort has been dedicated to exploring minimization methods for MRF energies (cf. [2] for a comparison of state-of-the-art algorithms). Generally, the optimization approaches

* This work was supported by the Austrian Science Fund under grant P18110-B15, the Austrian Research Promotion Agency within the VM-GPU project (no. 813396), and the Hausdorff Center for Mathematics.

are either based on message passing (e.g. loopy belief propagation by Pearl [3] and sequential tree-reweighted message passing by Kolmogorov [4]) or on graph cuts (α - β -swap and α -expansion, introduced by Boykov et al. [5], and the more recent “range moves” by Veksler [6] and LogCut by Lempitsky et al. [7]). Recently, Komodakis et al. proposed a fast optimization approach, based on the duality theory of Linear Programming [8]. The second optimization strategy for (1) uses the tools of the calculus of variations in a continuous context. Once an optimality condition for the energy is derived, the differential operators are discretized and a numerical scheme is used to minimize the energy. In contrast to the aforementioned discrete MRF approach, in the variational approach the discretization is postponed as long as possible.

This work will focus on limitations of the local optimization approaches used in the variational context. In order to circumvent these limitations, we will introduce a novel optimization strategy inspired by the discrete α -expansion and LogCut algorithms. In contrast to local methods, such an optimization strategy allows large moves and therefore is less likely to get stuck in bad local minima. Unlike combinatorial optimization approaches, the solution space does not have to be discretized and the algorithm does not induce systematic metrication errors. The proposed variational technique also facilitates high-performance implementations on massively parallel GPUs and permits an extension to higher-order “interactions” at low costs in time and memory.

After proposing the novel optimization strategy in Section 2, we evaluate this technique for optical flow estimation in Section 3. Experiments in Section 4 illustrate state-of-the-art results.

2 Energy Minimization by Solving Binary Subproblems

In order to simplify the discussion, we will only consider continuous energy functionals of the form

$$\min_{\mathbf{u}} \left\{ \int_{\Omega} \Psi(D\mathbf{u}(\mathbf{x}), D^2\mathbf{u}(\mathbf{x}), \dots) d\mathbf{x} + \lambda \int_{\Omega} \rho(\mathbf{u}(\mathbf{x}), \mathbf{x}) d\mathbf{x} \right\}, \quad (2)$$

where $\Omega \subset \mathbb{R}^n$ and $\mathbf{u} : \Omega \rightarrow \mathbb{R}^m$. The lefthand term, which corresponds to E_{prior} in (1), penalizes non-smooth solutions via a potentially robust error function Ψ , while the righthand term enforces data consistency, and the free parameter λ allows balancing the influence of the two terms. Solving such equations in a continuous setting entails deriving the Euler-Lagrange equations and finding an iterative numerical scheme to calculate the solution. These schemes are typically based on gradient descent and therefore every iteration will only result in a local improvement of the current solution. Hence, these methods are prone to get stuck in bad local minima. Moreover, the speed of convergence might be quite slow. Combinatorial optimization approaches for MRFs, on the other hand, are not restricted to small, local moves. For a lot of MRF energies, α -expansion seems to offer a good trade-off between runtime and optimality properties. The central idea of the α -expansion algorithm is to approximately minimize an (NP-hard) multi-label problem, by repeatedly solving a series of binary subproblems.

For each subproblem, the global optimum can efficiently be computed using a max-flow/min-cut algorithm [5].

Inspired by [9], where Chambolle shows close links between the Total Variation (TV) model and binary MRFs in the context of image denoising, and [10], where Nikolova et al. show how to find globally optimal solutions for certain nonconvex optimization problems by restating them as convex problems, we will pose the minimization problem (2) as a sequence of binary subproblems. Each of these subproblems can conceptually be understood as a continuous version of an α -expansion move, i.e. the current solution is changed to a proposed alternative solution, wherever this is energetically favorable. Repeatedly solving this binary problem for varying proposed solutions, i.e. performing *cycles*, as it is called in the α -expansion terminology, yields increasingly accurate results.

In order to formulate (2) as a binary problem, let $\alpha, \beta : \Omega \rightarrow \mathbb{R}^m$ be two arbitrary but *fixed* candidate solutions we will subsequently refer to as “proposed solutions” or just as “proposals.” Moreover, let $\phi : \Omega \rightarrow \{0, 1\}$ be a binary function to selectively combine α and β to a new solution

$$\mathbf{u}(\mathbf{x}) = (1 - \phi(\mathbf{x})) \alpha(\mathbf{x}) + \phi(\mathbf{x}) \beta(\mathbf{x}) . \quad (3)$$

The function ϕ is free to vary across Ω , as long as the fused solution \mathbf{u} fulfills the regularity requirements posed by the considered energy. Plugging the combined solution (3) into the model (2) yields

$$\min_{\phi \in F} \left\{ \int_{\Omega} \Psi(D\mathbf{u}(\phi(\mathbf{x}), \mathbf{x}), D^2\mathbf{u}(\phi(\mathbf{x}), \mathbf{x}), \dots) d\mathbf{x} + \right. \quad (4)$$

$$\left. \lambda \int_{\Omega} (1 - \phi(\mathbf{x})) \rho(\alpha(\mathbf{x}), \mathbf{x}) + \phi(\mathbf{x}) \rho(\beta(\mathbf{x}), \mathbf{x}) d\mathbf{x} \right\}$$

with $F = \{\phi : \Omega \rightarrow \{0, 1\}\}$, where we used the observation that

$$\rho[(1 - \phi(\mathbf{x})) \alpha(\mathbf{x}) + \phi(\mathbf{x}) \beta(\mathbf{x}), \mathbf{x}] = (1 - \phi(\mathbf{x})) \rho[\alpha(\mathbf{x}), \mathbf{x}] + \phi(\mathbf{x}) \rho[\beta(\mathbf{x}), \mathbf{x}] . \quad (5)$$

Therefore, instead of directly estimating the solution \mathbf{u} , the goal is now to find the best possible *fusion* of the two fixed proposals α and β , by optimizing with respect to ϕ . This construction is strongly related to the generalized α -expansion presented in [7], which comprises the core of the LogCut algorithm.

Unfortunately, the binary problem is not convex and hence hard to solve in a continuous context. Therefore, similar to [10], we first expand the range of ϕ to the interval $[0, 1]$ and then solve this relaxed problem in a continuous setting. Despite the fact that such a relaxation clearly invalidates (5), the observed continuous solutions are close to binary (see Fig. 3). This indicates that in the vast majority of the domain there is a clear preference for either α or β , so we argue that thresholding the continuous solution ϕ leads to a reasonable binary fusion of α and β . To avoid a solution procedure that merely oscillates between different proposals, we only accept a fused result, if the energy (4) decreases, compared to using either α or β alone.

3 Application to Optical Flow Estimation

Estimating the optical flow between two consecutive images of a scene requires finding corresponding points in the images. As mentioned above, this is an ill-posed, inverse problem, so solely relying on the optical flow constraint, i.e. to assume that the intensities remained constant, does not provide sufficient information to infer meaningful flow fields. Horn and Schunck were the first to apply the tools of the calculus of variations to impose a quadratic regularization on the displacement field [11]. Black and Anandan [12] introduced methods from robust statistics to avoid the quadratic penalization of changes in the flow field and of violations of the optical flow constraint, thereby allowing for flow discontinuities and outliers in the data term, respectively. For a systematic survey on variational formulations of the optical flow problem, we refer to Aubert et al. [13].

While robust error functions have the desirable property to limit the influence of outliers, most of them are not convex and therefore hard to optimize. Due to the fact that gross outliers are rare in optical flow estimation, several authors proposed a Total Variation (TV) regularization of the flow field and an L^1 data fidelity term [14,15,16]:

$$\min_{\mathbf{u}} \left\{ \sum_{d=1}^2 \int_{\Omega} |\nabla u_d| \, d\mathbf{x} + \lambda \int_{\Omega} \rho(\mathbf{u}(\mathbf{x}), \mathbf{x}) \, d\mathbf{x} \right\} \quad (6)$$

with $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))^T$ and $\rho(\mathbf{u}(\mathbf{x}), \mathbf{x}) = |I_1(\mathbf{x} + \mathbf{u}(\mathbf{x})) - I_0(\mathbf{x})|$, where I_0 and I_1 are the two input images. Since the data fidelity term ρ is nonlinear in \mathbf{u} , a local linearization is required at some point in the solution procedure, which limits the approach to recovering small displacements. To circumvent this restriction, the estimation procedure is typically performed on a scale pyramid of the input images, cf. [14] for an in-depth discussion. Yet, such a scale pyramid strategy often fails to recover the flow for small holes in foreground objects (which allow to observe a distant background), or for fine structures in front of a background which is moving due to ego motion. This effect can be seen around the moving leaves in the “Schefflera” sequence of the Middlebury optical flow evaluation dataset [17]. Figure 1(a) shows the input image 0 of the “Schefflera” sequence; the corresponding ground truth flow can be seen as color-coded image¹ in Fig. 1(b). Figure 1(c) shows a color-coded flow field, estimated using a scale pyramid-based implementation of the TV- L^1 optical flow model (6).

Before applying the new solution strategy, in a short technical detour, we follow the approach of Aujol et al. [18] and introduce an auxiliary displacement field \mathbf{v} , yielding a strictly convex approximation of (6):

$$\min_{\mathbf{u}, \mathbf{v}} \left\{ \sum_{d=1}^2 \int_{\Omega} |\nabla u_d| \, d\mathbf{x} + \frac{1}{2\theta} \sum_{d=1}^2 \int_{\Omega} (u_d - v_d)^2 \, d\mathbf{x} + \lambda \int_{\Omega} \rho(\mathbf{v}, \mathbf{x}) \, d\mathbf{x} \right\}, \quad (7)$$

¹ The hue encodes the direction of the flow vector, while the saturation encodes its magnitude. Regions of unknown flow (e.g. due to occlusions) are colored black.



Fig. 1. (a) input image 0 of the Middlebury “Schefflera” sequence; (b) color-coded ground truth flow (hue = direction, intensity = magnitude, black = unknown); color-coded flows, estimated using: (c) a continuous TV- L^1 flow model; (d) the proposed optimization strategy (AAE = 2.91°)

where the small, positive constant θ ensures a tight coupling of \mathbf{u} and \mathbf{v} . In contrast to the original energy (6), this convex approximation can be solved using a dual approach. Our proposed optimization strategy can now easily be applied by replacing \mathbf{v} with a binary fusion of two flow proposals $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\min_{\mathbf{u}, \phi} \left\{ \sum_{d=1}^2 \int_{\Omega} |\nabla u_d| \, d\mathbf{x} + \frac{1}{2\theta} \sum_{d=1}^2 \int_{\Omega} (u_d - [(1 - \phi) \alpha_d + \phi \beta_d])^2 \, d\mathbf{x} + \lambda \int_{\Omega} (1 - \phi) \rho(\boldsymbol{\alpha}, \mathbf{x}) + \phi \rho(\boldsymbol{\beta}, \mathbf{x}) \, d\mathbf{x} \right\}. \quad (8)$$

The relaxed version of (8) is a minimization problem in two variables, \mathbf{u} and ϕ . We therefore have to perform an alternating minimization procedure:

1. For ϕ fixed, solve for every u_d :

$$\min_{u_d} \left\{ \int_{\Omega} |\nabla u_d| \, d\mathbf{x} + \frac{1}{2\theta} \int_{\Omega} (u_d - [(1 - \phi) \alpha_d + \phi \beta_d])^2 \, d\mathbf{x} \right\} \quad (9)$$

2. For \mathbf{u} fixed, solve for ϕ :

$$\min_{\phi} \left\{ \frac{1}{2\theta} \sum_{d=1}^2 \int_{\Omega} (u_d - [(1 - \phi) \alpha_d + \phi \beta_d])^2 \, d\mathbf{x} + \lambda \int_{\Omega} (1 - \phi) \rho(\boldsymbol{\alpha}, \mathbf{x}) + \phi \rho(\boldsymbol{\beta}, \mathbf{x}) \, d\mathbf{x} \right\} \quad (10)$$

The subproblem (9) is the well understood image denoising model of Rudin, Osher, and Fatemi [19]. For this model, Chambolle proposed an efficient and globally convergent numerical scheme, based on a dual formulation [20]. In practice, a gradient descent/reprojection variant of this scheme performs better [9], although there is no proof for convergence. In order to make this paper self-contained, we reproduce the relevant results from [20,9]:

Proposition 1. *The solution of (9) is given by*

$$u_d = [(1 - \phi) \alpha_d + \phi \beta_d] - \theta \nabla \cdot \mathbf{p}_d. \quad (11)$$

The dual variable \mathbf{p}_d is obtained as the steady state of

$$\begin{aligned}\tilde{\mathbf{p}}_d^{k+1} &= \mathbf{p}_d^k + \frac{\tau}{\theta} \nabla (\theta \nabla \cdot \mathbf{p}_d^k - [(1 - \phi) \alpha_d + \phi \beta_d]) \quad , \\ \mathbf{p}_d^{k+1} &= \frac{\tilde{\mathbf{p}}_d^{k+1}}{\max \left\{ 1, |\tilde{\mathbf{p}}_d^{k+1}| \right\}} \quad ,\end{aligned}\tag{12}$$

where k is the iteration number, $\mathbf{p}_d^0 = \mathbf{0}$, and $\tau \leq 1/4$.

The subproblem (10) permits a direct solution.

Proposition 2. *The solution of (10) is given by clamping*

$$\tilde{\phi} = \begin{cases} \frac{(\mathbf{u} - \boldsymbol{\alpha})^T (\boldsymbol{\beta} - \boldsymbol{\alpha}) + \lambda \theta (\rho(\boldsymbol{\beta}, \mathbf{x}) - \rho(\boldsymbol{\alpha}, \mathbf{x}))}{(\boldsymbol{\beta} - \boldsymbol{\alpha})^T (\boldsymbol{\alpha} - \boldsymbol{\beta})} & \text{where } \boldsymbol{\alpha} \neq \boldsymbol{\beta} \\ 0 & \text{elsewhere} \end{cases}\tag{13}$$

to the range $[0, 1]$:

$$\phi = \max \left\{ 0, \min \left\{ 1, \tilde{\phi} \right\} \right\} \quad .\tag{14}$$

Proof: Starting with the Euler-Lagrange equation of (10)

$$\frac{1}{\theta} (\mathbf{u} - [(1 - \phi) \boldsymbol{\alpha} + \phi \boldsymbol{\beta}])^T (\boldsymbol{\alpha} - \boldsymbol{\beta}) + \lambda [\rho(\boldsymbol{\beta}, \mathbf{x}) - \rho(\boldsymbol{\alpha}, \mathbf{x})] \quad ,\tag{15}$$

we try to solve for ϕ , yielding

$$\phi (\boldsymbol{\beta} - \boldsymbol{\alpha})^T (\boldsymbol{\alpha} - \boldsymbol{\beta}) = (\mathbf{u} - \boldsymbol{\alpha})^T (\boldsymbol{\beta} - \boldsymbol{\alpha}) + \lambda \theta (\rho(\boldsymbol{\beta}, \mathbf{x}) - \rho(\boldsymbol{\alpha}, \mathbf{x})) \quad .\tag{16}$$

Wherever $\boldsymbol{\alpha} = \boldsymbol{\beta}$, $\rho(\boldsymbol{\alpha}, \mathbf{x}) = \rho(\boldsymbol{\beta}, \mathbf{x})$, hence ϕ can arbitrarily be chosen in $[0, 1]$. Everywhere else, we can divide by $(\boldsymbol{\beta} - \boldsymbol{\alpha})^T (\boldsymbol{\alpha} - \boldsymbol{\beta})$, yielding (13). \square

Please note that the data residuals $\rho(\boldsymbol{\alpha}, \mathbf{x})$ and $\rho(\boldsymbol{\beta}, \mathbf{x})$ are just constants. They have to be calculated only once per fusion step, and the sole requirement is that their range is \mathbb{R}_0^+ , i.e. almost any cost function can be used. Once the relaxed version of problem (8) is solved, a final thresholding of ϕ is required to obtain the binary fusion of the two flow proposals $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Since the continuous solution of ϕ is already close to binary, the threshold μ is not critical. Our heuristic solution is to evaluate the energy of the original TV- L^1 energy (8) for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and a few different thresholds $\mu \in (0, 1)$. Finally, we select the threshold yielding the flow field with the lowest energy.

3.1 Extension to a Second-Order Prior

The presented approach is by no means limited to Total Variation regularization. As an example, we will apply the proposed technique to an optical flow model with a prior based on decorrelated second-order derivatives [21]. This second-order prior has the intrinsic property to penalize only deviations from piecewise

affinity. Since spatial second-order derivatives are not orthogonal and the local information of orientation and shape are entangled, a decorrelation is necessary. In [22], Danielsson et al. used circular harmonic functions to map the the second-order derivative operators into an orthogonal space. In two spatial dimensions, the decorrelated operator is given by

$$\diamond = \sqrt{\frac{1}{3}} \left(\frac{\partial^2}{\partial_x^2} + \frac{\partial^2}{\partial_y^2}, \sqrt{2} \left(\frac{\partial^2}{\partial_x^2} - \frac{\partial^2}{\partial_y^2} \right), \sqrt{8} \frac{\partial^2}{\partial_x \partial_y} \right)^T. \quad (17)$$

The magnitude of this operator, defined as the Euclidean vector norm

$$\|\diamond u\| = \sqrt{\frac{1}{3} \left[\left(\frac{\partial^2 u}{\partial_x^2} + \frac{\partial^2 u}{\partial_y^2} \right)^2 + 2 \left(\frac{\partial^2 u}{\partial_x^2} - \frac{\partial^2 u}{\partial_y^2} \right)^2 + 8 \left(\frac{\partial^2 u}{\partial_x \partial_y} \right)^2 \right]}, \quad (18)$$

measures the local deviation of a function u from being affine. Adapting the TV- L^1 flow model (6) to the new prior is a matter of replacing the TV regularization in (6–8) with the Euclidean norm of the new operator (18). Instead of minimizing the ROF energy (9), step 1 of the alternate optimization of \mathbf{u} and ϕ now amounts to solving

$$\min_{u_d} \left\{ \int_{\Omega} \|\diamond u_d\| \, d\mathbf{x} + \frac{1}{2\theta} \int_{\Omega} (u_d - [(1-\phi)\alpha_d + \phi\beta_d])^2 \, d\mathbf{x} \right\} \quad (19)$$

for every u_d , while keeping ϕ fixed.

Proposition 3. *The solution of (19) is given by*

$$u_d = [(1-\phi)\alpha_d + \phi\beta_d] - \theta \diamond \cdot \mathbf{p}_d. \quad (20)$$

The dual variable \mathbf{p}_d is obtained as the steady state of

$$\begin{aligned} \tilde{\mathbf{p}}_d^{k+1} &= \mathbf{p}_d^k + \frac{\tau}{\theta} [\diamond ([(1-\phi)\alpha_d + \phi\beta_d] - \theta \diamond \cdot \mathbf{p}_d^k)], \\ \mathbf{p}_d^{k+1} &= \frac{\tilde{\mathbf{p}}_d^{k+1}}{\max \{1, |\tilde{\mathbf{p}}_d^{k+1}|\}}, \end{aligned} \quad (21)$$

where k is the iteration number, $\mathbf{p}_d^0 = \mathbf{0}$, and $\tau \leq 3/112$. For a proof and further details please refer to [21].

Moreover, we employ the following standard finite differences approximation of the \diamond operator:

$$(\diamond u)_{i,j} = \begin{pmatrix} \sqrt{\frac{1}{3}} (u_{i,j-1} + u_{i,j+1} + u_{i-1,j} + u_{i+1,j} - 4u_{i,j}) \\ \sqrt{\frac{2}{3}} (u_{i-1,j} + u_{i+1,j} - u_{i,j-1} - u_{i,j+1}) \\ \sqrt{\frac{8}{3}} (u_{i,j} + u_{i+1,j+1} - u_{i,j+1} - u_{i+1,j}) \end{pmatrix}, \quad (22)$$

where (i, j) denote the indices of the discrete image domain, enforcing Dirichlet boundary conditions on $\partial\Omega$. For details on the discretization of $\diamond \cdot \mathbf{p}$, please consult [21].

In the continuous setting, such an extension requires minor adaptations of the solution procedure and incurs only a small increase of time and memory requirements. Most combinatorial optimization approaches, however, are limited to unary and pairwise clique potentials, hence such a second-order prior can not be used. Extending combinatorial optimization algorithms to higher-order cliques (e.g. as proposed in [5,23]) is either expensive in time and space or imposes restrictions on the potentials, e.g. [23] restricts the potentials to the Potts model.

4 Experiments

In this section, we first restrict the optical flow model to a single dimension (rectified stereo) in order to analyze its behavior in a simplified setting. In Section 4.2 we will use image sets from the Middlebury optical flow database [17] to illustrate that the proposed algorithm yields state-of-the-art flow estimates.

Most of the algorithm has been implemented in C++, with the exception of the numerical schemes of the solvers, which have been implemented using CUDA 1.0. All subsequent experiments have been performed on an Intel Core 2 Quad CPU at 2.66 GHz (the host code is single-threaded, so only one core was used) with an NVidia GeForce 8800 GTX graphics card, running a 32 bit Linux operating system and recent NVidia display drivers. Unless noted otherwise, in all our experiments the parameters were set to $\lambda = 50$ and $\theta = 0.1$.

4.1 Illustrative Stereo Experiment

A restriction of the optical flow model (8) to a single displacement u permits a direct comparison of the estimated solutions to the global optimum (cf. [24] for details on calculating the global optimum of this multi-label problem). Since the approach presented in [24] is based on a discrete solution space, we further restrict our method by using only constant disparity proposals in 0.5 pixel increments.

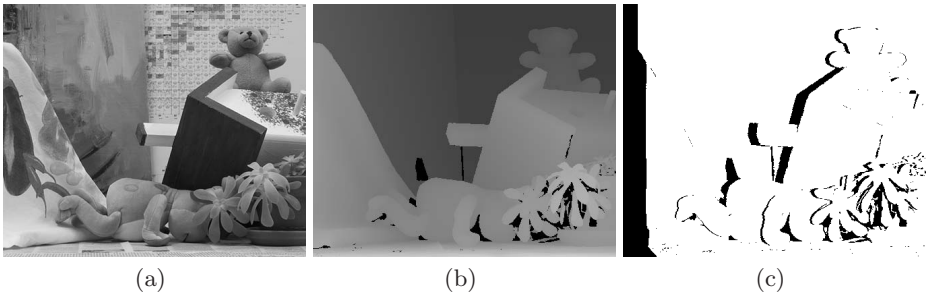


Fig. 2. (a) `im2` of the Middlebury “Teddy” stereo pair; (b) the corresponding ground truth disparity map; (c) a mask for the pixels, which are also visible in `im6`

Moreover, this (rectified) stereo setting simplifies discussing the effects caused by the relaxation and the seemingly asymmetric formulation.

All experiments in this section use the grayscale version of the “Teddy” stereo pair [25] and a set of constant disparity proposals in the range 0 to 59 pixels in 0.5 pixel increments. Figure 2(a) shows `im2` of the stereo pair, Fig. 2(b) the corresponding ground truth disparity, and Fig. 2(c) is the mask of non-occluded regions.

Relaxation and Thresholding. For every fusion step (4), the binary function $\phi : \Omega \rightarrow \{0, 1\}$ has to be optimized. Since this is a non-convex problem, we proposed to relax ϕ to the range $[0, 1]$, solve the continuous problem, and finally threshold ϕ . Obviously, this only leads to reasonable fusions of α and β , if the optimal ϕ is close to binary. Figure 3 shows, how a 64-bin histogram of the relaxed function ϕ evolves during a typical optimization procedure. Since ϕ is initialized with 0, in the beginning proposal α is chosen over β in the whole image. However, the “traces” in Fig. 3 illustrate that in several image regions the value of ϕ flips to 1, i.e. in these regions proposal β is energetically favored. Once the algorithm converged, the histogram of ϕ is close to binary, just as expected.

Symmetry. Since we initialize $\phi(\mathbf{x}) = 0$, the formulation appears to be asymmetric with respect to the flow proposals α and β , but in practice the effects of swapping the proposals is negligible. Figure 4(a) shows some intermediate solution of the “Teddy” disparity map, which is used as proposal α in the following experiment. The energy of α is 325963 and there clearly is room for improvement. The disparity map resulting from a fusion with the constant proposal $\beta(\mathbf{x}) = 20$ has an energy of 284748 and can be seen in Fig. 4(b). To highlight the image

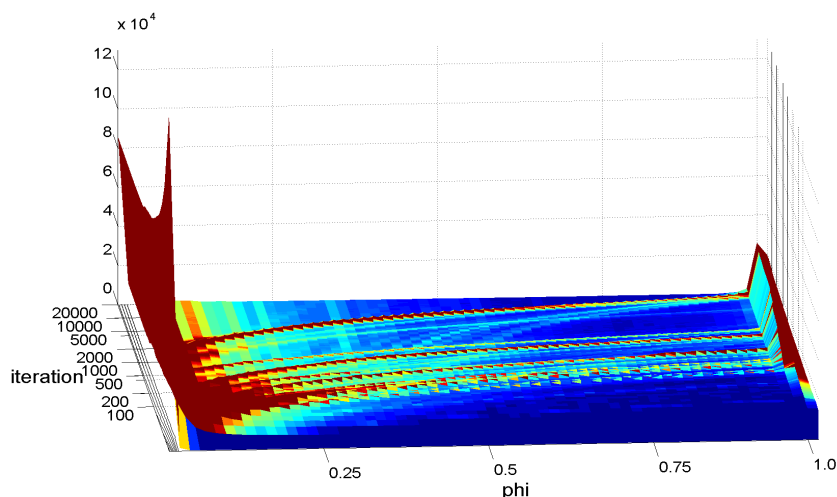


Fig. 3. The evolution of a 64-bin histogram of ϕ during the solution procedure. Starting at $\phi(\mathbf{x}) = 0$, i.e. with proposal α , several image regions flip to the alternative proposal β . It is clearly apparent that the converged histogram is close to binary. Please note that a logarithmic scale is used for the “iteration” axis.

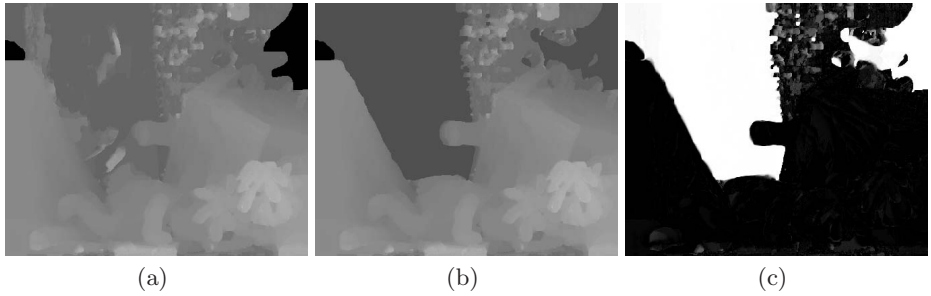


Fig. 4. Two intermediate disparity maps, before (a) and after (b) a binary fusion with the proposal $\beta(\mathbf{x}) = 20$. (c) shows the corresponding continuous optimum of ϕ .

regions that were switched to proposal β during this fusion step, Fig. 4(c) shows the continuous optimum of ϕ (before thresholding). Please note that ϕ is mostly binary, except for regions where neither α nor β are close to the true disparity. Repeating the experiment with α and β switched leads to visually indistinguishable results and an energy of 284727, i.e. the order of the proposals does *not* matter in the binary fusion step.

Disparity Estimation via Randomized Sweeping. In the discrete α -expansion algorithm [5], the minimization of a multi-label problem is performed by “agglomerating” an increasingly accurate solution by repeatedly allowing the current labeling to switch to an alternative label. One sweep through the label space is called a *cycle*. In the following experiment on the “Teddy” stereo pair, we

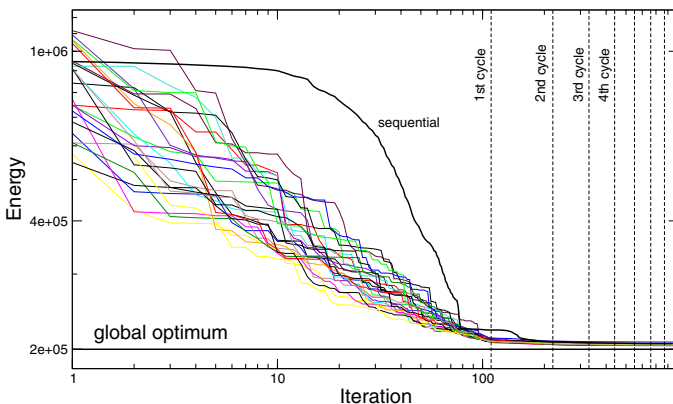


Fig. 5. Decrease of the energy of the flow field with every successful fusion. Results of randomized runs are shown as thin, colored lines; the thick, black line shows the progress for a sequential run. The dashed vertical lines delimit fusion cycles, the thick horizontal line marks the global optimum for the TV- L^1 flow model (6).

emulate this behavior by sweeping through a set of constant disparity proposals in the range 0 to 59 in 0.5 pixel increments. Figure 5 depicts, how the energy of the flow field decreases with every successful fusion step. Sweeping the “label space” in consecutive order of the displacements results in the thick, black line labeled “sequential”. The results for 25 distinct runs with a randomized order of the proposals are illustrated using thin, colored lines. The dashed vertical lines delimit fusion cycles, i.e. between two of those vertical lines every possible disparity is tested exactly once. The thick horizontal line represents the energy of the global optimum ($E = 199891$) for this model at $\lambda = 50$, see Fig. 6(c) for a disparity map of this solution. It is clearly apparent that after three or more cycles the particular sweeping order does not have a significant influence. After eight cycles, the mean energy of all runs is 205114, with the best run being roughly 2% better than the worst. Two exemplary disparity maps are shown in Figs. 6(a) and 6(b) – they differ from the global optimum (shown in Fig. 6(c)) mainly in occluded areas, i.e. areas, where Fig. 2(c) is black. The mean energy of the disparity maps, estimated using this approximated minimization technique, is 2.6% higher than the global optimum, which is consistent with empirical results reported for the discrete α -expansion algorithm.

Depending on the quality of the model, reducing the energy of a solution might not result in a lower error on the true problem. Hence, in order to compare the true errors of the global optimum and the approximated disparity fields, the results have been evaluated on the Middlebury stereo vision benchmark [25]. At an error threshold of 0.5, the global optimum mislabels 16.9% of the pixels in non-occluded regions; the estimated solution shown in Fig. 6(b) has a slightly larger error of 17.2%, which is consistent with the energy differences.

4.2 Optical Flow Estimation

For optical flow estimation, merely fusing a set of constant proposals, just like we did in the stereo case, is not feasible. One of the main reasons is that the expected range of motion is not known beforehand. Simply assuming very large ranges to be on the safe side either results in a coarse sampling of the solution space or in a huge number of constant flow proposals. Such a brute force method could be made tractable by using a scale pyramid, but this would still be inelegant and slow. However, the presented optimization strategy is not limited to fusing constant proposals – any flow field can be used as proposed solution. Thus, one obvious solution to the problem is to estimate a set of flow fields using a standard optical flow algorithm and fuse them. Therefore, before starting any fusion experiments, a set of 27 TV-regularized flows ($\lambda \in \{10, 25, 40, 65, 100, 150, 200, 500, 1000\}$, $\theta \in \{0.05, 0.1, 0.15\}$) and a set of 24 second-order prior regularized flows ($\lambda \in \{10, 25, 40, 55, 90, 200, 500, 1000\}$, $\theta \in \{0.05, 0.1, 0.15\}$) have been estimated using the algorithms described in [16,21]. Since the proposed optimization strategy is not limited to convex data terms, we used a truncated data term $\rho = \min\{1, 1 - r\}$, where r is the normalized cross-correlation, calculated on 3×3 patches across all color channels.

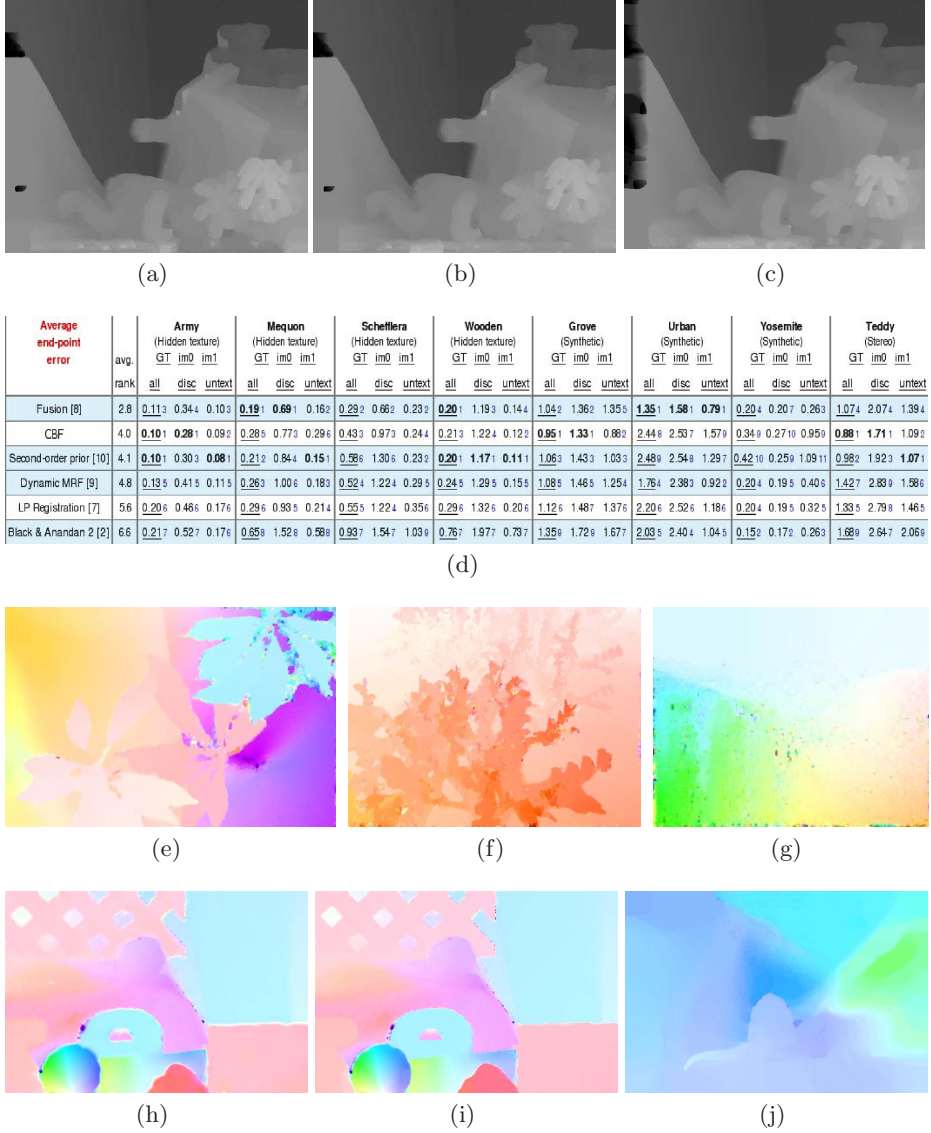


Fig. 6. First row: (a), (b) disparity maps, estimated by repeatedly fusing constant proposals using the proposed optimization strategy; (c) global optimum for the TV- L^1 model. Second row: average end-point error results for the Middlebury benchmark dataset; the proposed method, labeled “CBF”, was ranked 2^{nd} at the time of submission. Third row: (e–g) show color-coded flow fields for the sequences “Schefflera”, “Grove”, and “Yosemite”. Last row: (h) shows a color-coded flow field for the Middlebury “RubberWhale” sequence, estimated using the second-order prior (AAE = 3.14°); (i) and (j) show the color-coded flow fields for a TV-regularized fusion of TV and second-order prior flows for the “RubberWhale” (AAE = 2.87°) and the “Dimetrodon” (AAE = 3.24°) sequences.

A TV-regularized fusion of all 27 TV flows has been submitted to the Middlebury optical flow evaluation site, where at the time of submission this method (labeled “CBF”) was ranked second or better for 14 out of 16 error measures. Figure 6(d) shows the average end-point error of the six top-ranked algorithms, and Figs. 6(e)–6(g) show some color-coded results. Please visit the evaluation page at <http://vision.middlebury.edu/flow/eval/> for other error measures and further images. Due to the challenging diversity of the data sets, this was the only experiment where we used $\lambda = 90$ to improve the results on “Schefflera” and “Grove” at the cost of a rather noisy “Yosemite” result. On the “Urban” sequence (640×480 pixels), the outlined flow estimation procedure took 218s (138s for precalculating 27 TV-regularized flows and 80s for 6 fusion cycles).

Comparing the Figs. 6(e) and 1(c) indicates a slight improvement of the estimated flow, but since we only fuse TV- L^1 proposals, the correct solution is never “offered.” For further improvements, we have to resort to the brute force strategy of fusing a number of constant flow proposals, but since we already have a good estimate of the flow, the solution space is small and it quickly converges. The final result has an average angular error (AAE) of 2.91° (see Fig. 1(d)).

Furthermore, the second-order prior regularized optical flow algorithm (see Section 3.1) was used to fuse the 24 precalculated second-order flows. Figure 6(h) shows a color-coded result for the “RubberWhale” sequence of the Middlebury training dataset (AAE = 3.14°). Using *all* precalculated flows and a TV-regularized fusion algorithm yields even better results: Figs. 6(i) and 6(j) show the color-coded flow fields for the “RubberWhale” (AAE = 2.87°) and “Dimetrodon” (AAE = 3.24°) sequences, respectively.

5 Conclusion

The presented optimization strategy permits large optimization moves in a variational context, by restating the minimization problem as a sequence of binary subproblems. After verifying that the introduced approximations are reasonable, we showed that typical solutions for a stereo problem are within a few percent of the global optimum (in energy as well as in the true error measure). Finally, we showed that applying this optimization strategy to optical flow estimation yields state-of-the-art results on the challenging Middlebury optical flow dataset.

References

1. Mumford, D.: Bayesian rationale for energy functionals. In: Geometry-driven diffusion in Computer Vision, pp. 141–153. Kluwer Academic Publishers, Dordrecht (1994)
2. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(6), 1068–1080 (2008)
3. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)

4. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(10), 1568–1583 (2006)
5. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 1222–1239 (2001)
6. Veksler, O.: Graph cut based optimization for MRFs with truncated convex priors. In: *Proc. of the CVPR* (June 2007)
7. Lempitsky, V., Rother, C., Blake, A.: LogCut – efficient graph cut optimization for Markov random fields. In: *Proc. of the ICCV* (October 2007)
8. Komodakis, N., Tziritis, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic MRFs. In: *Proc. of the CVPR* (June 2007)
9. Chambolle, A.: Total variation minimization and a class of binary MRF models. *Energy Minimization Methods in Comp. Vision and Pattern Rec.* 136–152 (2005)
10. Nikolova, M., Esedoglu, S., Chan, T.F.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. on App. Math.* 66 (2006)
11. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artificial Intelligence* 17, 185–203 (1981)
12. Black, M.J., Anandan, P.: A framework for the robust estimation of optical flow. In: *Proc. of the ICCV*, pp. 231–236 (May 1993)
13. Aubert, G., Deriche, R., Kornprobst, P.: Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics* 60(1), 156–182 (2000)
14. Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision* 67(2), 141–158 (2006)
15. Bruhn, A., Weickert, J., Kohlberger, T., Schnörr, C.: A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision* 70(3), 257–277 (2006)
16. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L1 optical flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) *DAGM 2007*. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)
17. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M., Szeliski, R.: A database and evaluation methodology for optical flow. In: *Proc. of the ICCV* (2007)
18. Aujol, J.F., Gilboa, G., Chan, T.F., Osher, S.: Structure-texture image decomposition – modeling, algorithms, and parameter selection. *International Journal of Computer Vision* 67(1), 111–136 (2006)
19. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
20. Chambolle, A.: An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision* 20, 89–97 (2004)
21. Trobin, W., Pock, T., Cremers, D., Bischof, H.: An unbiased second-order prior for high-accuracy motion estimation. In: Rigoll, G. (ed.) *DAGM 2008*. LNCS, vol. 5096, pp. 396–405. Springer, Heidelberg (2008)
22. Danielsson, P.E., Lin, Q.: Efficient detection of second-degree variations in 2D and 3D images. *Journal of Visual Comm. and Image Representation* 12, 255–305 (2001)
23. Kohli, P., Kumar, P., Torr, P.H.: P3 & beyond: Solving energies with higher order cliques. In: *Proc. of the CVPR* (June 2007)
24. Pock, T., Schoenemann, T., Cremers, D., Bischof, H.: A convex formulation of continuous multi-label problems. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008*. LNCS, vol. 5304, pp. 792–805. Springer, Heidelberg (2008)
25. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: *Proc. of the CVPR*, vol. 1, pp. 195–202 (June 2003)