

A Segmentation Based Variational Model for Accurate Optical Flow Estimation

Li Xu, Jianing Chen, and Jiaya Jia

Department of Computer Science and Engineering
The Chinese University of Hong Kong
{xuli, jnchen, leojia}@cse.cuhk.edu.hk

Abstract. Segmentation has gained in popularity in stereo matching. However, it is not trivial to incorporate it in optical flow estimation due to the possible non-rigid motion problem. In this paper, we describe a new optical flow scheme containing three phases. First, we partition the input images and integrate the segmentation information into a variational model where each of the segments is constrained by an affine motion. Then the errors brought in by segmentation are measured and stored in a confidence map. The final flow estimation is achieved through a global optimization phase that minimizes an energy function incorporating the confidence map. Extensive experiments show that the proposed method not only produces quantitatively accurate optical flow estimates but also preserves sharp motion boundaries, which makes the optical flow result usable in a number of computer vision applications, such as image/video segmentation and editing.

1 Introduction

Accurate motion estimation is required for solving many computer vision problems, including moving object segmentation and video understanding. However, high-quality motion estimates are usually difficult to be obtained, especially for occluded pixels, discontinuous motion boundaries, and textureless regions.

In stereo matching [1–3], color-segmentation-based approaches have demonstrated their strong capability in handling textureless and occluded regions. These methods generally assume a specific (e.g. planar) model for each segment. Regularization is then applied to the model parameters. However, similarly employing the segmentation in modern optical flow frameworks is not easy, owing to the insufficiency of using the segmented regions to constrain the non-rigid motion in consecutive frames. Small-size segments were used in [4] to alleviate this problem. But it usually suffers from the following limitations. For one thing, small segments cannot faithfully represent the structure of natural scenes and thus weaken the regularization power brought forth by segmentation. For another, small patch size might result in poor estimation of the motion parameters due to the local aperture problem.

In this paper, we address two important issues of using segmentation in optical flow - that is, 1) how to know whether the motion model fits the flow in

a segment or not, and 2) how to handle the non-rigid motion while faithfully preserving the image structures. To this end, a three-step optical flow framework is proposed. In particular, we first segment the input images based on the color information and the initial motion estimate. The segmentation is then incorporated into a variational model to estimate the motion parameters combining the regularization terms. To reduce the possible segmentation errors caused by inappropriate motion parameterization for non-rigid objects, we compute a confidence map that represents the likelihood whether the parametric flow estimate for a specific pixel is trustworthy or not. This confidence map is taken into a final global optimization step to selectively and locally refine the problematic flow estimates. Our experimental results show that the proposed method is capable of handling both the rigid and non-rigid motions.

2 Related Work

Optical flow is a long studied problem [5–7]. Following the framework of Horn and Schunck [6], efforts have been recently put in improving the accuracy and efficiency using a variational model [8–12]. The main issue yet to be addressed is the recovery of high quality motion boundary in the presence of large occlusion.

There exist ways to alleviate the boundary problem. Black and Anandan [7] applied a robust function to handling the possible outliers and the motion discontinuity. This function is also employed in other methods [8, 10] and is solved as a modified L-1 norm minimization problem.

Anisotropic diffusion is a method using the gradient information to reduce over-smoothing in the processed region. Tschumperlé *et al.* [13] proposed a matrix-valued scheme for nonlinear diffusion in estimating the matrix-form motion tensor. The adaptive diffusion function returns small values at object boundary, which controls the smoothness over the motion discontinuity. Xiao *et al.* [11] extended the work by substituting the diffusing tensor with an adaptive bilateral filter, and controlled the diffusion process according to the occlusion detection. Although these methods can sharpen the motion boundary, they do not handle well large occlusions, possibly making the recovered motion boundary over-smoothed.

Segmentation- or layer-based approaches assume a parametric motion model (e.g., translational or affine model) for each segment. To handle non-rigid motion, the size of segments has to be small. Zitnick *et al.* [4] generated consistent segments between frames and enforced a translational model within each segment. Piece-wise parametric motion model is also used in [5, 7, 14] within small patches. This assumption may result in poor estimation of the motion parameters because of the local aperture problem. Black and Jepson [15] relaxed the affine constraint by adding local deformation to the parametric model, resulting in a non-parametric motion. Mémin and Pérez [16] combined the piece-wise parametric motion with local disturbance in a hierarchical setting to mix local flow field with different parameterizations. For these methods, as the model is no longer piece-wise parametric, the regularization power is weakened. The

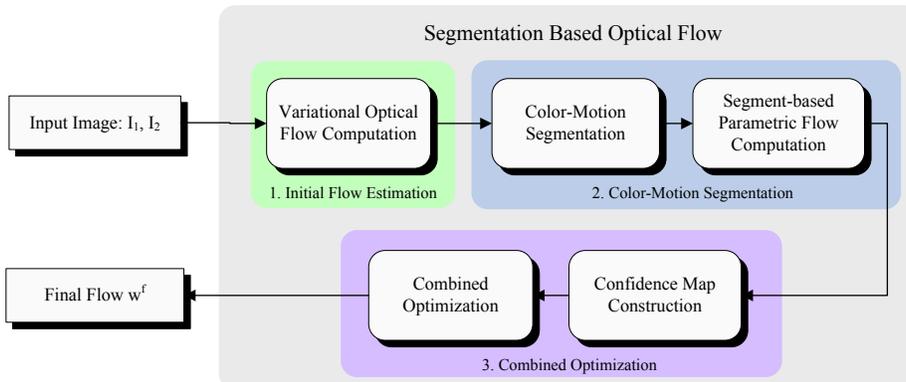


Fig. 1. Overview of our algorithm.

underlying difficulties of the segmentation-based methods are the handling of non-rigid motion and the robust estimation of parameters with the presence of occlusion. These difficulties hinder the widely studied color segmentation from being trivially employed in optical flow estimation.

Another topic related to optical flow is the motion segmentation which aims at extracting moving objects. The segmentation is usually accomplished based on motion [17, 18] or the combination of color and motion [19–22]. Joint estimation of motion and segmentation has recently been studied [17, 18, 21], where the contour evolving and motion estimation are iteratively performed. The motion segmentation methods cannot be directly applied to flow estimation since extracting a moving object does not need to accurately estimate flow for each and every pixel.

3 Our Approach

Given an image pair (I_1, I_2) that contains objects undergoing small spatially-variant, and possibly non-rigid motion, our objective is to estimate the motion flow vector $\mathbf{w}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}))^T$ for each pixel $\mathbf{x} = (x, y)^T$ in image I_1 . Our approach consists of three main steps. We illustrate the block diagram in Fig. 1. Briefly speaking, we first estimate an initial motion field based on a simple variational model. The initialized flow is then combined with the color information to generate segments. The flow field is refined by segmentation using a robust parameter estimation process. In order to handle the non-rigid motion, a confidence map is constructed measuring the confidence of using the motion parameterization in each segment. The final flow is obtained by relaxing the motion parameterization constraint in a global optimization step using the confidence map.

To visualize the dense flow, in this paper, we adopt the color coding in [23] where the chromaticity is used to distinguish the motion direction and the in-

tensity corresponds to the motion magnitude. Fig. 2 shows the reference color wheel and our *Yosemite* result represented by both flow arrows and colors.

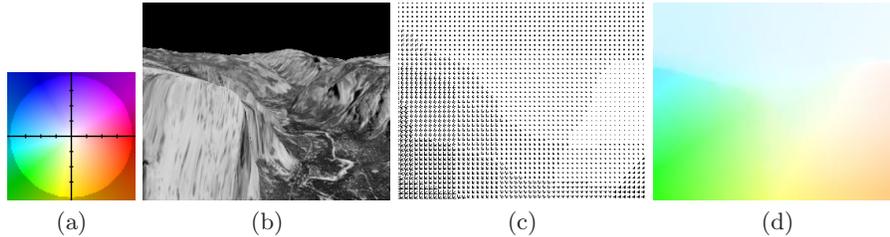


Fig. 2. Flow representation. (a) A reference color wheel. (b) One input image of the *Yosemite* sequence. (c) Traditional flow field representation using arrows. (d) Optical flow field with the color-coded representation.

3.1 Initial Flow

In this stage, we enforce the color constancy and use the variational model similar to that in [8, 10] to initialize the flow field. The data term is expressed as

$$E_{Data}(u, v) = \int_{\Omega_1} \sum_{c=1}^3 \Psi(|I_2(\mathbf{x} + \mathbf{w}, c) - I_1(\mathbf{x}, c)|^2, \epsilon_D) d\mathbf{x}, \quad (1)$$

where Ω_1 is the domain of image I_1 , $\Psi(x, \epsilon)$ is the Total Variation (TV) regularizer [8] defined as $\Psi(x, \epsilon) = \sqrt{x + \epsilon^2}$, and $I(\mathbf{x}, c)$ denotes the color of pixel \mathbf{x} in the c th channel of image I . The smoothness term is given by

$$E_{Smooth}(u, v) = \int_{\Omega_1} \Psi(\|\nabla u\|^2 + \|\nabla v\|^2, \epsilon_S) d\mathbf{x}, \quad (2)$$

where ∇ is the first-order derivative operator. In the rest of this paper, we denote $\Psi_D(x) = \Psi(x, \epsilon_D)$ and $\Psi_S(x) = \Psi(x, \epsilon_S)$ for simplicity's sake. The initial flow is estimated by minimizing the combined energy

$$E_0(u, v) = E_{Data}(u, v) + \alpha E_{Smooth}(u, v), \quad (3)$$

where α is a weight balancing the two terms. $E_0(u, v)$ is minimized by solving the corresponding Euler-Lagrange equations using the nonlinear multigrid scheme [10]. In this step, we compute the flow bidirectionally, i.e. a flow pair $(\mathbf{w}_1^0, \mathbf{w}_2^0)$, indicating mappings from I_1 to I_2 and from I_2 to I_1 respectively.

This initial flow estimates contain errors mostly in the occluded and textureless regions. One example is shown in Fig. 3 where the flow does not preserve clear edges and the whole map looks over-smoothed. We thus propose incorporating segmentation to improve it.

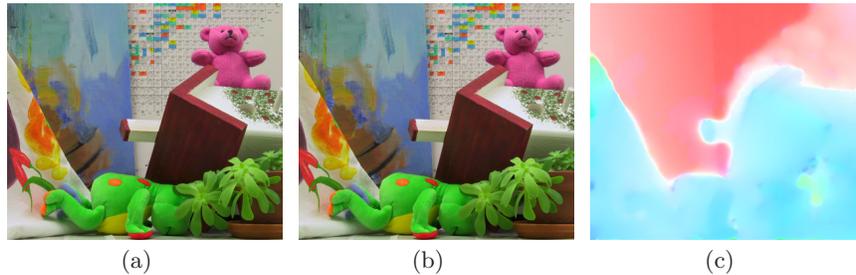


Fig. 3. An example of the initial flow. (a) and (b) show the image pair of the “Teddy” example from the Middlebury dataset [23]. (c) The dense flow result obtained in initialization. It is over-smoothed and contains errors around the object boundary.

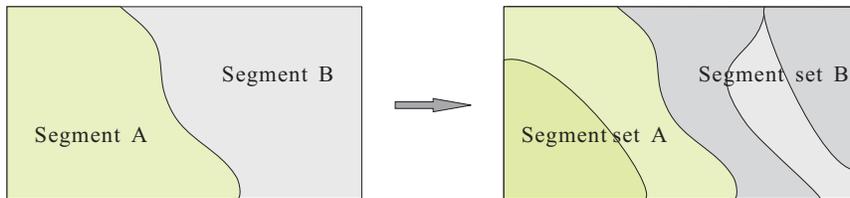


Fig. 4. Two-pass segmentation demonstration.

3.2 Color-Motion Segmentation

In our optical flow framework, the segments are produced counting in both the color and initial flow information, which reduces the possibility of mis-classifying a pixel to a segment only using color. Our method uses a two-pass segmentation scheme which first partitions the input images with regard to color to preserve edge structures, and then further “splits” each segment into more motion-distinctive patches. The two-pass segmentation is performed, in our approach, by the Mean-shift method [24] using the 3D color and 2D flow information respectively. The segmentation parameters all have fixed values in experiments (7 for spatial bandwidth, 6.5/0.02 for the range bandwidth of color/motion, 200 for minimum region size). Fig. 4 illustrates that segments A and B are produced using the color information. They are further split into more motion patches. A merging operation is performed to remove small patches near color discontinuities, as they are typically caused by occlusions and should not be treated as independent segments.

3.3 Parametric Flow Estimating Incorporating Segmentation

Regional information is used in stereo or the motion estimation by assuming a parametric model [7, 4] – that is, within each segment, all pixels are in comply with a single translational or affine motion model where the model parameters are estimated either by a regression method [7], or using plane fitting [2, 3].

However, these methods have inherent drawbacks when applied to noisy data in dense flow estimation. The regression method is known as possibly suffering from the local aperture problem. When the size of a segment is not large enough, this approach may lead to an unpredictable parameter estimate. The robust fitting technique depends too much on the initial variable values within each segment. In addition, both of the above methods do not enforce an intra-segment constraint, which is found quite useful in our method to reduce region-wise errors.

We estimate the parametric motions in a regularization framework, similar to that used in [14, 16]. Specifically, we define an affine model and denote by $\mathbf{a}_s = (a_{s0}, a_{s1}, a_{s2}, a_{s3}, a_{s4}, a_{s5})^T$ the vector of all affine parameters for segment s . The motion field $\mathbf{w}(\mathbf{a}_s, \mathbf{x}) = (u(\mathbf{a}_s, \mathbf{x}), v(\mathbf{a}_s, \mathbf{x}))^T$ in segment s is given by

$$\begin{aligned} u(\mathbf{a}_s, \mathbf{x}) &= a_{s0}x + a_{s1}y + a_{s2}, \\ v(\mathbf{a}_s, \mathbf{x}) &= a_{s3}x + a_{s4}y + a_{s5}. \end{aligned}$$

With the above parametric flow representation in each segment, the energy function w.r.t. all \mathbf{a} 's for image I_1 can be written as:

$$\begin{aligned} E_1(\mathbf{a}) &= \int_S \int_s \sum_{c=1}^3 \Psi_D(|I_2(\mathbf{x} + \mathbf{w}(\mathbf{a}_s, \mathbf{x}), c) - I_1(\mathbf{x}, c)|^2) d\mathbf{x} ds + \\ &\alpha \int_{\Omega_1} \Psi_S(\|\nabla u(\mathbf{a}_s, \mathbf{x})\|^2 + \|\nabla v(\mathbf{a}_s, \mathbf{x})\|^2) d\mathbf{x}, \end{aligned} \quad (4)$$

where S is the set of all segments in image I_1 . (4) is minimized in a coarse-to-fine manner using a Gaussian pyramid. The affine parameters \mathbf{a}^{k+1} in level $k+1$ are computed by adding increments to the estimated result in level k , i.e., $\mathbf{a}^{k+1} = \mathbf{a}^k + \Delta\mathbf{a}^{k+1}$. In each pyramid level, we use the Taylor expansion to approximate the increments by throwing away high-order terms. This gives us a new increment data term

$$E_{D'}(\Delta\mathbf{a}^{k+1}, \mathbf{x}) = \sum_{c=1}^3 \Psi_D(|(I_x^c)^k \cdot u(\Delta\mathbf{a}_s^{k+1}, \mathbf{x}) + (I_y^c)^k \cdot v(\Delta\mathbf{a}_s^{k+1}, \mathbf{x}) + (I_z^c)^k|^2) \quad (5)$$

where $(I_i^c)^k = \partial_i I_2(\mathbf{x} + \mathbf{w}(\mathbf{a}_s^k, \mathbf{x}), c)$, $i = \{x, y\}$, denoting the spatial derivatives. \mathbf{a}_s^k is the affine parameters for segment s estimated in level k . $(I_z^c)^k = I_2(\mathbf{x} + \mathbf{w}(\mathbf{a}_s^k, \mathbf{x}), c) - I_1(\mathbf{x}, c)$. It represents the temporal difference. The smoothness term is written as

$$E_{S'}(\Delta\mathbf{a}^{k+1}, \mathbf{x}) = \Psi_S(\|\nabla u(\mathbf{a}_s^k + \Delta\mathbf{a}_s^{k+1}, \mathbf{x})\|^2 + \|\nabla v(\mathbf{a}_s^k + \Delta\mathbf{a}_s^{k+1}, \mathbf{x})\|^2), \quad (6)$$

where ∇u and ∇v are approximated by the forward difference. In implementation, the smoothness term is further separated into two parts, i.e., the inter-segment and the intra-segment smoothness w.r.t. the locations of neighboring pixels in computing ∇u and ∇v . The inter-segment smoothness is imposed on the segment boundaries and is used to propagate information among regions. The intra-segment smoothness is enforced within each segment and is uniformly

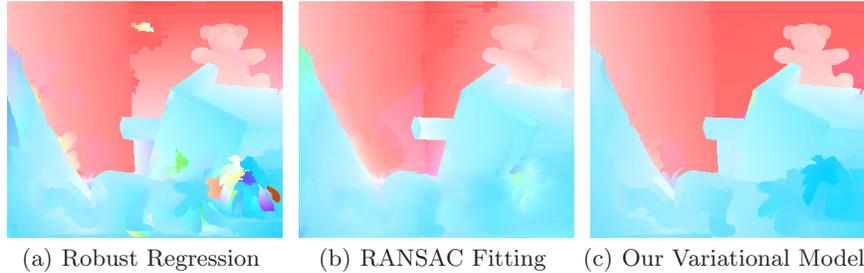


Fig. 5. Comparison of the estimated motion by three methods incorporating the same segmentation information. Our result not only preserves high-quality boundary, but also contains accurate motion estimate within small patches.

represented as $(a_{s_0}^{k+1})^2 + (a_{s_1}^{k+1})^2 + (a_{s_3}^{k+1})^2 + (a_{s_4}^{k+1})^2$. It regularizes the affine parameters and enforces a translational motion model. This is useful for the ubiquitous small-size textureless regions when the data term is not trustworthy.

The final energy to be minimized in pyramid level $k + 1$ combining $E_{S'}$ and $E_{D'}$ is expressed as

$$E(\Delta \mathbf{a}^{k+1}) = \sum_{\mathbf{x}} \{E_{D'}(\Delta \mathbf{a}^{k+1}, \mathbf{x}) + \alpha E_{S'}(\Delta \mathbf{a}^{k+1}, \mathbf{x})\}. \quad (7)$$

Since (7) is continuous, differentiable, and convex with respect to $\Delta \mathbf{a}^{k+1}$, we can use any gradient-based method (our system uses the Quasi-Newton method) to minimize it. In experiments, the flow estimation for each level of the pyramid only needs 5 – 7 iterations. The total number of variables (i.e., the affine parameters) to be updated is only $6N_s$ (N_s is the number of segments), which is much smaller than the number of dense flow vectors. So the optimization in this step is efficient and robust.

To demonstrate the effectiveness of the segmentation-combined variational model, we show in Fig. 5 a comparison of the motion results generated by robust regression, robust fitting, and our method based on the same initial flow and segmentation. The robust regression [7] is achieved by removing the regularization term in (4). The plane fitting result is obtained by performing RANSAC [25] for the initial flow to fit the affine parameters. We take 1000 iterations to produce the result. Comparably, these two methods have difficulties to get accurate flow estimates for small segments in the textureless or occluded regions. Our method, on the contrary, takes the advantages of segmentation as well as the intra-segment smoothness, thus can significantly reduce the errors.

3.4 Confidence Map Construction

Segmentation can improve flow estimation in textureless regions. However, it also induces problems for regions undergoing non-rigid motion. In this circumstance, the affine model in each segment could be over-restrictive. In order to alleviate

this problem, we propose constructing a confidence map to indicate how likely the estimated flow vectors in the above step are correct, respectively based on the pixel-wise motion coherence and segment-wise model confidence. In what follows, we first detect image occlusion using the motion information.

Occlusion detection. To compute the occlusion in I_1 , we simply warp I_2 to I_1 based on the motion vectors in the flow field of I_2 . If a pixel \mathbf{x} in I_1 does not receive projection from I_2 , we set its occlusion value $O(\mathbf{x})$ to 1; otherwise, $O(\mathbf{x})$ is set to 0. We do not use a global optimization to enforce smoothness since it is found that the detected map is already sufficiently usable for the purpose of our motion confidence evaluation.

Pixel-wise motion coherence. In the following discussion, we denote by $(\mathbf{w}_1^0, \mathbf{w}_2^0)$ the flow pair estimated in the initial step and $(\mathbf{w}_1^s, \mathbf{w}_2^s)$ the flow pair computed in the second step for images I_1 and I_2 respectively. We construct function $E_p(\mathbf{w}_1^i, \mathbf{x})$ for each point \mathbf{x} in flow field \mathbf{w}_1^i to measure the motion coherence:

$$E_p(\mathbf{w}_1^i, \mathbf{x}) = \exp\left(-\frac{\sum_{c=1}^3 |I_2(\mathbf{x} + \mathbf{w}_1^i(\mathbf{x}), c) - I_1(\mathbf{x}, c)|^2}{3\sigma_I^2}\right) \cdot \exp\left(-\frac{\|\mathbf{w}_1^i(\mathbf{x}) + \mathbf{w}_2^i(\mathbf{x} + \mathbf{w}_1^i(\mathbf{x}))\|^2}{\sigma_w^2}\right). \quad (8)$$

Here, the superscript $i = s$, denoting the confidence for the flow field in the 2nd step. $E_p(\mathbf{w}_1^s, \mathbf{x})$ is composed of two terms. $|I_2(\mathbf{x} + \mathbf{w}_1^s(\mathbf{x}), c) - I_1(\mathbf{x}, c)|^2$ models the color constancy between two matched pixels by a motion vector; $\|\mathbf{w}_1^i(\mathbf{x}) + \mathbf{w}_2^i(\mathbf{x} + \mathbf{w}_1^i(\mathbf{x}))\|^2$ models the motion coherence with respect to both images, similar to the cross check error defined in [3]. In all our experiments, $I_2(\mathbf{x} + \mathbf{w}_1^s(\mathbf{x}), c)$ and $\mathbf{w}_2^s(\mathbf{x} + \mathbf{w}_1^s(\mathbf{x}))$ are obtained using bilinear interpolation. The pixel-wise confidence for the flow computed in the second step is defined as:

$$C_p(\mathbf{w}_1^s, \mathbf{x}) = \begin{cases} \zeta & \text{if } O(\mathbf{x})=1, \\ E_p(\mathbf{w}_1^s, \mathbf{x}) & \text{otherwise,} \end{cases} \quad (9)$$

where $O(\mathbf{x})$ is the occlusion value. ζ is a constant to penalize the occluded pixels.

Segment-wise motion confidence. Only defining the pixel-wise motion coherence is not enough for the textureless segments with complex motion, since both the color constancy and motion coherence measure in E_p could have small values, which contradicts the true confidence definition. So we introduce the supplementary segment-wise confidence C_s , i.e. the confidence of the motion in a segment being affine, to handle the above problem. We define

$$C_s(\mathbf{w}_1^s, s) = \frac{\sum_{\mathbf{x} \in s} \exp(-\|\mathbf{w}_1^s(\mathbf{x}) - \mathbf{w}_1^0(\mathbf{x})\|^2) E_p(\mathbf{w}_1^0, \mathbf{x}) / \sigma_A^2 (1 - O(\mathbf{x}))}{\sum_{\mathbf{x} \in s} (1 - O(\mathbf{x}))}, \quad (10)$$

where s denotes a segment. $(1 - O(\mathbf{x}))$ is to exclude the occluded pixels in computing the confidence of a segment since the initial flow is usually erroneous

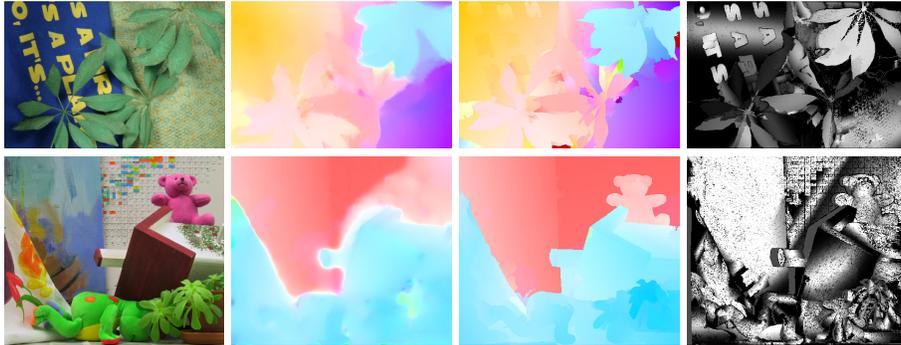


Fig. 6. Confidence maps for examples “Schefflera” (top row) and “Teddy” (bottom row). From left to right: one input image, initial optical flow (from step 1), optical flow with segmentation (from step 2), the constructed confidence map. Note that most pixels in the confidence map of “Teddy” are with high confidence. “Schefflera” contains non-rigid motion. So more flow vectors are problematic.

for these pixels. $E_p(\mathbf{w}_1^0, \mathbf{x})$ is defined in (8) by setting $i = 0$, modeling the pixel-wise flow confidence for the initial estimate. A small value of $E_p(\mathbf{w}_1^0, \mathbf{x})$ means we should not trust the initial flow. $\|\mathbf{w}_1^s(\mathbf{x}) - \mathbf{w}_1^0(\mathbf{x})\|^2$ measures how the flow in the second step is modified over that in initialization. If the initial flow is trustworthy, i.e., with large $E_p(\mathbf{w}_1^0, \mathbf{x})$, a large difference between $\mathbf{w}_1^s(\mathbf{x})$ and $\mathbf{w}_1^0(\mathbf{x})$ indicates that the affine model in a segment does not fit the actual pixel motion. So we should re-estimate the flow vectors for these pixels.

In (10), C_s returns a small value only if many pixels in a segment have high initial flow confidence and the corresponding initial flow vectors are quite different from those with segmentation (from step 2). It reflects that the parametric affine motion estimate in one segment is erroneous. The final confidence for \mathbf{w}_1^s is a combination of the two measures:

$$\mathbf{conf}(\mathbf{x}) = C_p(\mathbf{w}_1^s, \mathbf{x}) \cdot C_s(\mathbf{w}_1^s, s(\mathbf{x})), \quad (11)$$

where $s(\mathbf{x})$ denotes the segment that contains pixel \mathbf{x} . Two examples of $\mathbf{conf}(\cdot)$ are shown in Fig. 6 where dark pixels indicate the possible erroneous estimates generated in the segmentation step (step 2). In these maps, many low confidence pixels are in non-rigid bodies.

3.5 Final Variational Model

We integrate the estimated confidence map into a final flow refinement phase to correct the possible flow errors due to segmentation. The final energy function is defined as

$$E_2(u, v) = \int_{\Omega_1} (1 - O(\mathbf{x})) \sum_{c=1}^3 \Psi_D(|I_2(\mathbf{x} + \mathbf{w}, c) - I_1(\mathbf{x}, c)|^2) + \beta \mathbf{conf}(\mathbf{x}) \|\mathbf{w} - \mathbf{w}_1^s\|^2 + \alpha \Psi_S(\|\nabla u\|^2 + \|\nabla v\|^2) d\mathbf{x}. \quad (12)$$

There are three energy terms defined in $E_2(u, v)$. $(1 - O(\mathbf{x}))$ is to make color distance not be considered on the occluded pixels. $\beta \mathbf{conf}(\mathbf{x}) \|\mathbf{w} - \mathbf{w}_1^s\|^2$ imposes a soft constraint. When weight $\mathbf{conf}(\mathbf{x})$ has a large value, the flow computed with segmentation (in step 2) will be trusted. Otherwise, other energy terms will be more influential in estimating the flow vector.

$E_2(u, v)$ selectively refines the flow vectors computed in the segmentation step, where the confidence map provides an essential clue whether the flow estimated in this step is correct or not. We minimize (12) by solving the corresponding Euler-Lagrange equations, similar to how we minimize (3).

4 Experimental Results

Experiments on both the synthesized and real images were conducted. Parameters in all experiments are configured with constant values as shown in table 1.

(a) stage 1 & 2			(b) stage 3							
α	ϵ_D	ϵ_S	α	β	ϵ_D	ϵ_S	σ_I	σ_w	σ_A	ς
50	0.1	0.01	30	100	0.1	0.01	80	0.15	0.3	0.2

Table 1. Parameters used in our experiments

In regard to the parameter adjustment, specifically, ϵ_D and ϵ_S are set only for numerical stability. α and β are used to balance the energy terms. Note that the smoothness weight is set lower in our final stage as the segmentation information is incorporated. σ_I , σ_w and σ_A control the impact of the terms in constructing the confidence map. Larger values imply lower impact.

4.1 Quantitative Evaluation

Quantitative evaluation of the optical flow algorithm is conducted using the dataset in [23]. The overall rank of our method is high amongst all recorded optical flow algorithms on the Middlebury website based on the average angular error (AAE). We show in the second column of table 2 the average rank of the top eight algorithms at the moment we submit the data. Other columns on the right only show the AAE around motion discontinuities. Statistics show that our method has an advantage in faithfully preserving motion boundaries. The optical flow results are shown in Fig. 7.

For the computation speed, although the segmentation adds extra cost to the flow estimation, by using the multigrid scheme in the first and third steps, the total running time to process one image does not increase much. Typically, for an image with size 316×252 , the running time of our algorithm is about 15 seconds on a PC with an Intel Core2Due 2.4G CPU.

Algorithm	Avg.								
	Rank	Army	Mequon	Schefflera	Wooden Grove	Urban	Yosemite	Teddy	
Our Method	3.1	13.5 ₂	14.9 ₄	17.3 ₂	18.1₁	4.16₁	21.4 ₅	3.49 ₂	9.23₁
Fusion	3.3	13.7 ₃	8.91₁	9.68₁	19.8 ₂	4.82 ₅	17.3₁	5.78 ₈	13.6 ₄
SO prior	4.7	11.2₁	13.1 ₂	17.7 ₃	20.9 ₃	5.27 ₇	22.0 ₇	6.88 ₉	13.4 ₃
Dynamic MRF	4.8	15.0 ₄	15.3 ₅	17.8 ₄	23.7 ₄	4.63 ₄	19.1 ₄	5.29 ₅	17.8 ₇
LP Registration	5.8	16.8 ₅	13.8 ₃	17.8 ₄	24.5 ₅	4.56 ₃	21.4 ₅	5.48 ₇	17.9 ₈
B. & A.	6.4	18.7 ₆	21.9 ₇	23.7 ₆	30.0 ₆	5.23 ₆	18.2 ₂	4.44 ₃	14.3 ₅
2D-CLG	6.6	22.6 ₉	16.9 ₆	28.2 ₉	31.1 ₈	4.25 ₂	22.2 ₈	3.14₁	12.9 ₂
H. & S.	7.8	19.9 ₈	23.2 ₁₀	25.9 ₇	30.6 ₇	5.27 ₇	25.8 ₁₀	5.41 ₆	17.5 ₆

Table 2. Average Angular Error (AAE) comparison obtained from the Middlebury website [23]. The second column shows the average ranks and other columns on the right show AAEs around the motion discontinuities.

	Our results	Black and Anandan [7]
frame 10→0	0.88 %	3.33 %
frame 15→0	3.42 %	5.98 %
frame 20→0	6.13 %	9.26 %

Table 3. Average percentage of the mis-matched pixels (defined in Sect. 4.2) for the three sequences by warping frames 10, 15 and 20 to frame 0. For comparison, we also show the statistics from the method of Black and Anandan [7].

4.2 Image Warping Results

We also conducted experiments on natural image sequences to estimate the flow field. By warping a frame in the tail back to one at the head using the flow computed in all intermediate frames, we can evaluate the quality of optical flow in terms of the accumulated accuracy. Three different sequences are used¹ in experiments with the following evaluation criteria. We first compute the intensity difference between the warped frames and the original one. Then the average percentage of pixels whose intensity difference is greater than 0.1 of the maximum intensity is computed. The statistics are shown in Table 3, attained by warping frames 10, 15, 20 to frame 0, respectively.

One warping result is shown in Fig. 8 where we warp frame 15 back to frame 0. To produce the result in (b), we use the code from the authors by hand tuning the parameters. The close-ups in the bottom row show that our method faithfully preserves motion discontinuities and produces sharp warping boundaries.

¹ available at <http://www.cse.cuhk.edu.hk/~leojia/publication.html>

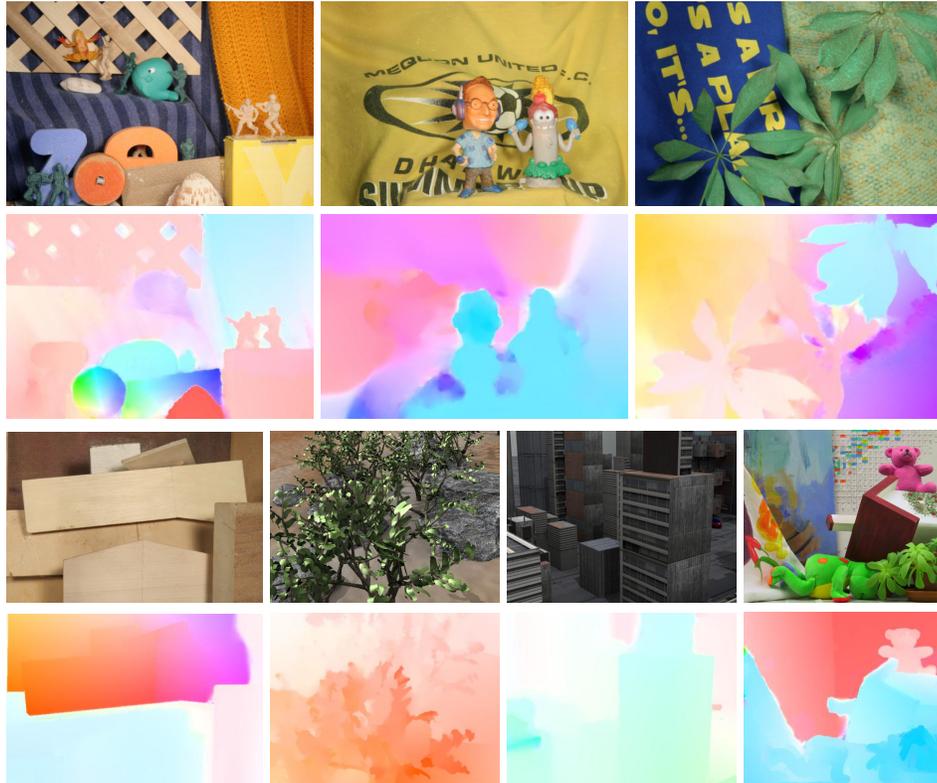
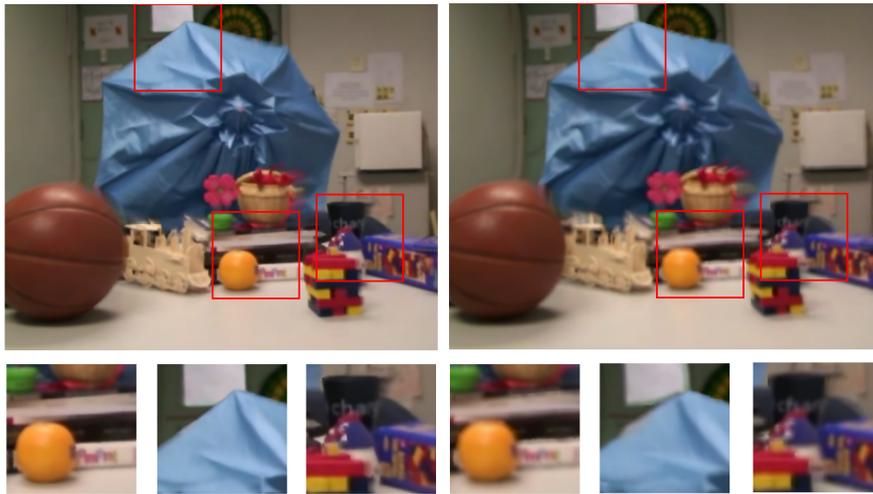


Fig. 7. Our optical flow estimates on a set of challenging examples. Incorporating segmentation and using parametric motion models do not degrade our results because of the multi-step estimation framework.

5 Conclusion

In this paper, we have described a segmentation-embedded optical flow framework which can be used to compute accurate flow field as well as high quality motion boundary. The proposed method accommodates the parametric and segmented motion estimation in a variational model. Then a confidence map is constructed to measure the confidence whether the segmentation and the corresponding motion model suit the flow estimation or not. This map enables the recognition of non-rigid motion and detection of the error caused by segmentation. Evaluation on the Middelbury data set validated the effectiveness of our method. Our segmentation-based method produces sharp motion boundary, having a clear advantage in applications such as video editing.



(a) Our result

(b) Black and Anandan [7]

Fig. 8. One example of warping frame 15 to frame 0.

Acknowledgements

We thank Guofeng Zhang for his comments on this paper. This work was fully supported by a grant from the Research Grants Council of Hong Kong (Project No. 412708).

References

1. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S.A.J., Szeliski, R.: High-quality video view interpolation using a layered representation. *ACM Trans. Graph.* **23**(3) (2004) 600–608
2. Sun, J., Li, Y., Kang, S.B.: Symmetric stereo matching for occlusion handling. In: *CVPR* (2). (2005) 399–406
3. Yang, Q., Wang, L., Yang, R., Stewénius, H., Nistér, D.: Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. In: *CVPR* (2). (2006) 2347–2354
4. Zitnick, C.L., Jojic, N., Kang, S.B.: Consistent segmentation for optical flow estimation. In: *ICCV*. (2005) 1308–1315
5. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *IJCAI*. (1981) 674–679
6. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artif. Intell.* **17**(1-3) (1981) 185–203
7. Black, M.J., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding* **63**(1) (1996) 75–104

8. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: ECCV (4). (2004) 25–36
9. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision* **61**(3) (2005) 211–231
10. Bruhn, A., Weickert, J.: Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In: ICCV. (2005) 749–755
11. Xiao, J., Cheng, H., Sawhney, H.S., Rao, C., Isnardi, M.A.: Bilateral filtering-based optical flow estimation with occlusion detection. In: ECCV (1). (2006) 211–224
12. Ben-Ari, R., Sochen, N.: Variational stereo vision with sharp discontinuities and occlusion handling. In: ICCV. (2007)
13. Tschumperlé, D., Deriche, R.: Diffusion tensor regularization with constraints preservation. In: CVPR (1). (2001) 948–953
14. Ju, S.X., Black, M.J., Jepson, A.D.: Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In: CVPR. (1996) 307–314
15. Black, M.J., Jepson, A.D.: Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(10) (1996) 972–986
16. Mémin, É., Pérez, P.: Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision* **46**(2) (2002) 129–155
17. Cremers, D., Schnörr, C.: Motion competition: Variational integration of motion segmentation and shape regularization. In: DAGM-Symposium. (2002) 472–480
18. Amiaz, T., Kiryati, N.: Piecewise-smooth dense optical flow via level sets. *International Journal of Computer Vision* **68**(2) (2006) 111–124
19. Black, M.J.: Combining intensity and motion for incremental segmentation and tracking over long image sequences. In: ECCV. (1992) 485–493
20. Khan, S., Shah, M.: Object based segmentation of video using color, motion and spatial information. In: CVPR (2). (2001) 746–751
21. Brox, T., Bruhn, A., Weickert, J.: Variational motion segmentation with level sets. In: ECCV (1). (2006) 471–483
22. Zhang, G., Jia, J., Xiong, W., Wong, T.T., Heng, P.A., Bao, H.: Moving object extraction with a hand-held camera. In: ICCV. (2007)
23. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. In: ICCV. (2007)
24. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5) (2002) 603–619
25. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6) (1981) 381–395