

# Estimating Piecewise-Smooth Optical Flow with Global Matching and Graduated Optimization

Ming Ye, Robert M. Haralick, *Fellow, IEEE*, and Linda G. Shapiro  
*Fellow, IEEE*

Ming Ye is with the Microsoft Corporation, One Microsoft Way, Redmond WA 98052. E-mail: mingye@microsoft.com.

Robert M. Haralick is with the Computer Science Department, CUNY Graduate Center, New York, NY 10016. E-mail: haralick@gc.cuny.edu.

Linda G. Shapiro is with the Computer Science and Engineering Department, University of Washington, Seattle, WA 98195. E-mail: shapiro@cs.washington.edu.

## Abstract

This paper presents a new method for estimating piecewise-smooth optical flow. We propose a global optimization formulation with three-frame matching and local variation and develop an efficient technique to minimize the resultant global energy. This technique takes advantage of local gradient, global gradient and global matching methods and alleviates their limitations. Experiments on various synthetic and real data show that this method achieves highly competitive accuracy.

## I. INTRODUCTION

Optical flow is a 2D image motion measure that has a wide range of applications in computer vision [14], video coding [16] and computer graphics [35]. Its accurate and efficient estimation is a long-standing difficult problem.

The fundamental assumption enabling optical flow estimation is *brightness conservation*:

$$I(x, y, t) = I(x + u\Delta t, y + v\Delta t, t + \Delta t), \quad (1)$$

where  $I(x, y, t)$  is the image intensity at point  $(x, y)$  and time  $t$ , and  $(u, v)$  is the optical flow vector. Depending on what variation of Eq.(1) is used, optical flow estimation methods can be classified into two main categories: matching-based and gradient-based<sup>1</sup>. Matching-based methods make direct use of Eq.(1). They can handle large motion and avoid tricky derivative calculations, but they often meet with computational difficulties and yield poor sub-pixel accuracy [2]. Gradient based methods use the linear approximation of Eq. (1)

$$I_x u + I_y v + I_t = 0, \quad (2)$$

a.k.a the Optical Flow Constraint Equation (OFCE) [2], where  $(I_x, I_y, I_t)$  is the spatiotemporal image intensity gradient. These methods have become the most popular because of the relatively low computational complexity and good accuracy, but they can break down in the presence of large motion and inaccurate derivative estimates [38].

Additional constraints on the flow are obtained from various flow field models [2], [4]. Among them we are particularly interested in *piecewise smooth* models [5], because they are applicable to general scenes and are indispensable building blocks, leading to more complex motion analysis in a bottom-up fashion[30]. Piecewise smooth models include

<sup>1</sup>Frequency/phase-based methods are often close to frequency-domain equivalents of the above two methods [2].

two types: local parametric and global optimization. Local parametric methods assume that within a certain region the flow field is described by a parametric model [4], with the simplest and most popular model being piecewise constant. Their accuracy and efficiency are among the best according to various comparative studies [2], [13], but they degrade or fail when local information becomes insufficient or unreliable. Global optimization methods try to strike a balance between overall brightness error and smoothness error by minimizing a global energy, which might be developed from regularization [5], [21] or Bayesian (MAP, MRF) [19], [6], [29] perspectives. They are less sensitive to poor local constraints, but existing techniques of this type, even those using robust estimators [5], tend to oversmooth the flow field.

Traditional techniques [2] usually require brightness conservation and flow smoothness to be satisfied in a strict (least-squares, LS) sense. They yield limited accuracy in practice where model violations due to motion discontinuities and abrupt intensity changes are abundant. As the limitations are widely recognized, a large number of recent efforts have been devoted to increasing robustness especially in the presence of motion discontinuities. Among these attempts, replacing LS estimators with robust estimators has achieved the most success [27], [28], [5], [21], [1], [26], [38]. Reformulating global optimization criteria in terms of MRF with line processes [22], [19], [15], [6], weak continuity [8], [5] or anisotropic diffusion [2], [7] is another widely pursued direction. Despite these advances, handling occlusion remains a tough problem for the field.

Unsatisfactory state-of-the-art performance is largely due to formulation defects and solution complexity. On one hand, approximate formulations are frequently adopted for ease of computation, with the consequence that the correct flow is unrecoverable even in ideal settings. As an example, many methods intended to preserve motion discontinuities use gradient-based brightness constraints, which break down at discontinuities due to derivative evaluation failure [38]. On the other hand, more sophisticated formulations typically involve large-scale nonconvex optimization problems, which are so hard to solve that the achievable accuracy might not be competitive with simpler methods. Motion estimation research has arrived at a stage at which a good collection of ingredients are available; but in order to significantly improve performance, both problem formulation

and solution methods need to be carefully considered.

In this paper we discuss a matching-based global optimization method with a practical solution technique. The formulation uses three-frame matching to reduce visibility problems at occlusions, and balances the strengths of brightness and smoothness errors according to local data variation. We develop a hierarchical three-step optimization strategy to solve the resulting energy minimization problem. Step I uses a high-breakdown robust local gradient-based method with a deterministic iterative implementation, which provides a high-quality initial flow estimate. Step II uses a global gradient-based method which efficiently improves the flow field coherence. Step III minimizes the original energy by greedy propagation. It corrects gross errors introduced by derivative evaluation and pyramid operations. In this process, merits tend to be inherited from all three techniques. As a result, high accuracy is achieved both on and off motion boundaries.

## II. FORMULATION

We assume the optical flow field  $V$  minimizes the global energy

$$E = \sum_{\text{all pixels } i} E_B(V_i) + E_S(V_i) \quad (3)$$

where  $V_i$  is the flow vector at pixel  $i$ ;  $E_B$  and  $E_S$  represent brightness conservation and flow smoothness respectively.

The traditional assumption that pixels are visible in all frames is a major source of gross errors in occlusion areas. Treating such violations as outliers [6] may prevent error from propagating to nearby regions, but does not provide constraints for occlusion pixels and thus does not help their motion estimation. Introducing additional fields to represent occlusion [10], [20] vastly increases the problem complexity. We observe that normally all points in a frame are visible in the previous *or* the next frame. Assuming constant motion within three frames, we define the matching error as the minimum of the backward and forward warping errors, i.e.,

$$e_W(V_i) = \min(|I_b(V_i) - I_i|, |I_f(V_i) - I_i|) \quad (4)$$

where  $I_i$  is the intensity of pixel  $i$  in the middle frame;  $I_b(V_i)$ ,  $I_f(V_i)$  are warped intensities in the previous and the next frames respectively. This error design is simple yet very

effective. Similar ideas have shown good results in recent independent studies [31], [39], [18]. We further use a robust error function  $\rho(x, \sigma)$  to resist other sources of outliers, yielding the brightness energy

$$E_B(V_i) = \rho(e_W(V_i), \sigma_{B_i}) \quad (5)$$

where  $\sigma_{B_i}$  is the local brightness variation scale. We define the smoothness error to be

$$E_S(V_i) = \frac{1}{8} \sum_{j \in N_i^8} \rho(|V_i - V_j|, \sigma_{S_i}) \quad (6)$$

where  $\sigma_{S_i}$  is the local flow variation scale. It requires a vector  $V_i$  to be consistent with its 8-connected neighbors  $\{V_j, j \in N_i^8\}$  and the robust error function prevents smoothing across motion boundaries. We choose to use the Geman-McClure robust error function  $\rho(x, \sigma) = x^2/(x^2 + \sigma^2)$  [5] in both the  $E_B$  and  $E_S$  terms for its redescending [5], [24] and normalizing properties. The first property ensures that the outlier influence tends to zero. We take errors exceeding

$$\tau = \sigma/\sqrt{3}, \quad (7)$$

where the influence function begins to decrease, as outliers [5]. The normalization property is desirable because the brightness and smoothness energies become comparable and their relative strengths can be adjusted locally by the scales  $\sigma_{B_i}, \sigma_{S_i}$ —where the observation is not trustworthy ( $\sigma_{B_i}$  is large), stronger smoothness is enforced, and vice versa. We gradually learn the scales during the optimization process (Section III). Previous global formulations [2], [5], [21], usually have a control parameter  $\lambda$  between  $E_B$  and  $E_S$ , and  $\sigma_B, \sigma_S, \lambda$  are all globally tuned parameters. Compared to such approaches, this locally adaptive scheme is more reasonable and reduces parameter tuning in experiments.

### III. OPTIMIZATION

Minimizing a global energy like Eq. (3) is very hard. Stochastic methods such as simulated annealing converge too slowly to be practical [6]. Deterministic methods [6], [21] have achieved more success, but they have limited capability for avoiding local minima and their performance depends on the initialization quality. Our approach to this problem is *hierarchical graduated optimization*. We create a  $P$ -level image pyramid  $I^p, p = 0, \dots, P-1$

and begin estimation from the top (coarsest) level  $P - 1$  with a zero initial flow field [4]. At every level the algorithm proceeds in three steps, each computing a finer approximation to the original energy.

#### A. Step I: Gradient-Based Local Regression

Suppose a crude flow estimate  $V_0$  is available and has been compensated for. Step I uses a robust gradient-based local regression method to compute the incremental flow  $\Delta V_0$ . For each pixel, we find the dominant translational motion in its  $W \times W$  neighborhood by solving the  $n = W * W$  set of OFCEs (Eq. (2)) under a least-median-of-squares (LMS) criterion [24]. The LMS solution is usually approximated by the estimate yielding the smallest criterion value in a set of trial estimates. In the recipe given by Rousseeuw and Levoy (RL) [24], a trial estimate is calculated from each pair of constraints and  $m$  estimates are generated by randomly drawing from the total  $C_n^2$  possibilities. They also suggest a reweighted least squares (RLS) procedure to improve the statistical efficiency of the LMS estimate. For each trial estimate, they calculate the variance of healthy residual errors as  $s = 1.4826 * \text{median}_{i=1}^n r_i^2$ , identify constraints agreeing with the trial estimate as those with residuals satisfying  $|r_i| \leq 2.5s$ , and update the estimate using LS on the inliers. Previous uses of LMS in motion estimation [1], [26], [38] uniformly apply the above algorithm to all pixels. This incurs unnecessary computation at the majority of the places where no outliers exist and least-squares suffices; meanwhile the accuracy is compromised at places of heavy contamination where a small trial set might not contain a good estimate.

By taking advantage of the piecewise smoothness property of optical flow and the reasonable accuracy of the LS estimate at the majority of pixels, we propose a deterministic adaptive algorithm to generate the trial set [39]. The idea is to start with the LS flow field and generate the trial solution set from neighboring values instead of by random drawing. In our experiments, we form a trial set  $V_{\{j\}}$  for each  $V_i$  using values at the corners and side midpoints of the  $W \times W$  box centered at site  $i$ , keep estimates yielding lower criterion values, and iterate until no update occurs.

This algorithm in effect provides an estimator whose complexity depends on the actual outlier contamination. It can be faster and yield more stable accuracy than algorithms based on random sampling. We show its advantages by comparing its results with the RL

algorithm using  $m = 30$  random samples; in this case the probability of having at least one good initial estimate is as high as 99.98% even with 50% outliers present [1]. On the TS sequence (Section IV), which only contains flow outliers at the motion boundary so that the RL algorithm can almost recover the LMS solutions, our method produces virtually the same accuracy with less computation. Fig. 2(e) shows the number of trials at each pixel as an intensity image. More trials, indicated by brighter colors, are carried out closer to the boundary where the motion is more complex. The number ranges from 1 to 13, as opposed to the fixed 30 in the RL case. Comparisons were also conducted on all the other sequences described in Section IV and our method always produced higher accuracy in less time. As a side note, it would be interesting to see how the trial set size could be used as an early cue for scene complexity.

### B. Step II: Gradient-Based Global Optimization

Recall that  $V_0$  is the initial flow estimate that has been compensated for and  $\Delta V_0$  is the incremental flow resulting from Step I. Step II improves the coherence of  $\Delta V_0$  by using a gradient-based global optimization method. The energy to minimize is a closer approximation to Eq. (3):

$$E(\Delta V) = \sum_i [\rho(e_G(\Delta V_i), \sigma_{B_i}) + \frac{1}{8} \sum_{j \in N_i^8} \rho(V_i + \Delta V_i - V_j - \Delta V_j, \sigma_{S_i})] \quad (8)$$

where  $e_G$  is the OFCE residual (Eq. (2)) and  $V_i$  is the  $i$ th vector of the initial flow  $V_0$ . The local scales  $\sigma_{B_i}, \sigma_{S_i}$  are important parameters which control the shape of  $E$  and hence the solution. Below we describe how to estimate  $\sigma$ 's from Step I's results.

Assume that normal errors are zero-mean Gaussian variables with standard deviation  $\tilde{\sigma}$ , and outliers have errors exceeding  $2.5\tilde{\sigma}$ . Contrasting this to Eq. (7) we can express  $\sigma$  in terms of  $\tilde{\sigma}$  as  $\sigma = 2.5\sqrt{3}\tilde{\sigma}$ . At each site  $i$ , we calculate  $\tilde{\sigma}_{S_i}$  as the sample standard deviation of ‘‘inliers’’ of smoothness errors  $\{V_j - V_i, j \in N_i^8\}$ , with inliers selected by the RLS procedure in Section III-A.  $\tilde{\sigma}_{S_i}$  is further bounded above by a globally constant value  $1.4826 * \text{median}_i \tilde{\sigma}_{S_i}$  to prevent erroneously large values and bounded below by 0.001 for stability. We use the OFCE residual as  $\tilde{\sigma}_{B_i}$  and limit its value to the range  $[0.01, 1.4826 * \text{median}_i \tilde{\sigma}_{B_i}]$ . The scale computation in Step III is similar except that  $\tilde{\sigma}_{S_i}$  is limited to  $[0.004, 0.02]$  and we adopt a globally constant  $\tilde{\sigma}_B$ , estimated from matching

errors  $e_{W_i}$  (Eq. (4)) by  $\max(0.08, 1.4826 * \text{median}_i e_{W_i})$ . Although ad-hoc, the above method has shown effectiveness in capturing local scale variation.

With the scales specified, we minimize the energy using Successive OverRelaxation (SOR) [8], [6]. Starting with the initial estimate  $\Delta V_0$ , on the  $k$ th iteration, each  $u$  component (and  $v$  similarly) is updated as  $u_i^k = u_i^{k-1} - \omega/T(u_i)\partial E/\partial u_i^{k-1}$ , where  $T(u_i) = I_x^2/\sigma_{B_i}^2 + 8/\sigma_{S_i}^2$ . SOR is good at removing high-frequency errors while very slow at removing low-frequency errors [36], [9]. In our case, the initial estimate has dominant high-frequency errors—it has good accuracy at most places but may lack coherence due to the use of local constraints—and hence the SOR procedure is very effective and converges fast. In addition, the update step size is adjusted by spatially varying local scales (through  $T(u_i)$ ), which further improves the efficiency in exploring the solution space.

### C. Step III: Global Matching

$V_1$ , sum of the initial estimate  $V_0$  and the incremental flow  $\Delta V_1$  from Step II, still exhibits gross errors at motion boundaries and other places with poor gradient estimates. We reduce its errors by solving the original matching-based formulation Eq. (3) through greedy propagation. We first calculate the energy  $E_B(V_i) + E_S(V_i)$  from  $V_1$  for all pixels. Then we iteratively visit each pixel, examining whether a trial estimate from a candidate set results in a lower global energy  $E$ . The candidate set consists of the 8-connected neighbors and their average, which were updated in the last visit. Once a pixel energy decrease occurs, we accept the candidate and update the locally affected energy terms. The simple scheme works reasonably well because bad estimates are confined to narrow areas in the initial flow  $V_1$ . It converged quickly in our experiments. It is worth mentioning that a similar greedy propagation scheme was successfully applied to solving a global matching stereo formulation in an independent study [33].

### D. Overall Algorithm

Operations on each pyramid level are illustrated in Fig. 1. When more than one pyramid level is used, we choose to skip Step III on the coarsest level. The consideration is that gradient-based methods suffice on the coarsest level since the data are substantially smoothed and the flow is small; applying the matching constraint can be harmful due to

the smoothing and possible aliasing.

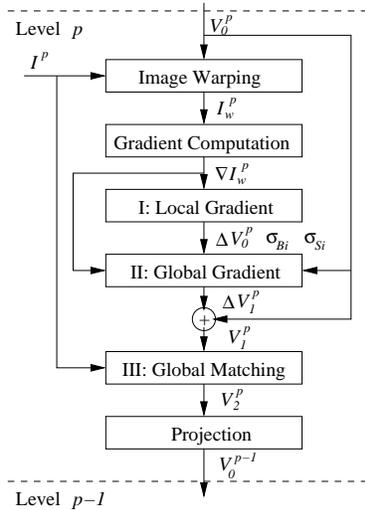


Fig. 1. System diagram (operations at each pyramid level)

From a practical point of view, the graduated scheme benefits from all three popular optical flow approaches. Step I (gradient-based local regression) generates high-quality initialization while leaving local ambiguities to be resolved later in more global formulations. Step II (gradient-based global optimization) improves the flow coherence; it converges fast because of the good initialization. Simple hierarchical schemes have the limitations that the projection and warping operations oversmooth the flow field, and errors in coarser levels are magnified and propagated to finer levels and are generally irreversible [3], [6]. These problems are much alleviated by Step III (matching-based global optimization)—it works on the original pyramid images and corrects gross errors caused by derivative computation, projection and warping.

#### IV. EXPERIMENTS

We estimate optical flow in the middle of every three frames. The number of pyramid levels is empirically determined and no other parameters are tuned. We also compare results with those produced by BA (Black and Anandan [6])—a good representative of previous dense regularization techniques. BA’s code is publicly available. It calculates flow on the second of two frames. It uses the same number of pyramid levels as ours and other parameters are set as suggested in [6]. The computation time of our algorithm depends on the motion complexity in the input data. It is typically close to that of BA.

Some sample CPU time values (in seconds on a PIII 900MHz PC running Linux) for our algorithm and BA are: 11.7 and 14.7 (Taxi), 29.5 and 27.4 (Flower Garden), 36.8 and 24.2 (Yosemite). Note that neither algorithm has been optimized for speed.

### A. Synthetic Data

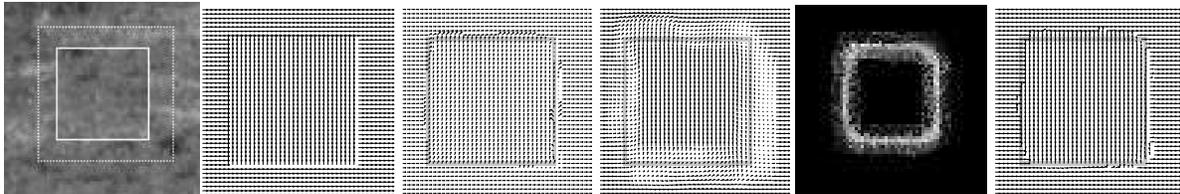


Fig. 2. TS sequence results (left to right: (a) to (f)). (a) middle frame with motion boundary highlighted. (b,c,d,f) show flows in the window outlined by the dotted line. (b) our estimate looks the same as the groundtruth. (c) BA estimate. (d) Initial LS estimate in Step I. (e) LMS trial set size in Step I (see Section III-A). (f) Step I final result. Step II result looks identical and is hence not shown separately.

Five data sets with flow groundtruth are used for quantitative evaluation. We use  $\mathbf{e}$ , the absolute  $u$  or  $v$  error, as the error measure. The motivations are (i) the  $u$  and  $v$  components, and positive and negative errors are treated symmetrically in optical flow estimation, and (ii) a 1-D measure is much easier to work with than a 2-D or higher-dimensional measure. We give the empirical cumulative distribution function (cdf) of  $e$  in addition to its mean  $\bar{e}$  (pixels). Better estimates should have cdf's closer to the ideal unit step function. To facilitate comparison with other techniques, we also report the popular average angular error  $e_{\angle}(\circ)$  [2][12]. It is important to point out that this error measure ought to be used with caution. One problem with it is that estimates having the same error magnitude may result in vastly different angular error values [1]. More fundamental problems with quantitative evaluation will be discussed at the end of this section.

The TS sequence (Fig. 2) gives an illustrative example. It is well textured and contains two occluding squares translating at exactly 1 pixel/frame. Therefore any formulation assuming brightness conservation and piecewise smoothness should fully recover the flow. Our method does achieve that (see Fig. 2(b) and curve “S3” in Fig. 3(a)). Now let us look at results from two gradient-based methods, BA and our Step I, which is a robust local gradient-based technique by itself [1], [26]. They both produce gross errors at motion boundaries due to gradient evaluation failure. In addition, Step I shows rounded corners

because the background motion becomes dominant there, while BA’s poor accuracy can be attributed to oversmoothing and slow convergence of its SOR procedure.

The Translating Tree (TT), Diverging Tree (DT) and Yosemite (YOS) sequences were obtained from Barron [2]. TT and DT [12] simulate translational camera motion with respect to a textured planar surface. TT’s motion is horizontal and DT’s is divergent. YOS’s motion is mostly divergent. The cloud part is excluded from evaluation [6], [1]. We use 2 levels of pyramid for TT and DT, and 3 levels for YOS. Error measures are given in the captions of Fig. 3. Many optical flow papers published after [2] report the  $e_{\perp}$  error on these data sets. Our TT result seems to be the best. The smallest DT error was given by Fleet and Jepson [12]:  $0.99^{\circ}$  at 61.0% density. Our error appears to be large because (i) much of our error comes from erroneous estimates at image borders, which are excluded from evaluation in [2] (in fact, if we exclude a 10-pixel wide margin, the errors drop to  $\bar{e} = 0.03$  and  $e_{\perp} = 1.64$  respectively) and (ii) our flow field is 100% dense while many other results in [2], [25], [26] are for selected fields. Some results on YOS are quoted in Table I. The first group assumes piecewise constant flow. The second groups assume stronger flow models such as local affine flow and constant flow in a considerable number of frames. These assumptions are appropriate for YOS and may lead to higher accuracy. The smallest error on YOS was reported from a simultaneous motion estimation and segmentation algorithm assuming an affine motion model in 9 frames [11]. Although using only a piecewise constant flow model and 3 frames, our method compares favorably with these techniques.

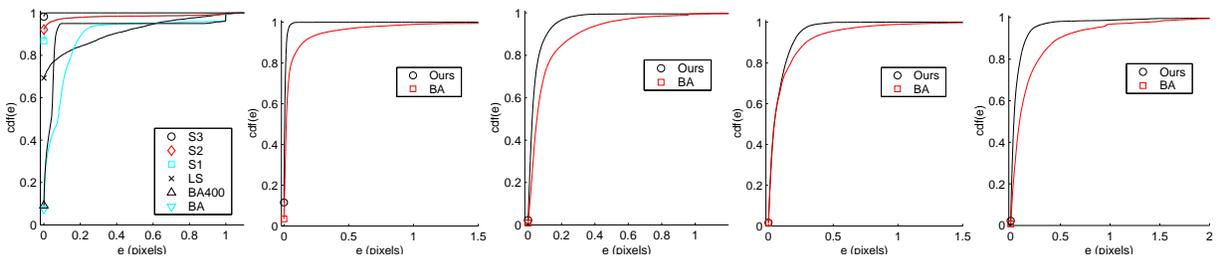


Fig. 3. Error  $e$  cdf curves for TS (a), TT (b), DT (c), YOS (d) and DTTT (e). ( $\bar{e}$ ,  $e_{\perp}$ ) values: TS: Ours ( $2.2e-4$ ,  $1.1e-2$ ), BA (0.12, 8.04); TT: Ours ( $9.8e-3$ , 0.05), BA (0.07, 2.60); DT: Ours (0.05, 2.60), BA (0.11, 6.36); YOS: Ours (0.08, 1.92), BA (0.12,  $2.71^2$ ); DTTT: Ours (0.08, 4.03), BA (0.20, 10.9).

<sup>2</sup>The value is slightly different from the one reported by Black and Anandan [5] most probably because their data are different from Barron’s and they calculated flow on the 14th frame instead of the 9th.

Technique	$e_{\perp}$ ( $^{\circ}$ )
Ye, Haralick and Shapiro (proposed)	1.92
Sim and Park [29]	4.13
Black and Anandan [5]	2.71
Szeliski and Coughlan [32]	2.45
Mémin and Pérez [21]	2.34
Black and Jepson [11]	2.29
Ju, Black and Jepson [17]	2.16
Bab-Hadiashar and Suter [1]	1.97
Farnebäck [11]	1.14

TABLE I

Comparison of various techniques on YOS (cloud part excluded) with Barron’s angular error

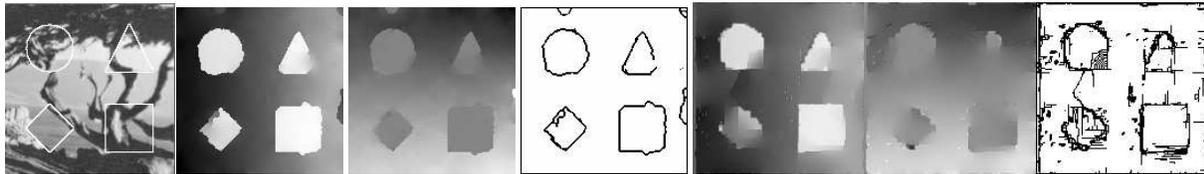


Fig. 4. DTTT sequence results (left to right): middle frame (motion boundaries highlighted), our horizontal, vertical flow and motion boundaries; BA horizontal, vertical flow and motion boundaries.

We examine the motion boundary performance of our method on the DTTT sequence (Fig. 4). It was generated from TT, DT and “cookie cutters”: image data inside the cookie cutters come from TT and those outside come from DT. We display the horizontal and vertical flow components as intensity images to show more details. Brighter pixels represent larger positive speeds. Two pyramid levels are used. Our result is smooth with crisp motion boundaries, and is better than BA both visually and quantitatively (see Fig. 4 and Fig. 3(e)). As a by-product of our method, motion boundaries are easily located in the smoothness error map.

Nonetheless, we do notice some gross errors near the boundaries, e.g., the right corner of the triangle is smoothed into the background. A closer look reveals that most of such errors happen in textureless regions, where even human viewers are unable to resolve the ambiguity. In such situations, the correctness of the “groundtruth” is questionable and so is the authority of quantitative evaluation based on it. Also noticeable is that our motion boundaries are not as smooth as one would like. This reflects the weakness of the simple optimization method in Step III and the need for more advanced techniques.

## B. Real Data

Results on four real image sequences: Taxi , Flower Garden, Traffic and Pepsi Can, are given in Fig. 5. Taxi [2] mainly contains three moving cars with the maximum speed about 3.0 pixels/frame. Motion in Flower Garden (from Black) is caused by camera translation and scene depth. The image speed of the front tree is as large as about 7 pixels/frame. Traffic (from Nagel) contains eleven moving vehicles with the maximum image speed at about 6 pixels/frame. Pepsi Can [5] has motion discontinuities due to camera translation and scene depths. The maximum image speed is about 2 pixels/frame. Two, three, three and two pyramid levels are used for the above four data sets respectively. Our technique shows consistent performance, yielding clear-cut motion boundaries and smooth flow within each layer.

## V. CONCLUSION

We have presented a matching-based global optimization approach to optical flow estimation with a practical solution technique. Contributions of our work include: (1) A simple and effective backward-forward matching scheme ameliorates the visibility problem at occlusions. (2) A global energy function balances the strength of brightness and smoothness errors according to local data variability. (3) Motion discontinuities can be reliably located at flow smoothness outliers. (4) A three-step graduated optimization strategy is developed to minimize the resultant energy. It takes advantage of gradient-based local regression, gradient-based global optimization and matching-based global optimization methods and reduces their limitations. (5) A deterministic algorithm is proposed to approximate the LMS robust estimator in the local gradient step. Experiments showed that it can be faster and more accurate than algorithms based on random sampling.

As an accurate and efficient low-level approach, the proposed method can facilitate higher-level motion analysis. Our flow estimates often have a layered look with motion boundaries located. They can aid motion segmentation, contour-based and layered representation [21], [35]. Model selection [34] is a crucial problem in automatic scene analysis, which is difficult because comparing a collection of models to the raw image data involves formidable computation. Our method can supply a higher ground for scene knowledge

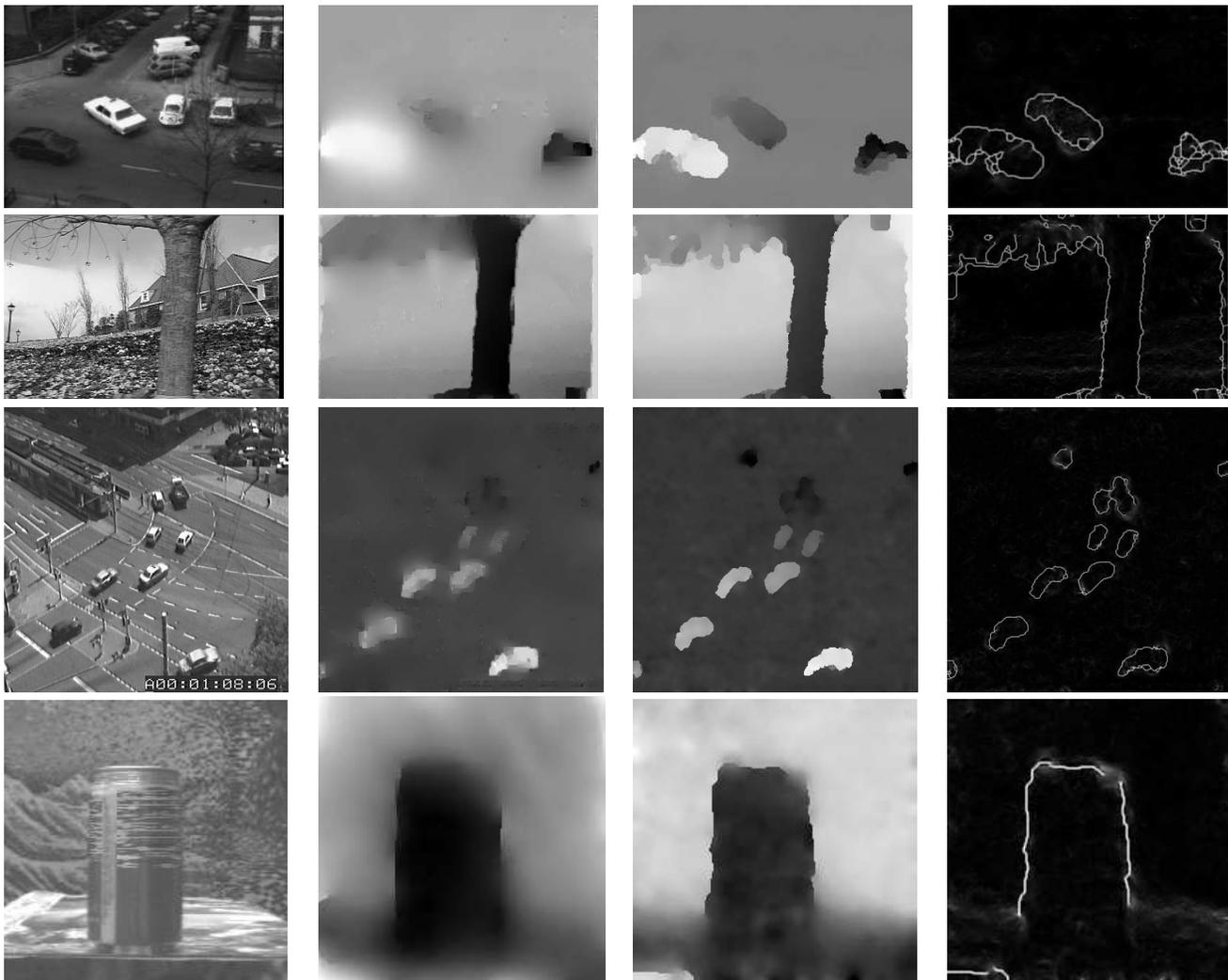


Fig. 5. Real data results. 1st column: middle frames. 2nd column: BA horizontal flow. 3rd column: our horizontal flow. 4th column: Our smoothness error ( $E_S$ ) map. 1st row: Taxi. 2nd row: Flower garden. 3rd row: Traffic. 4th row: Pepsi can.

learning. The backward-forward matching error, together with detected motion boundaries, can be used for occlusion reasoning and motion discontinuity-preserving image warping. Furthermore, some of our techniques can be useful to other low-level visual problems such as stereo matching, 3D surface reconstruction and image restoration.

To improve the formulation, the proper modeling of the three-frame matching error, the choice of robust estimator, and the learning of parameters should be investigated. To develop more global solutions to the optimization problem, methods such as graph cuts [9], full multigrid methods [23], Bayesian belief propagation (BBP) [36], and local minimization

methods alternative to SOR [23] are worth studying. Among all criteria that the global energy may arise from, the Bayesian approach is the most appealing in both theoretical and practical respects. The benefits of the Bayesian framework should be exploited and that may provide a graceful solution to two important problems: global optimization and uncertainty analysis [2], [37]. Finally, as we have pointed out in Section IV-A, existing evaluation methods are flawed and developing more convincing methods deserves great attention.

## REFERENCES

- [1] A. Bab-Hadiashar and D. Suter, "Robust optical flow estimation", *IJCV*, 29(1):59-77, 1998.
- [2] J. Barron, S. Beauchemin and D. Fleet, "Performance of optical flow techniques", *IJCV*, 12(1):43-77, 1994.
- [3] R. Battiti, E. Amaldi and C. Koch, "Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy", *IJCV*, 6(2):133-145, 1991.
- [4] J. Bergen, P. Anandan, K. Hanna and R. Hingorani, "Hierarchical model-based motion estimation", *Proc. ECCV*, 237-252, 1992.
- [5] M. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields", *CVIU*, 63(1):75-104, 1996.
- [6] M. Black, *Robust Incremental Optical Flow*, Ph.D. Dissertation, Yale Univ., 1992; Research Report YALEU/DCS/RR-923.
- [7] M. Black, G. Sapiro, D. Marimont and D. Heeger. "Robust anisotropic diffusion", *IEEE Trans. IP*, 7(3):421-432, 1998.
- [8] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA 1987.
- [9] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts", *PAMI*, 23(11):1-18, 2001.
- [10] R. Depommier and E. Dubois, "Motion estimation with detection of occlusion areas", *Proc. ICASSP*, 3:269-272, 1992.
- [11] G. Farnebäck, "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field", *Proc. ICCV*, 1:171-177, 2001.
- [12] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information", *IJCV*, 5:77-104, 1990.
- [13] B. Galvin, B. McCane, K. Novins, D. Mason and S. Mills, "Recovering motion fields: an evaluation of eight optical flow algorithms", *Proc. BMVC*, 1998.
- [14] R. Haralick and L. Shapiro, *Computer and Robot Vision*, Addison-Wesley publishing company, 1992.
- [15] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov random fields", *PAMI*, 15(12):1217-1232, 1993.
- [16] J. Jain and A. Jain. "Displacement measurement and its application in interframe image coding", *IEEE Trans. Communication*, 29(12):1799-1808, 1981.
- [17] S. Ju, M. Black and A. Jespon. "Skin and bones: multi-layer, locally affine, optical flow and regularization with transparency", *Proc. CVPR*, 307-314, 1996.

- [18] S. Kang, R. Szeliski, J. Chai, “Handling occlusions in dense multi-view stereo”, *Proc. CVPR*, 1:103-110, 2001.
- [19] J. Konrad and E. Dubois, “Bayesian estimation of motion vector fields”, *PAMI*, 14(9):910-927, 1992.
- [20] K. Lim, A. Das and M. Chong. “Estimation of occlusion and dense motion fields in a bidirectional bayesian framework”, *PAMI*, 24(5):712-718, 2002.
- [21] E. Mémín and P. Pérez, “Dense estimation and object-based segmentation of the optical flow with robust techniques”, *IEEE Trans IP*, 7(5):703-719, 1998.
- [22] D. Murray and B. Buxton, “Scene segmentation from visual motion using global optimization,” *PAMI*, 9(2):220-228, 1987.
- [23] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in C*, 2nd Ed., Cambridge Univ. Press, 1997.
- [24] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*, John Wiley and Sons, 1987.
- [25] J. Odobez and P. Bouthemy. “Robust multiresolution estimation of parametric motion models”, *Journal of Visual Comm. and Image Representation*, 6(4):348-365, 1995.
- [26] E. Ong and M. Spann, “Robust optical flow computation based on least-median-of-squares”, *IJCV*, 31(1):51-82, 1999.
- [27] B. Schunck, “Image flow segmentation and estimation by constraint line clustering”, *PAMI*, 11(10):1010-1027, 1989.
- [28] D. Shulman and J. Hervé, “Regularization of discontinuous flow fields”, *Proc. Workshop on Visual Motion*, 81-85, 1989.
- [29] D-G. Sim and R-H Park, “Robust reweighted MAP motion estimation”, *PAMI*, 20(4):353-365, 1998.
- [30] E. P. Simoncelli. *Distributed Analysis and Representation of Visual Motion*. Ph.D. dissertation, MIT, 1993.
- [31] S. Sun, D. Haynor and Y. Kim, “Motion estimation based on optical flow with adaptive gradients”, *Proc. ICIP*, 1:852-855, 2000.
- [32] R. Szeliski and J. Coughlan, “Hierarchical spline-based image registration”, *Proc. ICIP*, 194-201, 1994.
- [33] H. Tao, H. Sawhney and R. Kumar, “A global matching framework for stereo computation”, *Proc. ICCV*, 532-539, 2001.
- [34] P. Torr. “Geometric motion segmentation and model selection”, In J. Lasenby et al., editor, *Proc. The Royal Society of London*, 1321-1340, 1998.
- [35] P. Torr, R. Szeliski and P. Anandan. “An integrated bayesian approach to layer extraction from image sequences”. In *Proc. ICCV*, 983-990, 1998.
- [36] Y. Weiss and W. Freeman, “Correctness of belief propagation in gaussian graphical models of arbitrary topology”, *Neural Comp.*, 13(10):2173-200, 2001.
- [37] M. Ye and R. Haralick, “Image flow estimation using facet model and covariance propagation”, *Vision Interface: Real World Appl. of Computer Vision*, M. Cheriet and Y.H. Yang (eds.), World Sci., 209-241, 2000.
- [38] M. Ye and R. Haralick, “Optical flow from a least-trimmed squares based adaptive approach”, *Proc. ICPR*, 3:1052-1055, 2000.
- [39] M. Ye and R. Haralick, “Local gradient global matching piecewise smooth optical flow”, *Proc. CVPR*, 2:712-717, 2001.
- [40] M. Ye, R. Haralick and L. Shapiro. “Estimating optical flow using a global matching formulation and graduated optimization”. In *Proc. ICIP*, 2002.
- [41] M. Ye, *Robust Visual Motion Analysis: piecewise-smooth optical flow and motion-based detection and tracking*, Ph.D. Dissertation, Univ. of Washington, 2002.