International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 7, No. 6 (2014), pp. 273-284
http://dx.doi.org/10.14257/ijsip.2014.7.6.23

Large Displacement Optical Flow with Adaptive Feature Match

Xiaoping Zhang and Guoxin Li

Information Engineering School, Binzhou Vocational College, Binzhou, China
xiaopingzhang14@gmail.com

Abstract

This paper presents an accurate large displacement optical flow estimation approach by adaptively integrate the local feature match. Despite coarse-to-fine warping approach can handle large displacement optical flow; however, there is inherent limit for small object with large motion. And recently integration of feature match to the variational framework has relaxed the limit, but raised another problem of ambiguous feature matching due to poor feature descriptor. Address the aforementioned problem, in this paper we propose an adaptive integration approach of local feature match. The essence is that we only keep the robust feature and remove those unstable features (e.g, textureless region) to improve the flow accuracy. The adaptive approach substantially decreases the computational cost by remove uncertain features and leads to more robust performance by excluding unreliable matches. We qualitatively and quantitatively compared to the conventional flow methods on Middlebury and Sintel benchmark and show that we achieve more accurate and promising results.

Keywords: *Large Displacement Optical flow, Adaptive Feature Match, Energy Minimization, Confidence Measure*

1. Introduction

The estimation of accurate optical flow in image sequence is still an open and challenging problem in computer vision. Optical flow has broad applications, such as in motion estimation and video compression [16], object detection and tracking [5], robot navigation and visual odometry [14], micro air vehicles controlling [6], etc. Therefore, optical flow estimation is an extensive filed in computer vision community.

A critical but difficult problem for optical flow estimation, obviously, is constructing correspondences. Correspondences could be in totally different forms, *e.g.*, point correspondences, line correspondences, curve correspondences, even region correspondences. Sometimes, we can easily get some geometrical primitives from images, but sometimes is difficult. According to our best understanding, there are two major methodologies: "dense" approach, and "sparse" or "feature-based" approach. The dense approach tries to build correspondences pixel by pixel, while feature-based approach tries to associate different image features. These two ideas result in totally different taste of motion and structure analysis. The method proposed in this paper belong the "dense" approach by integrating both the sparse (feature level) and dense (pixel level) correspondence to handle the large displacement.

Despite most of the methods are using coarse-to-fine [2] strategies recently, situations where the coarse-to-fine heuristic does not work well in practice. Especially articulated motion as human motion (*e.g.*, small table tennis can move extremely fast, which violate the limit of structure detail of the small object). More recently, there has more intuitive research on the correct handling of large displacements. Based on the

ISSN: 2005-4254 IJSIP

optimization point of view, there are two different models: The first method is dense flow estimation with exhaustive search such as block matching [7]. However, it's known to provide poor results due to the ambiguous search. The second method is sparse flow estimation based on matching local image features such as SIFT [10], which suffers the weakness of spurious feature matching.

Address the aforementioned problems; in this paper we propose a novel adaptive approach to handle the large displacement optical flow estimation. In order to identify the regions that really need supplementary feature match, we initially compute the dense flow using a baseline approach (*e.g.*, Horn-Schunck framework), based on the estimated flow filed, we compute the warping error, and the locations with large warping error will be considered as candidate regions which need to integrate the local features. On the other hand, in order to avoid the unreliable local feature to be integrated, we compute the structure-tensor to remove the spurious matches around textureless region. Our adaptive approach can substantially improve the performance for large displacement.

2. Related Work

The optical flow literature is far too extensive and diverse to allow an exhaustive review here. Here we only focus on those literature related to large displacement optical flow. Based on the pioneer work of Lucas and Kanade [11] on iterative image registration, incremental coarse-to-fine strategies of Brox [2] or multi-scale focusing approaches [1] are recognized as the classical methods to address the problem of large displacement optical flow. The main drawback of such aforementioned hierarchical approaches is their intrinsic incapability to handle large displacements of small objects (*e.g.*, moving tennis): At coarse resolutions, where the motion becomes extremely too small to be estimated, such objects are no longer visible at all, which can never be recovered in the finer resolution.

In contrast, Liu [9] have recently proposed an approach named as SIFT flow that computes dense correspondence fields by using of richer descriptor, they compute a dense field of SIFT descriptors and then run an approximative discrete optimization via belief propagation from [13] on top of these descriptors. And Berg used a graph matching scheme as Integer Quadratic Programming (IQP) to combine descriptor matching with a regularity constraint. The drawbacks is computational costly and only generate sparse correspondence with an interpolation step with spline function. However, our goal is to estimate a dense flow field.

More recently, directly integrate initially computed point correspondences into the variational model [3] has been demonstrated as a state-of-art approach. Rather, while such an approach often strongly suffer from wrong feature matches, which result in a significant deterioration of the performance. Intuitively, for small displacements, it would be better to design an adaptive model that integrates such feature correspondences and variational scheme, i.e. the reliability and necessity of the supplementary matches to avoid uncertain and spurious matching, which is the key contribution of this paper.

3. Variational Model

Let $I_t(\mathbf{x}):(\Omega \in \mathrm{P}^2) \rightarrow \mathrm{P}^c$ be the image at time t, where $\mathbf{x} = (x, y)^T$ is the spatial location in the image domain Ω, and c represent of number of channels, *e.g.*, for gray scale

image we have $c=1$, and $c=3$ for color image. Our goal is to estimate the dense flow field $\mathbf{w}=(u,v)^T$ between two images at different time t. To this end, we formulate the energy functional as:

$$E(\mathbf{w}) = E_d(\mathbf{w}) + \alpha E_s(\mathbf{w}) + \beta E_m(\mathbf{w}) \tag{1}$$

where $E_d(\mathbf{w})$ is the data term to penalize the error from data constancy assumptions, $E_s(\mathbf{w})$ is the smoothness term to enforce spatial regularity of the flow field, and $E_m(\mathbf{w})$ is the matching term that flavors the estimate flow to be consistent with the sparse set of prior feature matches, and α, β are tuning parameters which can be adjusted empirically. We start to detail the aforementioned three energy terms as follows:

3.1. Data Term $E_d(\mathbf{w})$

The data term is based on common assumption of color constancy of corresponding pixels, can be represented by the energy:

$$E_d(\mathbf{w}) = \int_\Omega \Psi\left(| I_2(\mathbf{x}+\mathbf{w}) - I_1(\mathbf{x}) |^2\right) d\mathbf{x} \tag{2}$$

In order to achieve robustness against the deviation from the constancy assumption, the Charbonnier penalizer $\Psi(s^2) = \sqrt{s^2+\varepsilon^2}, \varepsilon = 0.001$ is employed, which allows to handle occlusions and other non-Gaussian deviations. To gain robustness against illumination changes, we also apply structure-texture decomposition [12, 15] to pre-process the input images.

3.2. Smoothness Term $E_s(\mathbf{w})$

Eqn.(2) enforce the matching of weakly descriptive feature (treat each pixel as a feature), just minimize the energy according to the brightness constancy result in many ambiguous solution (result from the aperture problem), which leads most of the flow estimation are not consistent with the true flow. Therefore, try to minimize distortion in flow and prefers solution with more smoothness, which is can be enforced by spatial regularity, that is, by penalizing the total variation of the flow field as:

$$E_s(\mathbf{w}) = \int_\Omega \Psi\left(| \nabla u(\mathbf{x}) |^2 + | \nabla v(\mathbf{x}) |^2\right) d\mathbf{x} \tag{3}$$

where $\Psi(\bullet)$ is the charbonnier penalizer as defined in Eqn.(2). This penalizer does act edge-preserving, which does not smooth the region across the boundary of the object in the image, so that the motion detail can be preserved.

3.3. Matching Term $E_m(\mathbf{w})$

In order to handle large displacement flow, we can use more descriptive features and neglect the regularity constraint. As Brox [3] indicated that point matching (*e.g.*, SIFT) can be done efficiently in a globally optimal manner by simple nearest neighbor. So that we can enforce the flow filed to be similar to the prior sparse vector field $\hat{\mathbf{w}}$ from point feature correspondences. Since $\hat{\mathbf{w}}$ is defined in a sparse manner rather than the whole image domain Ω, and we have to integrate it into the variational framework, we need to use an indicator function $\mathbf{1}_m(\mathbf{x})$ to indicate those sparse pixels that is necessarily needed, in other words,

$\mathbf{1}_m(\mathbf{x})$ is 1 if there is a feature descriptor available in image 1 at location x; otherwise it is 0. we also define another confidence function $\rho_m(\mathbf{x})$ to evaluate the reliability of the prior vector field $\hat{\mathbf{w}}$ by its matching score. Finally, considering outlier from wrong match, the same robust Charbonnier function Ψ as defined in Eqn.(2) is employed here to enforce the final estimate flow field $\mathbf{w}$ to be close to the prior descriptor match $\hat{\mathbf{w}}$. Hence, our matching term is formulated as:

$$E_m(\mathbf{w}) = \int_\omega \rho_m(\mathbf{x}) \cdot \mathbf{1}_m(\mathbf{x}) \cdot \Psi(|\mathbf{w}(\mathbf{x}) - \hat{\mathbf{w}}(\mathbf{x})|^2) d\mathbf{x} \quad (4)$$

Note that above term is very similar to the matching term in [3]. However, as [3] pointed out that large number of prior descriptor match actually result in more drawbacks than advantages due to the spurious matches, especially even worse along the the locations with small displacement, this is the main challenge of the contribution of this paper.

4. Adaptive Feature Matching

In this section we would like to explain how to choose the sparse feature matching which leads to sparse optical flow as $\hat{\mathbf{w}}$, As in Brox [3]'s method, they restrict the location of the features around the textured region of the image, which make sense since the texture- less region cannot generate meaningful feature and corresponding descriptor due to the lacking of variances of the photometric information. In addition to the texture restriction as in Brox [3]'s method, we propose a more elegant approach that employ the data term information to select the locations where the sparse feature is necessary and robust, which means that even though we can get some feature from the textured region, but at the same time the optical flow variational framework itself can estimate the flow accurately, if we push the sparse feature matching (sparse flow) into the variational framework, we may get worse optical flow, the reason is sparse flow somehow is uncertain and inaccurate due to the lacking of the regularization. Therefore, in order to achieve low computational cost and accurate flow estimation, we analyze both texture (structure) and data term (warping error) to adaptively choose the features and their corresponding matching.

In terms of analyzing data term for adaptive feature matching, firstly, we compute the initial flow through the pure variational framework, and those locations with large warping error will be considered as region of interest for generating the potential sparse feature. That is, we are looking for potential locations as $\{x : |I_1(x+\mathbf{v}) - I_2(x)| > T_{error}\}$. Therefore, by combing restrictions from both texture and data term, we can define the overall potential candidate locations as

$$x : \lambda_2(x) < \frac{1}{8}\bar{\lambda} \ \ and \ |I_1(x+\mathbf{v}) - I_2(x)| > T_{error} \quad (5)$$

where $\lambda_2(x)$ is the smaller eigen value of the structure of the image as $\nabla I(x) \nabla I^T(x)$, and $\bar{\lambda}$ is the average eigen value across the whole image, the ratio 1/8 is chosen as the same as in [3].

Now it's time to define reasonable descriptors to achieve robust matching, which should be invariant to scale, rotation, as well as viewpoint change. We investigate here two different methods: one based on Histogram of Gradient (HOG) descriptors [15], and the other one based on geometric blur (GB) [5]. The main requirement for a good descriptor matching method is that the descriptors are unique and informative enough to limit the number of false

matches. As [3] pointed out that HOG descriptors produce the fewest mismatches, whereas GB descriptors tend to capture more details, which lead to additional false matches. In this paper, HOG feature is used for all the experiments, we used the HOG descriptors for sparse matching. The merit of the HOG is that not only provide ice property to be more conservative in terms of false matches, but also the computational cost of the HOG descriptors is also the cheapest.

4.1. Confidence Function $\rho_m(x)$ Computation

Since we use two attributes (image texture/structure and warping error from the data term) to evaluate the reliability of a feature match for robust and efficient integration. First of all, we compute the first confidence function based on the feature match uniqueness as

$$\rho_m^1(x) = 1 - \frac{d_1(x)}{d_2(x)} \tag{6}$$

where $d_1(x)$ and $d_2(x)$ denote the distances of the best and the second best match, respectively. Similar to SIFT matching, take into account the second best match to give more weight to unique match to avoid ambiguous match, which means if $d_1(x)$ is much less than $d_2(x)$, then we treat this match as an unique match with large weight. Furthermore, in order to achieve the consistence of the confidence value, we constrain the value range of the weight, that is $\rho_m^1(x) \in [0,1]$, which is different to [3], in which they use $\frac{d_2(x) - d_1(x)}{d_1(x)}$ as the weight, where makes the confidence value sometimes is unexpected, even much larger one.

Secondly, we also estimate the another confidence weight by attribute of the warping error of the data term as

$$\rho_m^2(x) = 1 - e^{-\gamma \cdot (\frac{e(x)}{\bar{e}})^2} \tag{7}$$

where $e(x)$ is the warping error at location x, and $\bar{e} = mean(e(x))$ is the mean of the warping error, and γ is the coefficient to control the sharpness of the probability density distribution, which can be determined empirically. Our goal is to assign larger weight to the those matches that with large warping error, which means those locations are really need the feature matches to compensate their flow error, whereas if the warping error is small means the conventional variational framework is good enough to estimate the flow, if we push the integration of the feature match, it may makes the flow estimation even worse. Therefore, by employing the confidence weight from warping error, we can substantially decrease the computational cost, as well as improve the accuracy. Given the prior confidence weight from two independent measures, we need to fuse them together; we found that the most efficient fusion method for our problem is the product of the two measures, that is:

$$\rho_m(x) = \rho_m^1(x)\rho_m^2(x) \tag{8}$$

5. Minimization of the Energy Functional

The final step is to find the minimization of the energy functional as in Eqn. (1), which is non- convex and non-linear problem, we follow the concept from [2] that compute the optical

flow based on coarse-to-fine warping with a downsampling factor of 0.95, which makes the original energy as sequential convex energies. It is worth to note that the weight of prior matching term β should be adaptively decreased as the image goes to fine scale. We give them high impact at the beginning of the process (coarser scale), where the image resolution is very small and the feature correspondences dominate the photometric constancy. As the image resolution increases, the ratio between the number of feature correspondences and the increasing number of pixels in the image decreases and so does the impact of the feature correspondences as $\beta^i = \beta \cdot e^{-(c \cdot \frac{i}{i_{max}})^2}$, where i is the index of the iteration, and i_{max} is the max number of iterations for coarse-to-fine processing, which is related to the image size, and c is a constant to control the decrease speed, in this paper we found when $c = 2.5$ we can achieve nice result. In the continuous limit, this ratio goes to zero. We simulate this limit by running one last iteration with $\beta = 0$. The adaptive decreasing weight of prior feature matching at finer scales has helpful practical meaning. At coarse scales, feature matches leads the flow estimation close to large displacement which probably can't catch by the warping scheme. At finer scales, the prior feature matches is no longer needed, whereas warping scheme can provide more consistent flow estimation.

6. Experimental Results

In order to demonstrate the performance of our approach, we designed several experiments to show the significance of proposed adaptively integration method of the local feature matching. The parameter α, β were optimized to achieve the best performance for all the experiments. Typically, the range of α, β is $\alpha \in (0, 0.1), \beta \in (20, 150)$. In the first experiments, we would demonstrate the comparison to classical warping method, SIFT flow and LDOF flow as shown in Figure 1, all results indicate large displacements limit the performance of the classic warping methods. And SIFT flow shows salient discretization, quantization and discontinuities effect.

In the second experiment as shown in Figure 2, we use more images in Middlebury dataset which contains both small and large displacements to demonstrate the power of adaptive integration of local feature matches, in which the Middlebury training data set and the Average Angular Error (AAE) metric is explicitly evaluated as shown in Table 1, and also its corresponding standard deviation is evaluated to show the dispersion from the AAE. This experiment indicates ours is superior to conventional large displacement optical flow methods: SIFT flow [9] and LDOF method [3], in which the inaccuracy mainly suffers from the outlier of the wrong feature matches. Whereas we adaptively select those locations where the additional features are necessary for optical flow performance improvement, and get rid of the unnecessary matches which make the performance worse.

In the last experiment, we use the MPI Sintel [4] consists of simulated image sequences with large displacement (also includes motion blur, specular reflections, illumination, etc.). Since they can provide ground truth of flow estimation, so it's convenient to evaluate the performance quantitatively for large displacement. We have tested 13 sets of images contains many kinds of difficult motion even with serious occlusion. Our method outperforms both SIFT flow [9] and LDOF method [3] in terms of AAE metric, results and quantitative analysis are shown in Figure 3 and Table 2.

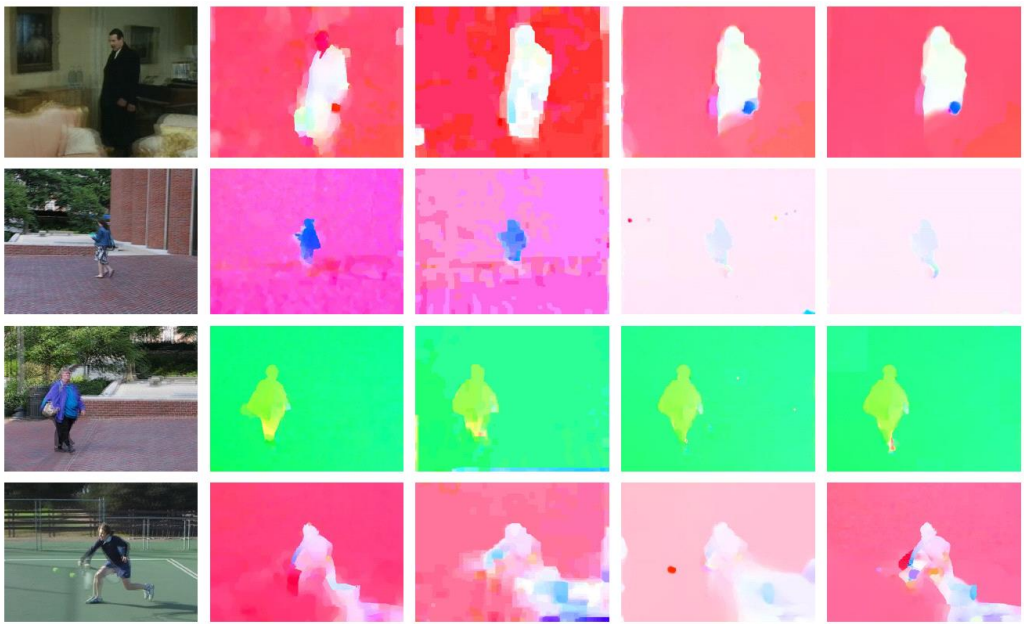
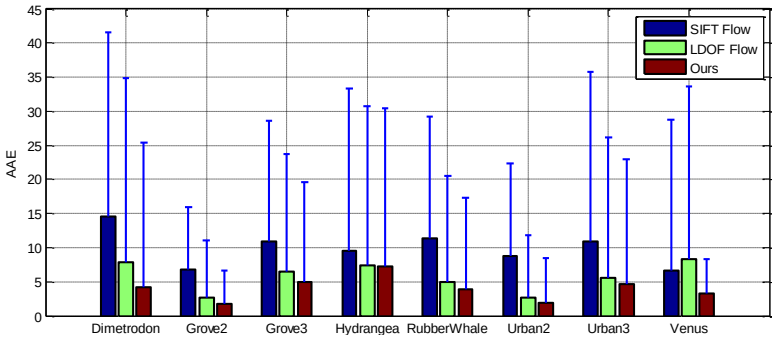


Figure 1. From Left to Right: 1) Superimposed input image; 2) Coarse-to-fine flow [2], 3) SIFT flow [8]; 4) LDOF [3]; 5) Result of ours. Image sequences from top to down are: maple11, people1, people2 and tennis respectively. Coarse-to-fine flow can't catch the fast moving small object as tennis, foot and hand, and SIFT/LDOF flow show noisy flow from outliers of spurious match, ours overcomes the above challenges and achieved better performance.

Table 1. Quantitative Comparison of Middlebury dataset as visualized in Figure 2. The blue line on the bottom figure indicates the standard deviation.

	IMAGES	SIFT FLOW	LDOF FLOW	OURS
1	Dimetrodon	14.59 ± 27.02	7.89 ± 26.93	**4.25 ± 21.20**
2	Grove2	6.75 ± 9.12	2.72 ± 8.36	**1.80 ± 4.85**
3	Grove3	10.87 ± 17.68	6.51 ± 17.19	**4.89 ± 14.71**
4	Hydrangea	9.49 ± 23.80	7.35 ± 23.40	**7.25 ± 23.13**
5	RubberWhale	11.37 ± 17.83	4.89 ± 15.58	**3.88 ± 13.47**
6	Urban2	8.75 ± 13.62	2.71 ± 9.12	**1.89 ± 6.61**
7	Urban3			
8	Venus			



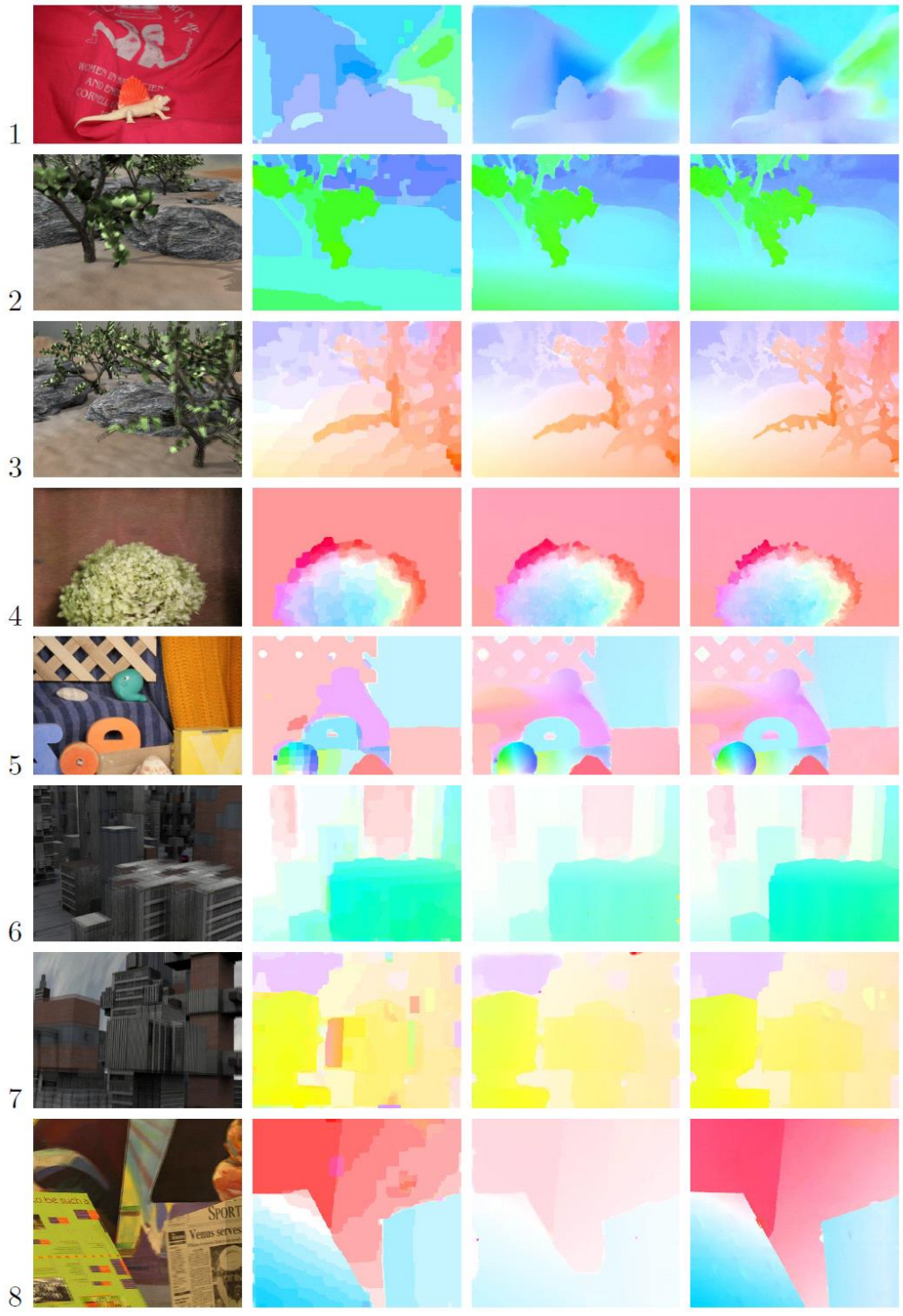


Figure 2. Qualitative results of Middlebury dataset, from left to right: 1) Superimposed input image; 2) SIFT flow [8]; 3) LDOF [3]; 4) Result of ours. Corresponding detail numbers and graphical illustration are shown in Table 1.

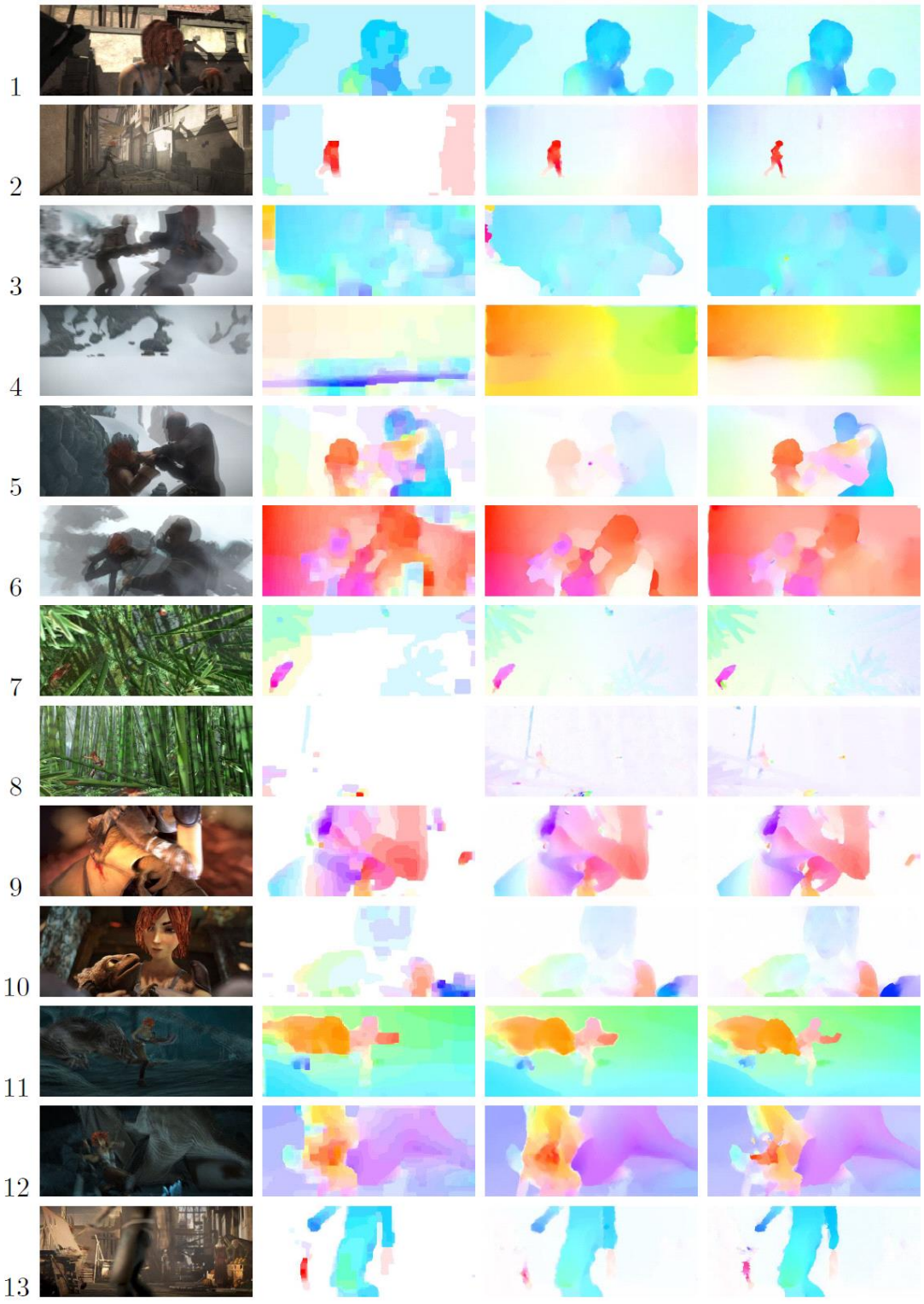
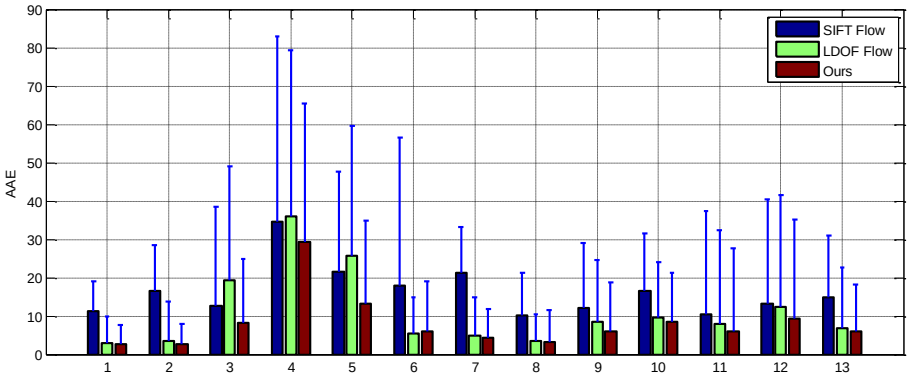


Figure 3. Qualitative results of Sintel dataset, from left to right: 1) Superimposed input image; 2) SIFT flow [8]; 3) LDOF [3]; 4) Result of ours. Corresponding detail numbers and graphical illustration are shown in Table 2.

Table 2. Quantitative comparison of Sintel dataset as visualized in Figure 3. The blue line on the bottom figure indicates the standard deviation.

	IMAGES	SIFT FLOW	LDOF FLOW	OURS
1	alley_1	11.31 ± 7.64	3.02 ± 6.80	**2.77 ± 4.96**
2	alley_2			
3	ambush_2	16.63 ± 11.85	3.43 ± 10.35	**2.77 ± 5.02**
4	ambush_4	12.76 ± 25.79	19.33 ± 29.61	**8.11 ± 16.67**
5	ambush_5			
6	ambush_6	34.66 ± 48.21	35.99 ± 43.40	**29.38 ± 35.94**
7	bamboo_1			
8	bamboo_2	21.55 ± 26.12	25.56 ± 33.92	**13.32 ± 21.60**
9	bandage_1	17.91 ± 38.61	6.09 ± 12.92	**5.29 ± 9.59**
10	bandage_2			
11	cave_2	21.28 ± 11.96	4.87 ± 9.95	**4.25 ± 7.66**
12	cave_4			
13	market_2	10.18 ± 11.13	3.57 ± 6.87	**3.18 ± 8.29**
		12.19 ± 16.71	8.55 ± 16.14	**6.02 ± 12.74**



7. Conclusion

This paper is trying to solve the large displacement optical flow by adaptively integrate local feature matching scheme. We have investigated the weakness of the conventional approaches, including the limit from classical warping scheme, outliers from spurious matching of SIFT flow and LDOF flow, finally we came up with our idea as only keep the robust feature and remove those unstable features(e.g, textureless region) to improve the accuracy of flow estimation, in which the robust feature is evaluated based on two attributes, analysis of image structure and warping error of initial flow. In a word, the adaptiveness of feature is implemented as a confidence measure to control how likely the feature should be integrated to the variational framework. Sufficient experiments on Middlebury and Sintel data set have qualitatively and quantitatively demonstrated the proposed method is superior to previous methods. Our future work will focus on improving the computation the confidence measure, and further how to use the variational framework to improve the feature matching.

References

[1] L. Alvarez, J. Weickert and J. S´anchez, "Reliable estimation of dense optical flow fields with large displacements", International Journal of Computer Vision, vol. 39, no. 1, (**2000**), pp. 41–56.
[2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping", In Computer Vision-ECCV 2004, (**2004**), pp. 25–36, Springer.
[3] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 3, (**2011**), pp. 500–513.
[4] D. J. Butler, J. Wulff, G. B. Stanley and M. J. Black, "A naturalistic open source movie for optical flow evaluation", In Computer Vision–ECCV 2012, (**2012**), pp. 611–625, Springer.
[5] E. Chen, Y. Xu, X. Yang and W. Zhang, "Quaternion based optical flow estimation for robust object tracking", Digital Signal Processing, vol. 23, no. 1, (**2013**), pp. 118–125.
[6] F. Expert and F. Ruffier, "Controlling docking, altitude and speed in a circular high-roofed tunnel thanks to the optic flow", In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, (**2012**), pp. 1125–1132.
[7] R. Li, B. Zeng and M. L. Liou, "A new three-step search algorithm for block motion estimation", Circuits and Systems for Video Technology, IEEE Transactions on, vol. 4, no. 4, (**1994**), pp. 438-442.
[8] C. Liu, J. Yuen and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 5, (**2011**), pp. 978-994.
[9] C. Liu, J. Yuen, A. Torralba, J. Sivic and W. T. Freeman, "Sift flow: dense correspondence across different scenes", In Computer Vision–ECCV 2008, (**2008**), pp. 28–42, Springer.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International journal of computer vision, vol. 60, no. 2, (**2004**), pp. 91–110.
[11] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", In IJCAI, vol. 81, (**1981**), pp. 674–679.
[12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms", Physica D: Nonlinear Phenomena, vol. 60, no. 1, (**1992**), pp. 259–268.
[13] A. Shekhovtsov, I. Kovtun and V. Hlav´a˘c, "Efficient mrf deformation model for non-rigid image matching", Computer Vision and Image Understanding, vol. 112, no. 1, (**2008**), pp. 91–99.
[14] F. Steinbruecker, J. Sturm, and D. Cremers, "Real-time visual odometry from dense rgb-d images", In Workshop on Live Dense Reconstruction with Moving Cameras at the Intl. Conf. on Computer Vision (ICCV), (**2011**).
[15] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, " An improved algorithm for TV-l 1 optical flow", In Statistical and Geometrical Approaches to Visual Motion Analysis, (**2009**), pp. 23–45, Springer.
[16] C. Zhang, Z. Chen, M. Li and K. Sun, "Direct method for motion estimation and structure reconstruction based on optical flow", Optical Engineering, vol. 51, no. 6, (**2012**), pp. 067004–067004–14.

