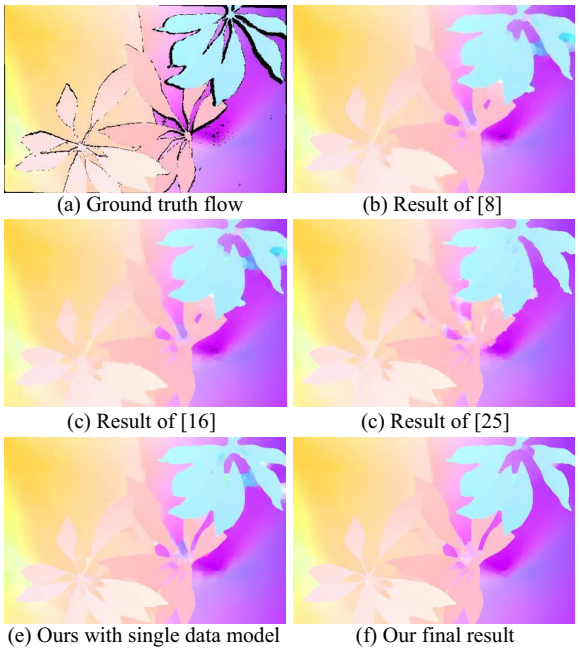
Optical Flow via Locally Adaptive Fusion of Complementary Data Costs

Tae Hyun Kim, Hee Seok Lee, and Kyoung Mu Lee
Department of ECE, ASRI, Seoul National University, 151-742, Seoul, Korea
{lliger9, ultra21, kyoungmu}@snu.ac.kr, http://cv.snu.ac.kr

Abstract

Many state-of-the-art optical flow estimation algorithms optimize the data and regularization terms to solve ill-posed problems. In this paper, in contrast to the conventional optical flow framework that uses a single or fixed data model, we study a novel framework that employs locally varying data term that adaptively combines different multiple types of data models. The locally adaptive data term greatly reduces the matching ambiguity due to the complementary nature of the multiple data models. The optimal number of complementary data models is learnt by minimizing the redundancy among them under the minimum description length constraint (MDL). From these chosen data models, a new optical flow estimation energy model is designed with the weighted sum of the multiple data models, and a convex optimization-based highly effective and practical solution that finds the optical flow, as well as the weights is proposed. Comparative experimental results on the Middlebury optical flow benchmark show that the proposed method using the complementary data models outperforms the state-of-the art methods.



(a) Ground truth flow (b) Result of [8] (c) Result of [16] (c) Result of [25] (e) Ours with single data model (f) Our final result

Figure 1. Our adaptive data fusion method consists of complementary data models that overcome the limitations of the single data model and provides the similar result to the ground truth flow.

1. Introduction

Optical flow estimation is used to find the pixel-wise displacement field between two images. It has been an active research topic in computer vision for decades. The estimation of accurate flow vectors has become a key step in numerous vision applications, such as dense 3D reconstruction and segmentations of video or motion [23, 18, 11]. However, optical flow estimation is difficult to solve because it is a highly ill-posed inverse imaging problem. To address this problem, traditional approaches used the energy minimization formulation composed of both data term and regularization as follows:

$$E = E_{data}(\mathbf{u}) + \lambda E_{reg}(\mathbf{u}), \tag{1}$$

where $\mathbf{u} = (u, v)^T$ denotes the optical flow field between two input images. E_{data} measures the data fidelity, E_{reg} enforces regularization of the flow field, and λ controls the tradeoff between the regularization and data fitting. Horn and Schunk [13] proposed the variational formulation for the first time in the optical flow estimation. Their model has been intensively investigated together with a coarse-to-fine strategy [1] in handling large motion between two images. However, their original model did not preserve the discontinuities in the displacement field and did not well reject the outliers in the data term, thus, various methods have been introduced to solve such problems. Black and Anandan [4] applied a robust statistic estimator, and Zach et al. [26] used an L1 data penalty term for the problems. However, the data terms were restricted to the brightness constancy in which the brightness of the corresponding pixels does not change, which is not applicable under illumination changes or noise.

Some recent studies have focused on improving the optical flow constraints beyond the brightness constancy. Brox et al. [6] averaged two data models with respect to the brightness constancy and gradient constancy, which as-

sumed that the intensity gradients of the corresponding points are same. Their model provided a robust estimator against illumination changes. Steinbrücker et al. [21] compared the classical data model based on the brightness constancy with arbitrary non-convex data models and showed the superiority of the complex data models, such as block matching. Werlberger et al. [24] also adopted the block matching and used truncated normalized cross correlation (NCC) as a pixel-wise data term to cope with the matching problems under illumination or exposure changes and to remove the ambiguity in the occluded regions.

Although many approaches have tried to improve the optical flow constraints and designed robust data models, the previously proposed data models still remained inadequate in practical situations because their inherent weaknesses. For example, the data models based on the gradient constancy or block matching such as the sum of absolute difference (SAD) are not valid when a geometrical transformation significantly changes the appearance in the target image. However, these models could complement each other, and thus, the problem of designing a locally adaptive data term is of the greatest importance.

The present study is related to the previous work of Xu et al. [25], where the selection of a data model provides low value in the total energy. Xu et al. selected the best data model for each pixel in the reference image from two data models which are the brightness and gradient constancy, and assumed that the weight variable is binary. They adopted the mean-field approximation for easy inference. However, their solution is intractable in some cases and it makes a drastic limitation to be a general data fusion model. Because of the small noise and uncertainty of each data model even if all assumptions of data models are not violated, the continuous weight variables must be allowed rather than the binary variables, to suppress the noise by averaging. We observed that the smoothness prior of the weight variables exhibited significant improvement, however, these additional cues rendered their solution infeasible.

Therefore, we propose a more general and unified variational framework that considers both the data fidelity and regularization to determine the locally varying continuous weight variables and thus, can be applicable to other vision problems. Figure 1 shows that our new optical flow estimation model based on the generalized data fusion framework significantly improves the accuracy. In addition, our model includes a novel data discriminability term, which defines the goodness of each data model to reduce the ambiguity in the homogeneous region. The proposed minimization procedure is very efficient and practical; thus, it can handle many data models, and the complexity increases linearly as the number of used data models increases.

We also provide a method to select the complementary data models to be used in the optical flow estimation. Although previous works [6, 27, 25] showed that more complementary models provide more accurate results, they are computationally inefficient. No study has been conducted on the data models whether they contain redundancy. Therefore, this study also proposes the method learning complementary data models where the number of selected models is made as few as possible based on the minimum description length (MDL) [12] and then fuses the chosen models.

Finally, experimental results verify our claims and show that the proposed approach outperforms the other conventional approaches and achieves very satisfactory performance in the Middlebury optical flow benchmark site.

2. Optical Flow Estimation Model Using Locally Adaptive Data Fusion

Most traditional optical flow estimation methods are based on the variational framework and it is easy to implement and parallelize on modern GPUs. Therefore, our proposed optical flow estimation method is also based on the variational framework with a robust data term and regularizer.

However, designing a robust single data model, that is reliable on the entire image domain, is almost impossible. Thus, designing a locally (pixel-wise) adaptive data term is desirable by the fusion of complementary data models while excluding the invalid data models. In addition, data discriminability which indicates the goodness of each data model to reduce the ambiguity in the textureless region and smoothness prior on the weight variable, are also required. Therefore, we can generalize (1) by employing the adaptive data fusion model as,

$$E = E_{data}(\mathbf{u}, \mathbf{w}) + \eta E_{dscr}(\mathbf{u}, \mathbf{w}) + \mu E_{reg}(\mathbf{w}) + \lambda E_{reg}(\mathbf{u}). \tag{2}$$

The set $\mathbf{w} = \{\mathbf{w}_l\}$ means a set of M weight variables, where $l = 1, 2, \ldots, M$. $E_{data}(\mathbf{u}, \mathbf{w})$ measures the data fidelity coupled with the flow fields and weight variables, $E_{dscr}(\mathbf{u}, \mathbf{w})$ measures the discriminability of each data model, and $E_{reg}(\mathbf{w})$ and $E_{reg}(\mathbf{u})$ enforces the regularization of the weights, respectively. The constant η is used to define the importance of discriminability term and μ controls the influence of regularization of the weight variables.

In the following sections, more details of the major factors for the adaptive data fusion, which are the data fidelity, the data discriminability, and regularization, are provided.

2.1. Data Fidelity

This study proposes an optical flow estimation model that combines the conventional but complementary optical flow constraints learnt by the method presented in Section 4.

One of the most important factors for the adaptive data fusion is data fidelity. For example, a data model with respect to the brightness constancy gives unreliable data cost at the true matching under illumination changes, shades or noise; thus, the energy minimization procedure can be over-fitted and can provide undesirable result to avoid high cost. However, other data models such as the gradient constancy or NCC can provide lower cost where the brightness constancy is invalid. In addition, a data model such as SAD can give lower cost when noise exists. Therefore, we can obtain the desired result and avoid over-fitting by favoring more reliable data models, which provide better data fidelity if given models are normalized to have similar costs in the true matching where the assumptions of the models are valid. Thus, we design E_{data} to minimize the weighted sum of the data models by

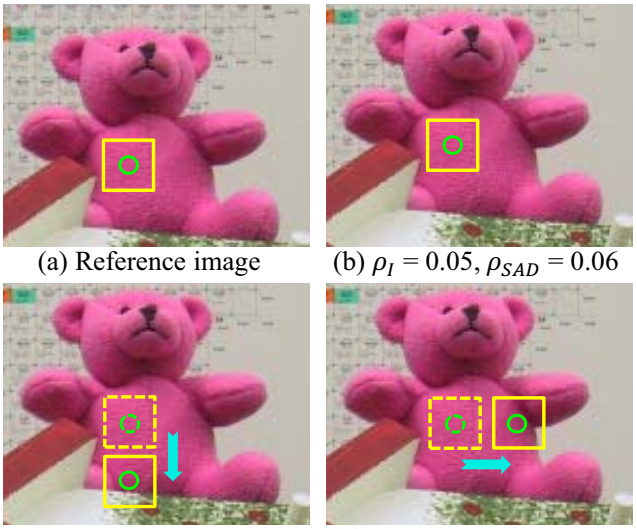
$$E_{data}(\mathbf{u}, \mathbf{w}) = \sum_{x} \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \rho_l(\mathbf{x}, \mathbf{u}), \tag{3}$$

where $\mathbf{x} \in \Re^2$ denotes the indices of the discrete locations in the image domain. The continuous weight variable, $\mathbf{w}_l \in \mathbf{w}$, has constraints, $\mathbf{w}_l(\mathbf{x}) \geq 0$ and $\sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) = 1$. This design allows to construct an integrated locally (pixel-wise) best suited data model.

2.2. Data Discriminability (Goodness)

In general, a data model giving a low cost value is preferred. However, favoring the model that gives the lowest cost may not be the best approach in some situations as shown in Figure 2. Two data models are considered at the corresponding points between the reference and target images. One model measures the SAD as ρ_{SAD}, whereas the other model measures the absolute difference of the brightness among the corresponding pixels as ρ_I. The corresponding point in the target image is shown in Figure 2(b), and both ρ_{SAD} and ρ_I are low because they are matched. However, if the matched point in the target image moves horizontally or vertically as shown in Figure 2(c)-(d), then ρ_{SAD} increases as the matched point moves, whereas ρ_I does not change. Thus ρ_I gives a low cost in a wide range and have high probability to generate many false positives.; However, ρ_{SAD} does not. Accordingly, ρ_{SAD} is a better model than ρ_I in this case where the region is homogeneous. Therefore, not only the data fidelity but also the data discriminability of the data model should be considered in our adaptive data fusion.

The notion of the discriminability term is similar to the concept of *good feature to track* in [20] and the Harris corner detector, which measures how reliable a point is under a given supporting region for feature matching. This constraint helps in choosing better data cues that enable unique and accurate feature localization in the matching.



(a) Reference image (b) $\rho_I = 0.05$, $\rho_{SAD} = 0.06$ (c) $\rho_I = 0.05$, $\rho_{SAD} = 0.21$ (d) $\rho_I = 0.06$, $\rho_{SAD} = 0.16$

Figure 2. Discriminability comparison of two different data models. (a) The box and circle centered at the same position in the reference image are used to measure ρ_{SAD} and ρ_I. (b) Corresponding point in the target image. Both ρ_{SAD} and ρ_I give reliable data costs. (c) and (d) Horizontally and vertically shifted corresponding point in the target image of (b). ρ_{SAD} changes considerably and gives a high data cost. ρ_I does not change and still yields a low cost.

Similar to the Harris corner detector, the discriminability of each data model can be measured by the smallest eigenvalue of the auto-correlation matrix corresponding to the auto-correlation function defined by

$$c_l(\mathbf{x}, \mathbf{u}_0) = \sum_{\mathbf{s} \in \mathcal{W}} (\rho_l(\mathbf{x}, \mathbf{u}_0) - \rho_l(\mathbf{x}, \mathbf{u}_0 + \mathbf{s}))^2, \tag{4}$$

where the given optical flow $\mathbf{u}_0$ can be obtained from the initial state of each level in the coarse to fine approach or from the previous result in the iterative optimization procedure. The function $c_l(\mathbf{x}, \mathbf{u}_0)$ is defined on the 5 x 5 window $\mathcal{W}$ centered at zero.

To obtain high discriminability, the smallest eigenvalue of the auto-correlation matrix of $c_l(\mathbf{x}, \mathbf{u}_0)$ should be large, and we define it as

$$e_l(\mathbf{x}, \mathbf{u}_0) = \min(|\nu_1|, |\nu_2|), \tag{5}$$

where ν_1 and ν_2 denote the two eigen values of the local auto-correlation matrix. Then the data discriminability term in (2) can be represented as

$$E_{dscr}(\mathbf{x}, \mathbf{u}) = \sum_{\mathbf{x}} \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \sum_{k=1, k \neq l}^{M} e_k(\mathbf{x}, \mathbf{u}_0). \tag{6}$$

Because the data model with a large discriminability yields a large $e_l(\mathbf{x}, \mathbf{u}_0)$, this is added to the costs of the other data models and prevents the other models from gaining more weights.

2.3. Regularization

As optical flow estimation is a highly ill-posed problem, regularization enforcing the smoothness of variables is necessary to obtain a reliable solution. In our energy model, two primal variables are the set of the weight variables **w** and the flow fields **u**, and the details of regularization for each variable are described in the following sections.

2.3.1 Regularization on w

Regularizing the weight variables is also an important factor for adaptive data fusion. In particular, if noise exists in the true matching, then favoring a model which gives the better data fidelity only may not be the best solution, but averaging or weighted summing of all the data models could give more reliable results. In addition, the abnormally low costs compared with those of the neighboring pixels from the same data model are likely false. Thus, regularization of the weight variables is necessary. Therefore, we allow the continuous, but not the binary, weight field to get a solution from the weighted average of the costs, and incorporate the smoothness prior on the weight variables to avoid assigning large weight to unreliable models. We design the regularization of the weight variables to change smoothly but to have sparse discontinuities. This process yields

$$E_{reg}(\mathbf{w}) = \sum_{\mathbf{x}} \sum_{l=1}^{M} |\nabla \mathbf{w}_l|. \tag{7}$$

2.3.2 Regularization on u

In general, conventional optical flow estimation models assume that the flow vectors vary smoothly and have the sparse discontinuities in the edges of reference image. Therefore, the edge map [18, 25, 23] is coupled to the total variation regularizer which allows discontinuities in the flow fields. The edge map of a colored reference image is given by the maximal color difference among neighboring pixels as

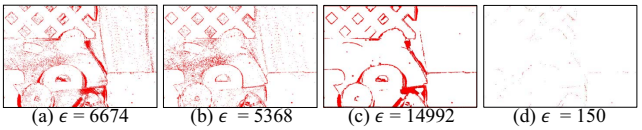
$$g(\mathbf{x}) = \exp(-\max(|\nabla I_R|,\ |\nabla I_G|,\ |\nabla I_B|)^{\kappa}), \tag{8}$$

where κ controls the magnitude of the difference between the homogeneous region and the edge and ∇I_R, ∇I_G and ∇I_B denote the pixel-wise derivatives of the RGB color channels, respectively. The regularization of the motion fields is formulated by

$$E_{reg}(\mathbf{u}) = \sum_{\mathbf{x}} g(\mathbf{x}) \cdot |\nabla \mathbf{u}|. \tag{9}$$

3. Learning of Complementary Data Models Based on MDL

Using many complementary data models could give better results, however, this process is inefficient because of the redundancy. Therefore, we study the nature of the data models to be used in the data fusion on the Middlebury training datasets where the ground truth of the motion fields is known.



(a) ϵ = 6674 (b) ϵ = 5368 (c) ϵ = 14992 (d) ϵ = 150

Figure 3. Sum of error, ϵ, on the RubberWhale sequence and the pixel whose data cost is high becomes red. (a) Single model. (brightness constancy) (b) Single model. (gradient constancy) (c) Single model. (5x5 SAD) (d) All 31 models.

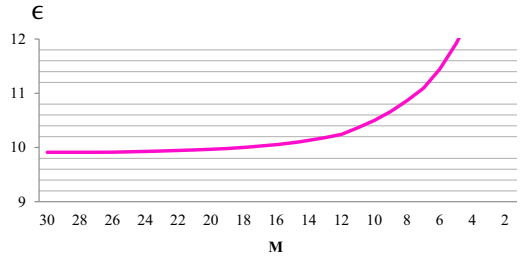


Figure 4. Trade-off curve between M and ϵ of the chosen M models. 31 models are used as candidates.

Prior to this study, we assume that the distribution of the data costs in the true matching is Gaussian and thus all $\rho_l(.)$ are normalized to have zero mean and unit variance. This setting is used in all experiments in our study. By normalization, fair comparison of the costs among the data models in the true matching is possible, and we can presume that the matching cost is unreliable when the cost is over one ($\sigma = 1$), otherwise it is reliable. Thus, we define the error function of the data model by

$$f_l(\mathbf{x}) = \begin{cases} C, & \text{if } \rho_l(\mathbf{x}, \mathbf{u}_{gt}) > 1 \\ 0, & \text{otherwise.} \end{cases}, \tag{10}$$

where $\mathbf{u}_{gt}$ denotes the ground truth motion and C is a positive constant. In addition, we compose L candidate data models which have been used in conventional methods for comparison. The candidate data models consists of total 31 models such as the brightness, gradient and block matching for the gray and RGB color channels. In addition, different sizes of blocks are used for block matching. (See the supplementary material for more details.)

For a set of data models to be complementary, at least one data model should give reliable cost where the others could not. Therefore we should minimize the sum of error function with complementary data models, and it is given by

$$\epsilon = \sum_{\mathbf{x}} \sum_{l=1}^{L} q_l(\mathbf{x}) f_l(\mathbf{x}), \tag{11}$$

where $q_l(\mathbf{x}) \in \{0, 1\}$ denotes the pixel-wise integer indicator function and $\sum_{l=1}^{L} q_l(\mathbf{x}) = 1$. As illustrated in Figure 3(a)-(c), ϵ of a single data model could be quite high,

however, with the aid of complementary data models, ϵ is reduced significantly as shown in Figure 3(d). So, the aim of the current study is to minimize ϵ with fewer data models by removing redundancy.

To use as few models as possible in literally describing data while minimizing redundancy, we employe the MDL concept [12] for the formulation of model selection as

$$F = \sum_{\mathbf{X}} \sum_{l=1}^{L} q_l(\mathbf{X}) f_l(\mathbf{X}) + \gamma \delta_l, \tag{12}$$

where $\mathbf{X}$ denotes the indices of the discrete locations in the entire image domain of the training datasets and the indicator function δ_l is defined as,

$$\delta_l = \begin{cases} 1, \text{ if } \sum_{\mathbf{X}} q_l(\mathbf{X}) \geq 1 \\ 0, \text{ otherwise.} \end{cases},$$

and thus, $\sum_{l=1}^{L} \delta_l$ indicates the number $M (\leq L)$ of chosen models in (12) and γ controls the strength of its importance. The first term in (12) is designed to choose complementary set of the data models and the second term is used to minimize the redundancy of the chosen data models based on MDL.

If γ is given, the function F can be minimized by [9], and our M data models to be used in the optical flow estimation can be learned. A trade-off between M and ϵ with chosen models is allowed, thus, we can determine the preferred set of data models from the curve shown in Figure 4. In our experiments, we use eight($M = 8$) models learned from this curve which gives about 10% higher error compared with that of using full 31 models. The learned eight models are complementary as expected. To be robust against geometrical changes, the three data models of the red, blue and gray channels are based on the brightness constancy in [26]. In addition, to be robust against illumination changes, the four models of the green and blue channels are based on the vertical and horizontal gradient constancy in [25]. Finally, the 5x5 SAD used in [21] for the green channel is selected, which is robust against noise. The finally chosen eight models are listed in Table 1.

Data model:	$\lvert I_r(\mathbf{x}) - I_t(\mathbf{x}+\mathbf{u})\rvert$	$\lvert \nabla I_r(\mathbf{x}) - \nabla I_t(\mathbf{x}+\mathbf{u})\rvert$	5x5 SAD
Type	Brightness constancy	Gradient constancy	Block matching
Channel	R, B, Gray	G, B(2 vertical, 2 horizontal)	G

Table 1. Eight chosen data models. I_r and I_t denote the reference and target image, respectively.

4. Optimization

The proposed optical flow estimation model introduced in the previous section includes the regularization, weighted sum of the multiple data models and the data discriminability, and the final objective function of this study is as follows:

$$\min_{\mathbf{u},\mathbf{w}} \sum_{\mathbf{x}} \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \rho_l(\mathbf{x}, \mathbf{u}) + \eta \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \sum_{k=1, k\neq l}^{M} e_k(\mathbf{x}, \mathbf{u}_0) + \mu \sum_{l=1}^{M} |\nabla \mathbf{w}_l| + \lambda \cdot g(\mathbf{x}) \cdot |\nabla \mathbf{u}|, \tag{13}$$

where $\mathbf{w}_l(\mathbf{x}) \geq 0$, and $\sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) = 1$.

Because regularization and discriminability terms are convex, if all the data models are convex then (13) can be a jointly convex problem [10]. However, the data models are generally not convex, and thus, the proposed model is also not convex and it suffers from some computational difficulties. To address this problem, quadratic relaxation method is widely used [2, 26, 21]. Our minimization is also based on this technique, and we decouple the convex and non-convex parts by introducing an auxiliary variable $\mathbf{v}$ linked to $\mathbf{u}$ as,

$$\min_{\mathbf{u},\mathbf{v},\mathbf{w}} \sum_{\mathbf{x}} \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \rho_l(\mathbf{x}, \mathbf{v}) + \eta \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \sum_{k=1, k\neq l}^{M} e_k(\mathbf{x}, \mathbf{u}_0) + \frac{(\mathbf{u}-\mathbf{v})^2}{2\theta} + \mu \sum_{l=1}^{M} |\nabla \mathbf{w}_l| + \lambda \cdot g(\mathbf{x}) \cdot |\nabla \mathbf{u}|. \tag{14}$$

If θ is set to a very small number, then the minimization of (14) is close to (13). Using decomposition, the function of $\mathbf{u}$ and $\mathbf{w}$ becomes convex with respect to fixed $\mathbf{v}$. Moreover $\mathbf{v}$ can be minimized globally with respect to fixed $\mathbf{u}$ and $\mathbf{w}$ with a complete search for all pixels. Therefore this minimization problem can be optimized by alternating the steps.

4.1. Continuous Optimization of u and w

For $\mathbf{v}$ to be fixed, using total variations makes the function non-differentiable. However, with the aid of convexity, the duality principle is applied in the minimization of $\mathbf{u}$ and $\mathbf{w}$. Thus, to solve (14) when $\mathbf{v}$ is fixed, we adopt the first-order primal dual algorithm [7] which is known to be fast with optimal convergent rate. The primal dual update step

is as follows.

$$\begin{cases} \mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \sigma \mathbf{A}\mathbf{K}\mathbf{u}^n}{\max(\mathbf{1}, \mathbf{p}^n + \sigma \mathbf{A}\mathbf{K}\mathbf{u}^n)} \\ \mathbf{u}^{n+1} = \frac{\lambda\theta(\mathbf{u}^n - \tau(\mathbf{A}\mathbf{K})^T \mathbf{p}^{n+1}) + \tau \mathbf{v}}{\lambda\theta + \tau} \\ \mathbf{r}_l^{n+1} = \frac{\mathbf{r}_l^n + \sigma \mathbf{K}\mathbf{w}_l^n}{\max(\mathbf{1}, \mathbf{r}_l^n + \sigma \mathbf{K}\mathbf{w}_l^n)} \\ \mathbf{w}_l^{n+1}(\mathbf{x}) = \mathbf{w}_l^n(\mathbf{x}) - \tau(\mathbf{K}^T \mathbf{r}_l^{n+1})(\mathbf{x}) \\ \qquad\qquad - \tau \dfrac{\rho_l(\mathbf{x}, \mathbf{v}) + \eta \sum_{k=1, k\neq l}^{M} e_k(\mathbf{x}, \mathbf{u}_0)}{\lambda} \\ \mathbf{w}^{n+1} = \Pi_{\mathbf{w}}(\mathbf{w}^{n+1}), \end{cases} \tag{15}$$

where $n \geq 0$ indicates the iteration number, and $\mathbf{p}$ and $\mathbf{r}_l$ denote the dual variables on the vector space. The constant update steps, σ and τ control the convergence rate, as defined in [7]. $\mathbf{K}$ is a continuous linear operator that calculates the difference between the neighboring pixels, and the diagonal matrix $\mathbf{A}$ is the weighting matrix denoted as $\mathbf{A} = \text{diag}(g(\mathbf{x}))$. Because $\mathbf{w}$ has some constraints, the orthogonal projection $\Pi_{\mathbf{w}}$ projects $\mathbf{w}$ onto a unit simplex [17]. This projection converges with M iterations at most, and the detail is shown in Algorithm 1. (See supplementary material for more details on the minimization of $\mathbf{u}$ and $\mathbf{w}$.)

Algorithm 1 Algorithm of projection onto a unit simplex

1: $T = \{1, ..., M\}$
2: $\mathbf{w}_l(\mathbf{x}) \leftarrow \mathbf{w}_l(\mathbf{x}) - (\sum_l \mathbf{w}_l(\mathbf{x}) - 1)/|T|$, if $l \in T$
3: $T \leftarrow T - \{l\}$, if $\mathbf{w}_l(\mathbf{x}) < 0$
4: $\mathbf{w}_l(\mathbf{x}) \leftarrow 0$, if $l \notin T$
5: Repeat steps 2-4 until $\sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) = 1$, for all $\mathbf{x}$

4.2. Continuous Optimization of v

For arbitrary and non-convex data terms, $\mathbf{v}$ can be optimized globally by performing a complete search when both $\mathbf{u}$ and $\mathbf{w}$ are fixed. It allows the computation of the large displacement optical flow [21] and it also makes possible to avoid staying in the local minima. However, the complete search is slow. Therefore, we adopt an efficient warping scheme that allows the arbitrary data term using the second-order Taylor expansion [24]. Our data models are approximated as follows:

$$\begin{aligned} \rho_l(\mathbf{x}, \mathbf{u}) &\approx \rho_l(\mathbf{x}, \mathbf{u}_0) + \nabla \rho_l(\mathbf{x}, \mathbf{u}_0)^T (\mathbf{u} - \mathbf{u}_0) \\ &+ \frac{1}{2} (\mathbf{u} - \mathbf{u}_0)^T \nabla^2 \rho_l(\mathbf{x}, \mathbf{u}_0) (\mathbf{u} - \mathbf{u}_0), \end{aligned} \tag{16}$$

where $\nabla \rho_l(\mathbf{x}, \mathbf{u}_0)$ is the first-order derivative and $\nabla^2 \rho_l(\mathbf{x}, \mathbf{u}_0)$ is a diagonal matrix whose entries are only positive second-order derivatives of $\rho_l(\mathbf{x}, \mathbf{u})$ at $(\mathbf{x} + \mathbf{u}_0)$. By eliminating the non-diagonal and the negative diagonal entries, we can ensure that the Hessian matrix is positive semi-definite and guarantee that the approximated function is convex near $(\mathbf{x} + \mathbf{u}_0)$, which results in,

$$\begin{aligned} \mathbf{v} &= \arg\min_{\mathbf{v}} \frac{(\mathbf{u} - \mathbf{v})^2}{2\theta} + \sum_{l=1}^{M} \mathbf{w}_l(\mathbf{x}) \cdot \rho_l(\mathbf{x}, \mathbf{v}) \\ &\approx \frac{\mathbf{u} - \theta \sum_{l=1}^{M} w_l (\nabla \rho_l(\mathbf{x}, \mathbf{u}_0) - \mathbf{u}_0^T \nabla^2 \rho_l(\mathbf{x}, \mathbf{u}_0))}{1 + \theta \sum_{l=1}^{L} w_l \nabla^2 \rho_l(\mathbf{x}, \mathbf{u}_0)}. \end{aligned} \tag{17}$$

4.3. Occlusion Detection and Postprocessing

The data models used in our optical flow estimation have weakness in occlusion, thus occlusion handling in the post-processing is beneficial. Generally, cross checking of the optical flows is effective however, it doubles the computational cost [19]. Checking the pixels that violate the mapping uniqueness constraint [25, 15, 5] is another method for occlusion detection and the state of the occlusion variable in the current study is defined as follows:

$$o(\mathbf{x}) = \min(\frac{\max(N(\mathbf{x} + \mathbf{u}) - 1), 0)}{2}, 1), \tag{18}$$

where $N(\mathbf{x} + \mathbf{u})$ denotes the number of pixels in the reference image that corresponds to a pixel located at $(\mathbf{x} + \mathbf{u})$ in the target image. The state $o(\mathbf{x})$ contains one of the three values $\{0, 0.5, 1\}$. Because the estimated optical flow could have some errors and a rounding off technique is used to count the number on the discrete grid, we regard the pixels as ambiguous when $N(\mathbf{x}+\mathbf{u}) = 2$ and $o(\mathbf{x})$ is equal to 0.5 in this case. This occlusion handling method violates the mapping uniqueness constraint, however, it is quite useful and performs well in practical situations. To fill the occluded pixels and remove the artifacts in the homogeneous regions, we apply the joint bilateral filter with the occlusion states and the similarity of color and optical flow fields that follows [19].

4.4. Implementation

Algorithm 2 Overall procedure of the proposed optical flow estimation algorithm

Input: Two color images I_r and I_t
Output: Continuous fields $\mathbf{u}, \mathbf{w}$
1: Build pyramids for both reference and target images.
2: Set initial values of continuous primal and dual variables, $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{p}, \mathbf{r}$ in the coarse level.
3: **for** $n = 1$ to 100 **do**
4: Continuous minimization of $\mathbf{u}$ and $\mathbf{w}$ (Sec. 4.1)
5: Continuous minimization of $\mathbf{v}$ (Sec. 4.2)
6: **end for**
7: Occlusion detection and postprocessing (Sec. 4.3)
8: Propagate variables to the next pyramid level if exists
9: Repeat steps 3-8 from coarse to fine pyramid level

The proposed optical flow estimation model is based on the quadratic approximations of the original data model to

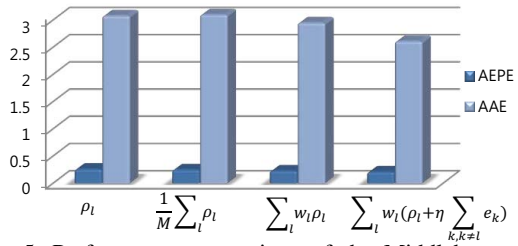


Figure 5. Performance comparison of the Middlebury training datasets. The comparisons are made using the methods with single data term, which has the best score, mean of data models, and the proposed data fusion with and without the discriminability term.

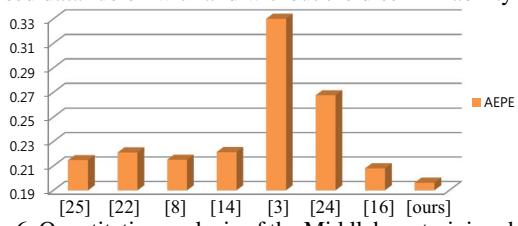


Figure 6. Quantitative analysis of the Middlebury training datasets. We compare our method with the state-of-the-art methods based on the variational framework.

ensure convexity where $\mathbf{u} \approx \mathbf{u}_0$. Initially these approximations are only valid for small displacements, and thus the proposed model is embedded into the coarse-to-fine strategy to cope with the large displacements. Furthermore we estimate five affine motions from the flow fields similar to RANSAC, and update the initial flow field when the affine motion yields lower energy, $E_{data}(\mathbf{u}, \mathbf{w})$, than the propagated motion field from the coarser level. To obtain accurate results, we use a scale factor of 0.9 to build the image pyramids. Three image warps that use intermediate flow vectors are performed in a single pyramid level as in [22]. The overall procedure of the proposed method is summarized in Algorithm 2.

The parameters are determined empirically, $\lambda = 0.25$, $\theta = 0.1$, $\eta = 0.02$, and $\mu = 1$ from the curve in Figure 7. Furthermore the proposed optical flow estimation model is implemented in C++ on the GPU with CUDA, and the computation time is significantly reduced using parallelization. It takes about 150 seconds for Urban2 sequence on the GPU.

5. Experimental Results

The end point error (EPE) and angular error (AE) of the flow estimation results are measured using the various data models and shown in Figure 5. AEPE and AAE denote the averaged errors of the entire Middlebury training datasets. In the evaluation, the single data model giving the best result, simple mean of data models and their fusion by the proposed method with and without the data discriminability are compared. The results suggest that the proposed fusion method with discriminability outperforms others and provides significantly better results.

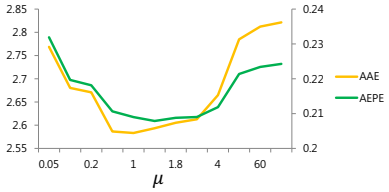


Figure 7. Precision change by varying μ.

Figure 7 shows the plots of AEPE and AAE by varying μ. The curve drops rapidly and begins to stabilize. This result shows that relying only on the regularization of $\mathbf{u}$ does not give a good accuracy compared with taking into account the smoothness of the weight variables $\mathbf{w}$. Figure 8 shows the pixel-wise weights, and the weights shown in Figures 8(a)-(c) are related with the gradient constancy, and the weight shown in Figure 8(d) is obtained from the brightness constancy and that in Figure 8(e) represents the weight of SAD. The arrows indicate the shaded regions in the reference image, and the data models related with the gradient constancy gain more weights than the other models, as expected, and the data model shown in Figure 8(c) has less weight where the data cost is high in Figure 3(a).

Our results are compared with the state-of-the-art optical flow estimation methods based on the variational framework [25, 22, 14, 3, 8, 16, 24], and the AEPE of each method is shown in Figure 6. The proposed approach outperforms the other methods in the Middlebury training datasets. Also, the results of the Middlebury test sets are shown in Figure 9. Our proposed method ranks the first among published papers in the AEPE and second in the AAE at the time of submission. The results are also available in the evaluation site [1].

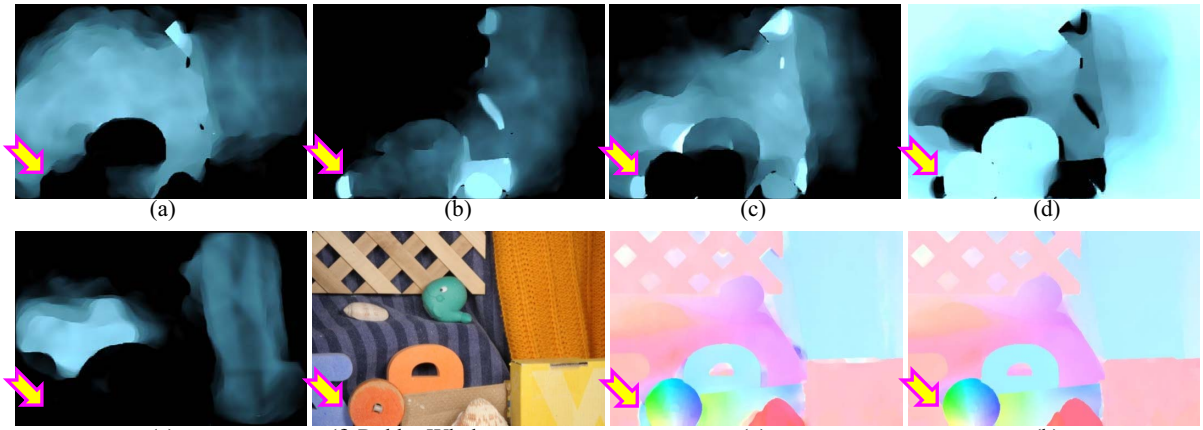
6. Discussion and Conclusion

This study has presented a novel optical flow estimation method that fuses various data models. By providing the locally adaptive data term, the limitation of a single data model can be overcome. This study also provided an efficient and practical solution, and the proposed optical flow estimation model showed competitive results compared with the state-of-the art methods. In addition, a method for learning the set of data models based on MDL was presented which provided the data models to be used in our optical flow estimation. The proposed model can be generalized and applied to other problems that requires the incorporation of numerous data models. It can resolve difficulties in designing a robust data function.

Acknowledgments

This research was supported in part by the Forensic Research Program of the National Forensic Service (NFS), Ministry of Security and Public Administration, Korea.

[1] http://vision.middlebury.edu/flow/eval/



(a) (b) (c) (d)

(e) (f) RubberWhale sequence (g) (h)

Figure 8. (a)-(e) Five weight maps of the eight chosen models. The maps of three models that have lesser weights are omitted. (g) Flow from the single data model with brightness constancy. (h) Flow from our final result.

Average endpoint error	avg. rank	Army (Hidden texture) GT im0 im1			Mequon (Hidden texture) GT im0 im1			Schefflera (Hidden texture) GT im0 im1			Wooden (Hidden texture) GT im0 im1			Grove (Synthetic) GT im0 im1			Urban (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Teddy (Stereo) GT im0 im1		
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
ComplexFlow [80]	2.6	**0.07** 1	0.20 2	**0.05** 1	**0.15** 1	0.51 3	0.12 5	**0.18** 1	**0.37** 1	**0.14** 1	0.10 2	0.49 3	0.06 2	**0.41** 1	**0.61** 1	0.21 2	0.23 2	0.66 2	**0.19** 1	0.10 4	0.12 8	0.17 11	**0.34** 1	0.80 4	0.23 2
ours	6.6	0.08 7	0.21 3	0.06 5	0.16 5	0.53 4	0.12 5	0.19 2	**0.37** 1	**0.14** 1	0.14 7	0.77 21	0.07 4	0.51 4	0.78 5	0.25 3	0.31 4	0.76 3	0.25 6	0.11 10	0.12 8	0.21 29	0.42 7	0.78 2	0.63 12
MDP-Flow2 [69]	7.5	0.08 7	0.21 3	0.07 14	**0.15** 1	**0.48** 1	**0.11** 1	0.20 4	0.40 4	**0.14** 1	0.15 17	0.80 26	0.08 11	0.63 13	0.93 14	0.43 14	0.26 3	0.76 3	0.23 5	0.11 10	0.12 8	0.17 11	0.38 3	0.79 3	0.44 4

Figure 9. EPE of the Middlebury benchmark at the time of submission. Three top-performing results are listed, and our method is the second in terms of EPE.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal on Computer Vision*, 2, 1989. 1

[2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition–modeling, algorithms, and parameter selection. *Int. J. Comput. Vision*, 67(1):111–136, Apr. 2006. 5

[3] A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 97, 2011. 7

[4] M. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proceedings of IEEE International Conference on Computer Vision*, 1993. 1

[5] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 2003. 6

[6] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of European Conference on Computer Vision*, 2004. 1, 2

[7] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40, 2011. 5, 6

[8] Z. Chen, J. Wang, and Y. Wu. Decomposing and regularizing sparse/non-sparse components for motion field estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 7

[9] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 96, 2012. 5

[10] M. Ehrgott. *Multicriteria optimization*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 2000. 5

[11] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 1

[12] P. D. Grünwald. *The Minimum Description Length Principle (Adaptive Computation and Machine Learning)*. The MIT Press, 2007. 2, 5

[13] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17, 1981. 1

[14] K. Jia, X. Wang, and X. Tang. Optical flow estimation using learned sparse model. In *Proceedings of IEEE International Conference on Computer Vision*, 2011. 7

[15] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. In *Proceedings of IEEE International Conference on Computer Vision*, 2001. 6

[16] P. Krähenbühl and V. Koltun. Efficient nonlocal regularization for optical flow. In *Proceedings of the 12th European conference on Computer Vision - Volume Part I*, ECCV'12, pages 356–369, Berlin, Heidelberg, 2012. Springer-Verlag. 7

[17] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of α^n. *Journal of Optimization Theory and Applications*, 50, 1986. 6

[18] R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 1, 4

[19] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011. 6

[20] J. Shi and C. Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994. 3

[21] F. Steinbrücker, T. Pock, and D. Cremers. Advanced data terms for variational optic flow estimation. In *Proceedings of International Workshop on Vision, Modeling, and Visualization*, 2009. 2, 5, 6

[22] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 7

[23] M. Unger, M. Werlberger, T. Pcok, and H. Bischof. Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 1, 4

[24] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2, 6, 7

[25] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2012. 2, 4, 5, 6, 7

[26] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Proceedings of DAGM conference on Pattern recognition*, 2007. 1, 5

[27] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In *In EMMCVPR*, 2009. 2