# FusionFlow: Discrete-Continuous Optimization for Optical Flow Estimation

Victor Lempitsky
Microsoft Research Cambridge

Stefan Roth
TU Darmstadt

Carsten Rother
Microsoft Research Cambridge

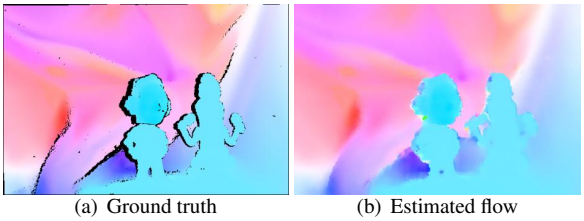
## Abstract

*Accurate estimation of optical flow is a challenging task, which often requires addressing difficult energy optimization problems. To solve them, most top-performing methods rely on continuous optimization algorithms. The modeling accuracy of the energy in this case is often traded for its tractability. This is in contrast to the related problem of narrow-baseline stereo matching, where the top-performing methods employ powerful discrete optimization algorithms such as graph cuts and message-passing to optimize highly non-convex energies.*

*In this paper, we demonstrate how similar non-convex energies can be formulated and optimized discretely in the context of optical flow estimation. Starting with a set of candidate solutions that are produced by fast continuous flow estimation algorithms, the proposed method iteratively fuses these candidate solutions by the computation of minimum cuts on graphs. The obtained continuous-valued fusion result is then further improved using local gradient descent. Experimentally, we demonstrate that the proposed energy is an accurate model and that the proposed discrete-continuous optimization scheme not only finds lower energy solutions than traditional discrete or continuous optimization techniques, but also leads to flow estimates that outperform the current state-of-the-art.*

## 1. Introduction

Optical flow has been an important area of computer vision research, and despite the significant progress made since the early works [14, 20], flow estimation has remained challenging to this date. Two challenges dominate recent research: firstly, the issue of choosing an appropriate computational model, and secondly, computing a good solution given a particular model. Papenberg *et al*. [23], for example, suggested a complex continuous optimization scheme for flow estimation. Despite its success, this approach is limited by the fact that the spatial regularity of flow is modeled as a convex function. Consequently, the estimated flow fields are somewhat smooth and lack very sharp discontinuities that exist in the true flow field, especially at motion boundaries. Black and Anandan [3] addressed this problem by introducing non-convex penalty functions based on robust statistics that are tolerant towards such outliers. While this formulation allowed for realistic discontinuities, the corresponding non-convex energy was hard to minimize. Despite using annealing, the optimization could often get trapped in poor local optima. Roth and Black [25] studied the spatial statistics of optical flow and found the derivative statistics to be very heavy-tailed. This validated the assumptions made by [3] and implied that the convex energies used by the majority of today's flow approaches only provide a relatively crude approximation to the statistics of the flow. They also suggested a Markov random field (MRF) model motivated by these statistics, but due to the corresponding non-convex energies, inference has remained very challenging, and the flow fields estimated using a continuous optimization approach still suffer from smoothed discontinuities.



(a) Ground truth (b) Estimated flow

Figure 1. With a new, more accurate MRF energy and a powerful discrete-continuous optimization approach, the proposed method is capable of accurate flow recovery from challenging image sequences such as "Mequon" from [1] (*hue = direction, saturation = magnitude, black in GT = "unknown"*).

Surprisingly, for narrow-baseline stereo matching these difficulties have been addressed quite a while ago, and the need for non-convex energy functions has not only been recognized but, in contrast to optical flow, also been widely addressed. Their use has been facilitated by modern discrete optimization algorithms, such as graph cuts [7] or message passing [28], which are often able to obtain nearly-global optima of such energies [21]. Most top-performing stereo techniques rely on discrete optimization for minimizing non-convex energies. Even though disparity estimation has a lot in common with optical flow estimation and can, in fact, be regarded as a particular constrained case of it, surprisingly little knowledge has been transferred to the optical flow problem. Only a few authors have attempted to use dis-

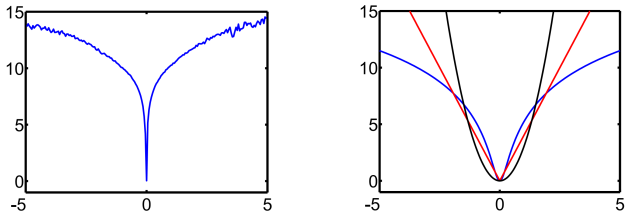


Figure 2. *(Left)* Negative log-density of the x-derivative of the horizontal flow (from [25]). *(Right)* Charbonnier potentials (red, *c.f*. [9]) are not a good fit of these statistics, but better than quadratics (black). Here, we instead use the negative-log of a Student-t distribution (Lorentzian) (blue, *c.f*. [3]).

crete optimization for optical flow [7, 10, 11, 19, 27], however, none of them reported state-of-the-art performance.

This work presents a new approach to the estimation of optical flow called *FusionFlow*. Similar to [3] and motivated by the statistics of optical flow [25], we employ a heavy-tailed spatial term as well as a robust data term, which permit modeling sharp discontinuities and occlusions in the flow. Similar to stereo methods, we rely on discrete optimization to minimize the resulting highly non-convex energy, but our optimization is not purely discrete. Instead, it relies on a set of continuous-valued candidate flow fields, which are obtained via fast and simple flow algorithms such as the original Lucas-Kanade [20] and Horn-Schunck [14] methods. These *proposal* solutions are subsequently combined together using discrete optimization. Following [19], the optimal combination (*fusion*) of each pair of solutions with respect to the considered energy is computed from the minimum cut on a special graph [4, 5, 16]. Afterwards, the fused solution is locally improved using gradient descent. The suggested optimization scheme relies on discrete algorithms in order to avoid poor local minima of the energy, but operates with continuously-valued proposal solutions. This avoids the pitfalls of purely discrete approaches, namely the huge number of labels as well as discretization artifacts.

We first motivate our new non-convex energy in Sec. 2, and then introduce the proposed discrete-continuous optimization approach in Sec. 3. It allows for more efficient and robust optimization compared to hierarchical coarse-to-fine continuous frameworks, which are very often used for flow estimation. In Sec. 4 we show that our non-convex energy combined with our optimization procedure leads to high-quality flow results, superior to typical convex formulations. At the time of publication, the proposed algorithm is the top-ranking approach across the majority of evaluation metrics of the Middlebury optical flow benchmark [1].

### 1.1. Background and Related Work

In many cases, optical flow estimation has been posed as a problem of energy minimization. The use of energy minimization dates back to the work of Horn and Schunck [14], where given a pair of images $I(x, y, t)$ and $I(x, y, t+1)$ the optical flow field $(u, v)$ is computed by minimizing an energy of the form

$$E(u,v) = \int_\Omega D\left(I(x+u(x,y), y+v(x,y)) - I(x,y)\right) + S\left(||\nabla u(x,y)||^2 + ||\nabla v(x,y)||^2\right)\, dx\, dy\,. \quad (1)$$

Here $\Omega$ denotes the image region, and $(x, y)$ denotes the image coordinate. The first term, the so-called *data term*, embodies the *brightness constancy assumption*, which states that corresponding pixels in the two frames should have similar brightness. The second term, also called *spatial term*, imposes spatial regularity on the resulting optical flow field by penalizing spatially varying flow. This is necessary, because flow estimation is severely underconstrained and suffers, for example, from the aperture problem. Historically, [14] proposed using quadratic penalties for both data and spatial terms, and further linearized the brightness constancy assumption. This makes the energy convex, and after spatial discretization quite easy to optimize.

Nevertheless, this simple formulation has a variety of problems. The linearized brightness constancy assumption allows estimating only very small motions, which has prompted the use of hierarchical coarse-to-fine estimation schemes, in which the solution from a coarser level is used to warp the frames toward each other. Interestingly, these can be interpreted as numerical schemes for approximately minimizing energies with non-linearized, thus non-convex, brightness constancy terms [23]. Another problem relates to the fact that a quadratic spatial term leads to overly smooth flow estimates without any discontinuities, as they naturally occur at motion boundaries. Even though robust non-convex penalty functions have been proposed [3] and are motivated by the flow statistics (see Fig. 2), they were hard to optimize with traditional continuous optimization approaches. Many subsequent methods thus often used slightly robust, but still convex penalty functions (*e.g*. [9, 23]), which still lead to smoothed discontinuities.

Some approaches recover more accurate motion discontinuities using explicit discontinuity modeling either using line processes [3] or explicit segmentation of the flow fields [8, 22]. In both cases, the estimation of discontinuities is alternated with flow estimation, which also makes these approaches vulnerable to getting trapped in local minima.

A number of authors have formulated the problem of flow estimation using a Markov random field, in which case flow is usually computed using maximum a-posteriori (MAP) estimation [3, 13, 18]. Inference was performed using a number of approximative techniques, such as stochastic or deterministic relaxation, all of which are vulnerable to getting trapped in relatively poor local optima.

A number of attempts have been made to adapt discrete optimization methods for optical flow computation [7, 10, 11, 19, 27]. All of them, however, suffered from

the problem of *label discretization*. While in stereo it is relatively easy to discretize the disparities, this is not the case for optical flow, where at each pixel a two-dimensional flow vector has to be described using discrete labels.

Our approach may also be related to layered-based correspondence methods such as [2]. These methods assume that the scene decomposes into a small set of layers with few parameters (such assumption, however, rarely holds for real scenes). The energy, which is dependent both on the non-local layer parameters as well as on the pixel assignments to layers, is minimized by alternating discrete optimization updating pixel assignments and continuous optimization updating layer parameters. Despite the use of discrete algorithms, such optimization still often gets stuck in poor local minima.

## 2. Energy Formulation

Following a number of optical flow approaches (*e.g.* [3, 13, 18]) as well as a large body of work in stereo, we model the problem of optical flow estimation using pairwise Markov Random Fields. Flow estimation is performed by doing maximum a-posteriori (MAP) inference. As we will see, this provides a good trade-off between the accuracy of the model as well as inference, *i.e.* optimization, tractability. The posterior probability of the flow field $\mathbf{f}$ given two images, $I_0$ and $I_1$, from a sequence is written as

$$\begin{aligned} p(\mathbf{f}|I^0, I^1) \;&= \frac{1}{Z} \prod_{\mathbf{p}\in\Omega} \exp(-D_{\mathbf{p}}(f_{\mathbf{p}};\; I^0, I^1)) \cdot \\ &\prod_{(\mathbf{p},\mathbf{q})\in\mathcal{N}} \exp(-S_{\mathbf{p},\mathbf{q}}(f_{\mathbf{p}}, f_{\mathbf{q}}))\;, \end{aligned} \tag{2}$$

where $f_{\mathbf{p}} = (u_{\mathbf{p}}, v_{\mathbf{p}})$ denotes the flow vector at pixel $\mathbf{p}$, the set $\mathcal{N}$ contains all pairs of adjacent pixels, and $Z$ is a normalization constant. Before specifying the model in more detail, we obtain an equivalent energy function by taking the negative logarithm of the posterior and omitting constants:

$$E(\mathbf{f}) = \sum_{\mathbf{p}\in\Omega} D_{\mathbf{p}}(f_{\mathbf{p}};\; I^0, I^1) + \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{N}} S_{\mathbf{p},\mathbf{q}}(f_{\mathbf{p}}, f_{\mathbf{q}})\;. \tag{3}$$

This can be viewed as a spatially discrete variant of Eq. (1).

**Data term.** The first term of the energy, the so-called data term, measures how well the flow field $\mathbf{f}$ describes the image observations. In particular, it models how well the corresponding pixels of $I^0$ and $I^1$ match. Traditionally, it is modeled based on the brightness (or color) constancy assumption, *e.g.*, $D_{\mathbf{p}}(f_{\mathbf{p}};\; I^0, I^1) = \rho_d(||I^1(\mathbf{p} + f_{\mathbf{p}}) - I^0(\mathbf{p})||)$. Note that we do not employ a linearized constraint as our optimization method does not require this. We found, however, that such simple color matching is heavily affected by illumination and exposure changes, in particular by shadows. To make the data term robust to these effects, we remove lower spatial frequencies from consideration:

$$H^i = I^i - G_\sigma * I^i, \qquad i \in 0, 1\;, \tag{4}$$

where $G_\sigma$ is a Gaussian kernel with standard deviation $\sigma$. Based on these filtered images we define

$$D_{\mathbf{p}}(f_{\mathbf{p}};\; I^0, I^1) = \rho_d(||H^1(\mathbf{p} + f_{\mathbf{p}}) - H^0(\mathbf{p})||)\;. \tag{5}$$

Here $||\cdot||$ denotes the Euclidean distance of the RGB color values. $H^1(\mathbf{p}+f_{\mathbf{p}})$ is computed using bicubic interpolation.

As suggested by the probabilistic interpretation from Eq. (2), the penalty function $\rho(\cdot)$ should model the negative log-probability of the color distance for natural flow fields. Although the statistics of color constancy for optical flow have not been studied rigorously so far, the presence of effects such as occlusions and specular reflections suggests that a robust treatment using heavy-tailed distributions is necessary to avoid penalizing large color changes unduly (cf. [3]). Therefore, we use the Geman-McClure robust penalty function:

$$\rho_d(x) = \tfrac{x^2}{x^2+\mu^2}\;, \tag{6}$$

which has a similar shape as a truncated-quadratic penalty $f(x) = \min(ax^2, 1)$ that has been very successfully used in the stereo literature, but is differentiable everywhere.

**Spatial term.** As is usual in a pairwise MRF formulation of flow, the spatial term penalizes changes in horizontal and vertical flow between adjacent pixels:

$$S_{\mathbf{p},\mathbf{q}} = \rho_{\mathbf{p},\mathbf{q}}\left(\frac{u_{\mathbf{p}} - u_{\mathbf{q}}}{||\mathbf{p}-\mathbf{q}||}\right) + \rho_{\mathbf{p},\mathbf{q}}\left(\frac{v_{\mathbf{p}} - v_{\mathbf{q}}}{||\mathbf{p}-\mathbf{q}||}\right)\;,$$

where $||\mathbf{p}-\mathbf{q}||$ is the Euclidean distance between the pixel centers of $\mathbf{p}$ and $\mathbf{q}$. As the flow differences between adjacent pixels closely approximate the spatial derivatives of the flow, we use the spatial statistics of optical flow [25] to motivate suitable penalty functions. Roth and Black [25] showed that the statistics of flow derivatives are very heavy-tailed and strongly resemble Student t-distributions. We therefore choose the penalty to be the (scaled) negative log of a Student-t distribution (see also Fig. 2):

$$\rho_{\mathbf{p},\mathbf{q}}(x) = \lambda_{\mathbf{p},\mathbf{q}} \log\left(1 + \tfrac{1}{2\nu^2}x^2\right)\;. \tag{7}$$

Motivated by the success of stereo approaches, we assume that the flow field discontinuities tend to coincide with image color discontinuities. Hence we make the smoothness weight $\lambda_{\mathbf{p},\mathbf{q}}$ spatially-dependent and set it to a lower value if the pixel values $I^0(\mathbf{p})$ and $I^0(\mathbf{q})$ are similar.

While being of high fidelity, as we will see shortly, the proposed MRF energy is more difficult to optimize than the energies used in recent popular optical flow methods such

as [9, 23]. While the non-linearized color constancy assumption already makes the objective non-convex [23], the penalty functions of data and spatial term in our formulation are robust, thus non-convex as well. Finally, the data term works with the high frequency content of images, which only adds to its non-linearity. Therefore, as we demonstrate in the experimental section, traditional continuous optimization schemes based on coarse-to-fine estimation and gradient descent often end up in poor local minima. Also, the proposed energy is harder to optimize than many energies used in stereo matching, since the value at each pixel spans a potentially unbounded 2D rather than a bounded 1D domain, making it infeasible for purely discrete techniques to sample it densely enough. This suggests the use of a new, more powerful optimization scheme that combines the merits of discrete and continuous-valued approaches.

# 3. Energy Minimization

## 3.1. Graph cut methods for energy minimization

Over the last years, graph cut based methods have proven to be invaluable for the minimization of pairwise MRF energies in the case of discrete labels (*e.g.*, $x_{\mathbf{p}} \in \{0, 1, \dots N-1\}$), which take the form:

$$E(\mathbf{x}) = \sum_{\mathbf{p}\in\Omega} D_{\mathbf{p}}(x_{\mathbf{p}}) + \sum_{\mathbf{p},\mathbf{q}\in\mathcal{N}} S_{\mathbf{p},\mathbf{q}}(x_{\mathbf{p}}, x_{\mathbf{q}})\,. \qquad (8)$$

These methods rely on the fact that in the case of MRFs with binary labels ($N$=2) finding the global minimum can be reduced to computing the minimal cut on a certain network graph [5, 12]. The existing algorithms based on the maxflow/mincut duality find the minimal cut (and hence the global minimum of the MRF energy) very efficiently [6]. The graph construction proposed in [12, 17], however, worked only for MRFs with a certain *submodularity* condition imposed on the pairwise terms. In the cases when these submodularity conditions are not met, a *partial* global optimum may still be computed via minimum cut on an extended graph [4, 5, 16]. Here, partiality implies that the label cannot be determined from the minimal cut for some of the nodes. However, the remaining (labeled) nodes are assigned the same label as they have in the global optimum. The number of nodes with unknown labels depends on the structure of the problem: connectivity, number of submodularity constraints violated, *etc*. [26].

Various ways have been suggested to extend graph cut based methods to MRFs with multiple labels ($N$>2) [7, 15, 19]. In particular, recently [19] suggested the *fusion move* approach to be used in this context. The fusion move considers two given $N$-valued labelings $\mathbf{x}^0$, $\mathbf{x}^1$ and introduces an auxiliary binary-valued labeling $\mathbf{y}$. The set of all possible auxiliary labelings naturally corresponds to the set of *fusions* of $\mathbf{x}^0$ and $\mathbf{x}^1$ (*i.e.*, labelings of the original MRF, where each node $\mathbf{p}$ receives either label $x^0_{\mathbf{p}}$ or label $x^1_{\mathbf{p}}$):

$$\mathbf{x}^f(\mathbf{y}) = (\mathbf{1}-\mathbf{y})\cdot\mathbf{x}^0 + \mathbf{y}\cdot\mathbf{x}^1\,, \qquad (9)$$

where the product is taken element-wise. This induces a binary-labeled MRF over the auxiliary variables with the energy defined as:

$$E^f(\mathbf{y}) = E(\mathbf{x}^f(\mathbf{y})) = \sum_{\mathbf{p}\in\Omega} d_{\mathbf{p}}(y_{\mathbf{p}}) + \sum_{\mathbf{p},\mathbf{q}\in\mathcal{N}} s_{\mathbf{p},\mathbf{q}}(y_{\mathbf{p}}, y_{\mathbf{q}}),$$
$$\text{where}\quad d_{\mathbf{p}}(i) = D_{\mathbf{p}}(x^i_{\mathbf{p}}),\quad s_{\mathbf{p},\mathbf{q}}(i,j) = S_{\mathbf{p},\mathbf{q}}(x^i_{\mathbf{p}}, x^j_{\mathbf{q}}).$$

The minimum of this binary-valued MRF can be computed via minimum cut on the extended graph. It corresponds to the fusion of $\mathbf{x}^0$ and $\mathbf{x}^1$ that is optimal with respect to the original energy from Eq. (8). Consequently, the energy of this optimal fusion will not be higher (and in most cases lower) than the energy of both $\mathbf{x}^0$ and $\mathbf{x}^1$. This crucial property of the fusion move algorithm can be enforced even when the obtained optimal auxiliary labeling is not complete (in this case all the unknown labels are taken from the original solution with lower energy).

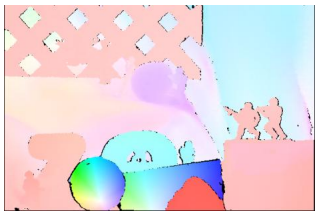
The fusion move generalizes the $\alpha$-expansion move and the $\alpha\beta$-swap move proposed in [7]. Thus, the popular $\alpha$-expansion algorithm may be regarded as subsequent fusions of the current labeling with different constant labelings. Our FusionFlow framework relies on discrete-continuous optimization that also uses fusion moves, here to combine proposal solutions. In particular, it relies on the fact that fusion moves can be applied even in the case of MRFs with continuous-valued labels such as in Eq. (3), which avoids discretization artifacts.

## 3.2. Discrete-continuous energy minimization

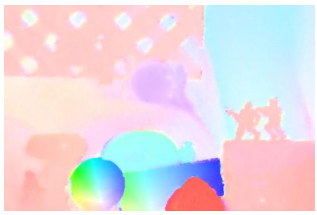
**Proposal solutions.** The discrete part of our algorithm proceeds by first computing many different continuous-valued flow fields that serve as *proposal* solutions, and then iteratively fusing them with the current solution using graph cuts. After each fusion different parts of the proposal solution are copied to the current solution so that the energy goes down (or stays equal). The success of the method thus depends on the availability of good proposal solutions.

It is important to note that the proposal solutions need not to be good in the whole image in order to be "useful". Instead, each solution may contribute to a particular region in the final solution, if it contains a reasonable flow estimate for that region, no matter how poor it is in other regions. This suggests the use of different flow computation methods with different strengths and weaknesses for computing the proposals. In our experiments, we used three kinds of the proposal solutions. Firstly, we used solutions obtained using the Lucas-Kanade (LK) method [20]. Due to the properties of the method, such solutions often contain good results for textured regions but are virtually useless in textureless regions. Secondly, we used solutions obtained using

(a) First Frame

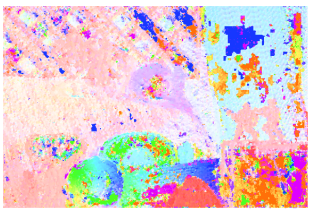


(b) Ground truth



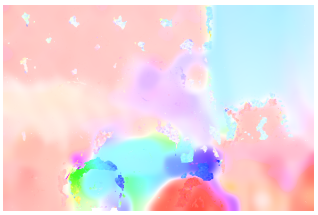
(c) Final Solution, Energy=2041

(d) Solution 1, Energy=7264



(e) Solution 2, Energy=44859

(f) Auxiliary variables



(g) Fused solution, Energy=6022

Figure 3. Results for the *Army* sequence from [1]. The bottom row shows **the first step of our discrete optimization** (*where "Lucas-Kanade meets Horn-Schunck"*). Here, a randomly chosen initial solution (d) (computed with Horn-Schunck) is fused with another randomly chosen proposal solution (e) (computed with Lucas-Kanade). The graph cut allows to compute the optimal fused solution (g) with much lower energy, which is passed on to the next iteration. The optimal auxiliary variables (f) show which regions are taken from Solution 1 (black) and from Solution 2 (white). In this example 99.998% of the nodes were labeled by the minimum cut on the extended graph.

the Horn-Schunck (HS) method [14]. Such solutions often contain good results for regions with smooth motion, but motion discontinuities are always severely oversmoothed. Finally, we also used constant flow fields as proposals.

To obtain a rich set of proposal solutions, we use the LK and HS methods with various parameter settings. For HS we vary the strength of the regularization ($\lambda \in \{1, 3, 100\}$). Since both methods should be applied within a coarse-to-fine warping framework to overcome the limitations of the linearized data term (of the proposals, not of our energy), we also vary the number of levels in the coarse-to-fine hierarchy ($l \in \{1, \ldots, 5\}$). Finally, for all LK solutions and a few HS solutions we produce shifted copies (by shifting $\pm 2^{l-1}$ and $\pm 2^l$ pixels in each direction). For the LK method, this corresponds to the use of a family of non-centralized windows and, hence, gives better chances of providing correct flow values near flow discontinuities, and as we found reduces the energy of the solution. These variations result in about 200 proposals (most of them, however, are shifted copies and do not take much time to compute). 64 constant flow fields are also added to the set of proposals. The choice of the constants is discussed below.

It is important to note that other (potentially more efficient) approaches for obtaining proposal solutions may also be considered. We also experimented with proposal solutions by performing gradient descent from different starting points. We found this to lead to good results (given a sufficient number of minima), but did not pursue it in our final experiments due to its computational inefficiency.

**Discrete optimization.** As described above, the proposal solutions are fused by computing the minimal cut on the extended graph. The process starts with the LK and HS proposal fields only. One of these proposal fields is randomly chosen as an initial solution. After that, the remaining LK and HS proposals are visited in random order, and each of them is fused with the current solution (as described in Sec. 3.1). An example of such a fusion (during the first iteration of the process) is shown in Fig. 3.

After all LK and HS solutions are visited, the motion vectors of the obtained fused solution are clustered into 64 clusters using the k-means algorithm. The centers of the clusters $\mathbf{c}_i \in \mathbb{R}^2$ are used to produce constant proposal flow fields $f_i(\mathbf{p}) \equiv \mathbf{c}_i$. Note that more sophisticated proposals dependent on the current solution may be suggested and our constant solutions are just one step in this direction.

The created constant proposals are added to the LK and HS proposals and the fusion process continues until each proposal is visited twice more. At this point the procedure typically converges, *i.e.*, the obtained fused solution can no longer be changed by fusion with any of the proposals.

We should note that the number of unlabeled nodes during each fusion was always negligible (we never observed it exceeding 0.1% of the nodes). Each fusion is guaranteed not to increase the energy, and in practice the resulting solution always has an energy that is much smaller than the energy of the best proposal.

**Continuous optimization.** After fusing flow fields using discrete optimization, we perform a continuous optimization step that helps "cleaning up" areas where the proposal solutions were not diverse enough, which for example may happen in relatively smooth areas. In order to perform continuous optimization, we analytically compute the gradient of the same energy we use in the discrete step, $\nabla_{\mathbf{f}} E(\mathbf{f})$, and use a standard conjugate gradient method [24]

to perform local optimization. The gradient of the spatial term is quite easily derived; computing the gradient of the data term relies on the fact that $H^1(\mathbf{p} + f_\mathbf{p})$ is computed using bicubic interpolation, which allows computing the partial derivatives w.r.t. $u_\mathbf{p}$ and $v_\mathbf{p}$. We should note that the gradient bears resemblance to the discretization of the Euler-Lagrange equations for the objective used in [23].

Since the discrete optimization step avoids many of the poor local optima that are problematic for purely continuous optimization methods, the combination of discrete and continuous optimization leads to local minima with a substantially lower energy in most of our experiments.

## 4. Evaluation

To evaluate the proposed approach and in particular the efficiency of the proposed energy minimization scheme in obtaining low-energy states of Eq. (3), we performed a number of experiments using the recent Middlebury optical flow benchmark dataset [1]. Since our method is applicable to color images and based on 2 frames, we used the 2-frame color versions of the datasets, and left the extension to multi-frame sequences for future work. In all experiments, we used an 8-neighborhood system for the spatial term and the following parameters, which were chosen to give good performance on challenging real-world scenes: $\sigma = 1.5$ for the high-pass filter, $\mu = 16$ and $\nu = 0.2$ for the MRF potentials; $\lambda_{\mathbf{p},\mathbf{q}} = 0.024$, if the sum of absolute differences between $I(\mathbf{p}, t)$ and $I(\mathbf{q}, t)$ was less than or equal to 30, and $\lambda_{\mathbf{p},\mathbf{q}} = 0.008$ otherwise.

We evaluated the proposed method on 8 benchmarking sequences and found that it outperforms other methods in the benchmark, particularly on the challenging real world scenes as they are shown in Fig. 1, Fig. 3 and Fig. 4. At the moment of publication, the proposed method was top-ranked on 10 out of 16 available performance measures including the average angular error (AAE) and the average end-point error. The only sequence, where our method does not perform well is the Yosemite sequence. This is mainly due to the fact that our parameters were chosen to give good performance on real-world sequences (increasing the smoothness weight by 16x without changing other parameters lowers the AAE on Yosemite from 4.55 to 2.33 degrees).

**Proposed energy vs. baseline energy.** To evaluate the advantage of the proposed energy, we also considered a simple "baseline" energy that is quite similar to the objectives in popular continuous methods [9, 23]. In particular, we again use an 8-neighborhood for the spatial term, but here with convex Charbonnier potentials ($Ch(x, \phi) = \phi^2\sqrt{1 + x^2/\phi^2}$), which are similar to the absolute difference measure (see also Fig. 2). The trade-off weight $\lambda$ was not adapted spatially. For the data term, we used gray-scale images as input from which we did not remove the low frequencies, and also relied on Charbonnier potentials. We optimized the energy using the approach proposed here, and tuned the parameters of the baseline model using grid search on "RubberWhale" (see Fig. 5). As can be seen in Fig. 5a and c, the proposed energy clearly outperforms the baseline model visually and quantitatively, even on the sequence used to tune the baseline energy parameters. In particular, the robust spatial potentials employed here allow to recover sharp discontinuities, while at the same time recovering smooth flow fields in continuous areas (Fig. 5c). Also, the robust data potentials allowed to attain better performance in occlusion areas. Finally, ignoring low frequency image content substantially improved the results in areas with shadows, such as on "Schefflera", from which the baseline model suffers.

**Our optimization vs. other schemes.** For our energy, we also compared the proposed discrete-continuous optimization method with a baseline continuous and a baseline discrete optimization scheme. For baseline continuous optimization, we employed a hierarchical coarse-to-fine estimation framework with 5 levels (*c.f.* [3]), where at each pyramid level we used the gradient descent scheme described in Sec. 3.2. As a baseline discrete algorithm, we ran $\alpha$-expansion (i.e. "conventional" graph cuts) [7]. To make the comparison more favorable for this baseline algorithm, we estimated the minimum and maximum horizontal and vertical flow from our high-quality solution (note that such an accurate estimate cannot be obtained directly from proposal solutions), and discretized this range uniformly to give about 1000 labels (*i.e.*, 4 times larger than the number of proposals that the fusion approach uses).

We found that neither of the two baseline energy minimization schemes gave good results for our energy consistently across the benchmark datasets, neither in terms of the energy, nor in terms of flow accuracy (see Tab. 1). In particular, the continuous baseline algorithm failed on most of the datasets (see, *e.g.*, Fig. 5b). This suggests that the proposed energy is simply too difficult for standard continuous optimization approaches, for example because of the non-convex potentials and the removal of low frequencies in the data term. While the behavior of the continuous optimization could be improved, for example using deterministic annealing (*c.f.* [3]), such heuristics rarely work well across a wide-range of datasets. The baseline discrete algorithm obviously suffered from the uniformity of discretization, especially for the datasets with large motion (*e.g.*, "Urban").

**How many proposals?** In our experiments we focused on the accuracy of optical flow estimation, thus using extensive numbers of proposals and an exhaustive number of iterations within continuous optimization. As a result, our unoptimized MATLAB implementation takes more than an hour for processing a single frame pair.

It is important, nevertheless, to emphasize that flow fu-

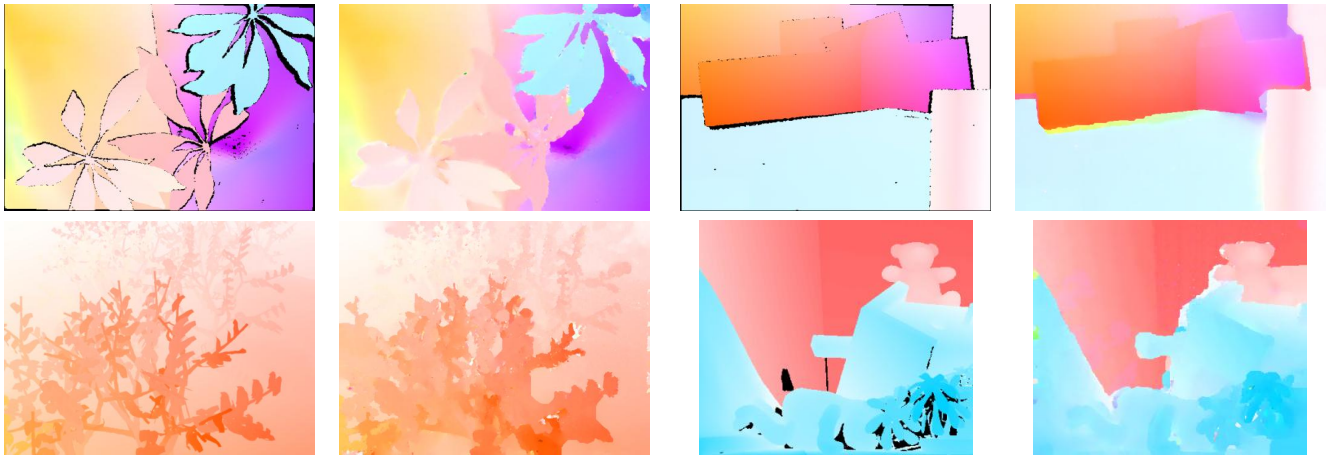


Figure 4. Example results on the benchmark datasets (right) along with ground truths (left).

| Optimiz. | Army | | Mequon | | Schefflera | | Wooden | | Grove | | Urban | | Yosemite | | Teddy | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ | AAE | $E(\mathbf{f})$ |
| Our (full) | **4.43** | **2041** | **2.47** | **3330** | **3.70** | **5778** | **3.68** | **1632** | 4.06 | **17580** | 6.30 | **5514** | 4.55 | **1266** | 7.12 | **9315** |
| Our (discrete) | 4.97 | 2435 | 4.83 | 4375 | 5.14 | 7483 | 5.24 | 2180 | **4.00** | 21289 | **6.27** | 6568 | **4.03** | 1423 | **6.68** | 10450 |
| Baseline cont. | 7.97 | 4125 | 52.3 | 21417 | 36.1 | 24853 | 16.8 | 7172 | 64.0 | 78122 | 46.1 | 26517 | 23.2 | 4470 | 63.9 | 31289 |
| Baseline disc. | 5.61 | 3038 | 5.19 | 6209 | 5.36 | 8894 | 4.94 | 2782 | 9.03 | 44450 | 18.7 | 17770 | 5.67 | 1995 | 9.13 | 15283 |

Table 1. Comparison of different optimization techniques: our full discrete-continuous, our discrete (*i.e.*, fusion of proposals without continuous improvement), baseline continuous, baseline discrete. Shown are the flow accuracy (average angular error) and the energy achieved. On the 8 test datasets [1], our optimization scheme consistently outperforms the two baseline algorithms.

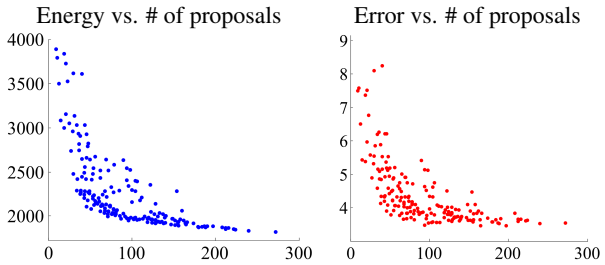


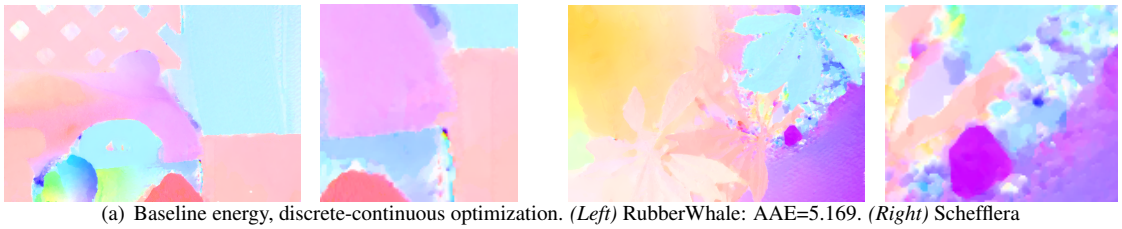

Figure 6. The energy and the average angular error (AAE) after the discrete optimization step for different subsets of proposals for "Rubber Whale" [1]. The $x$-coordinate of each dot corresponds to the number of proposals in the subset. The rightmost point on each plot corresponds to the full set of proposals. The plots suggest that sets of proposals that are 5 time smaller would do almost as well in our experiments.

sion idea is useful in scenarios where speed matters. Fig. 3 shows one such example, where fusion of just two motion fields (each computed with real-time methods) improves the result over both of them. Note that one fusion takes fractions of a second. Fig. 6 further explores the trade-off between the number of proposals and the quality of the solution for one of the sequences from [1]. It demonstrates that a five-fold reduction in the number of proposals leads to solutions that are only slightly worse than those computed with the full set of proposals.

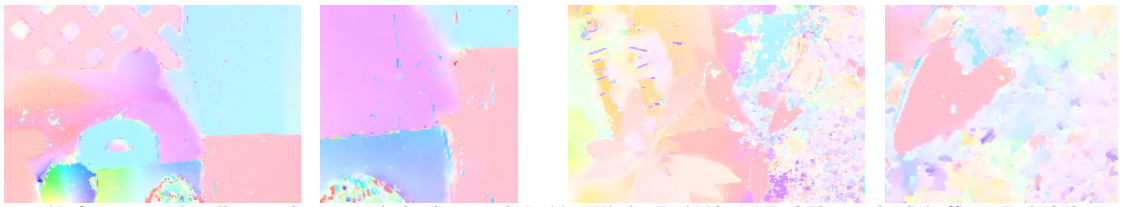
We also tried to fuse both the discrete and the discrete-continuous results with the known ground truth on "Rubber Whale" to determine the quality of our proposals. We found a mild energy reduction in the discrete case (from $E = 1821$ to 1756), but hardly any reduction in the discrete-continuous case (from $E = 1613$ to 1610). This shows that our proposals appear to be very appropriate, as even knowing the ground truth will not lower the energy much, and that the continuous improvement step substantially improves the results toward the ground truth (especially in smoothly varying areas). Furthermore, this also suggests that the proposed optimization gets very close to the ground truth (as much as is permitted by the model) and that further accuracy gains will require more involved models.

## 5. Conclusions

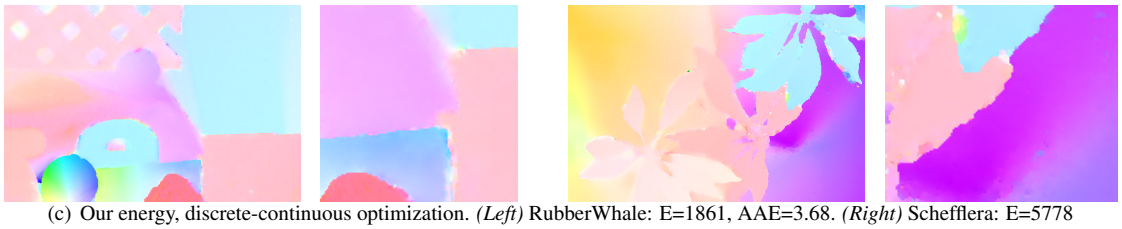
In this paper we proposed a new energy minimization approach for optical flow estimation called FusionFlow that combines the advantages of discrete and continuous optimization. The power of the optimization method allowed us to leverage a complex, highly non-convex energy formulation, which is very challenging for traditional continuous optimization methods. The proposed energy formulation was motivated by the statistics of optical flow, borrows from the stereo literature, and is robust to brightness changes, such as in shadow regions. Experimentally, at the moment of publication our approach is the top-performing method on the Middlebury optical flow benchmark across a variety of complex real-world scenes.



(a) Baseline energy, discrete-continuous optimization. *(Left)* RubberWhale: AAE=5.169. *(Right)* Schefflera



(b) Our energy, baseline continuous optimization. *(Left)* RubberWhale: E=4149, AAE=6.72. *(Right)* Schefflera: E=24853



(c) Our energy, discrete-continuous optimization. *(Left)* RubberWhale: E=1861, AAE=3.68. *(Right)* Schefflera: E=5778

Figure 5. Results for different energies and optimization methods. Each part shows the estimated flow field as well as a detail of the result.

Future work should consider whether more efficient proposal solutions can be developed that offer a similar diversity, but using many fewer proposals to reduce run-time. With very few proposals, very fast flow estimation might even be possible, at least without continuous refinement.

## References


[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *ICCV 2007* `http://vision.middlebury.edu/flow/`.

[2] S. Birchfield, B. Natarjan, and C. Tomasi. Correspondence as energy-based segmentation. *Image Vision Comp.*, 25(8):1329–1340, 2007.

[3] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, 1996.

[4] E. Boros, P. Hammer, and G. Tavares. Preprocessing of unconstrained quadratic binary optimization. Technical Report RUTCOR RRR 10-2006.

[5] E. Boros and P.L.Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.

[6] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, 26(9):1124–1137, 2004.

[7] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 23(11):1222–1239, 2001.

[8] T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. *ECCV 2006*, v. 1, pp. 471–483.

[9] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *IJCV*, 61(3):211–231, 2005.

[10] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *CVPR 2004*, v. 1, pp. 261–268.

[11] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. *IPMI 2007*.

[12] D. Greig, B. Porteous, and A. Seheult. Exact MAP estimation for binary images. *J. Roy. Stat. Soc. B*, 51(2):271–279, 1989.

[13] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *TPAMI*, 15(12):1217–1232, 1993.

[14] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, 1981.

[15] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *TPAMI*, 25(10):1333–1336, 2003.

[16] V. Kolmogorov and C. Rother. Minimizing non-submodular functions with graph cuts — A review. *TPAMI*, 29(7):1274–1279, 2006.

[17] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *TPAMI*, 24(2):147–159, 2004.

[18] J. Konrad and E. Dubois. Multigrid Bayesian estimation of image motion fields using stochastic relaxation. *ICCV 1988*, pp. 354–362.

[19] V. Lempitsky, C. Rother, and A. Blake. LogCut - Efficient graph cut optimization for Markov random fields. *ICCV 2007*.

[20] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, pp. 674–679, 1981.

[21] T. Meltzer, C. Yanover, and Y. Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. *ICCV 2005*, v. 1, pp. 428–435.

[22] É. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *IJCV*, 46(2):129–155, 2002.

[23] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *IJCV*, 67(2):141–158, 2006.

[24] C. E. Rasmussen. `minimize.m`. `http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/`, 2006.

[25] S. Roth and M. J. Black. On the spatial statistics of optical flow. *IJCV*, 74(1):33–50, 2007.

[26] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. *CVPR 2007*.

[27] A. Shekhovtsov, I. Kovtun, and V. Hlavac. Efficient MRF deformation model for non-rigid image matching. *CVPR 2007*.

[28] J. Sun, N.-N. Zhen, and H.-Y. Shum. Stereo matching using belief propagation. *TPAMI*, 25(7):787–800, 2003.