Vol 1 No 2 (2019) *e*-ISSN: 2682-8383

**Journal of Engineering Technology and Applied Physics**

# A Dense Optical Flow Field Estimation with Variational Refinement

Wei Yong Eng[1,*], Yang Lang Chang[2], Tien Sze Lim[1] and Voon Chet Koo[1]
[1] *Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia*
[2] *National Taipei University of Technology, Taipei, Taiwan.*
*engweiyong@gmail.com
https://doi.org/10.33093/jetap.2019.1.2.3

***Abstract* – Optical flow has long been a focus of research study in computer vision community. Researchers have established extensive work to solve the optical flow estimation. Among the published works, a notable work using variational energy minimization has been a baseline of optical flow estimation for a long time. Variational optical flow optimization solves an approximate global minimum in a well-defined non-linear Markov Energy formulation. It works by first linearizing the energy model and uses a numerical method specifically successive over-relaxation (SOR) method to solve the resulting linear model. An initialization scheme is required for optical flow field in this iterative optimization method. In the original work, a zero initialization is proposed and it works well on the various environments with photometric and geometric distortion. In this work, we have experimented with different flow field initialization scheme under various environment setting. We found out that variational refinement with a good initial flow estimate using state-of-art optical flow algorithms can further improve its accuracy performance.**

***Keywords—Optical flow, variational formulation, dense estimation, Markov Random Field, flow initialization***

## I. Introduction

Optical flow estimation is an ill-posed problem in computer vision. Optical flow estimation is based on brightness constancy constraint, also known as Optical Flow Constraint (OFC) as in Eq. (1).

$$I_t(x,y) = I_{t+1}(x+u, y+v) \tag{1}$$

An image pair are captured at two temporal step $t$ and $t+1$. A flow field $f(u,v)$ is estimated for all the pixel $(x,y) \in R^2$ in the image domain $R^2$. It is an underdetermined problem as it contains two unknowns $(u,v)$ in a single equation. Lucas *et al.* [1] made assumption that the flow is constant in a small local window. This tranform the formulation into a overdetermined problem and can be solved using a least-squared method. Their method made use of a local window and is also categorized under local method. Horn *et al.* [2] introduce a $L_2$ norm smoothness prior on the flow field by penalizing the non-smooth flow estimate. Their method made a prior distribution assumption on the flow field and favouring a smooth flow field. Their method assumpt a prior global distribution and also categorized under global method. A $L_1$ norm is introduced by [3] as $L_2$ penalises too strongly especially in the object boundary and occlusion region. A pyramid Lucas Kanade method [4] is proposed to estimate the optical flow in a coarse-to-fine framework. [4] first estimate the flow in a coarse image and further refine them in a finer image iteratively. A local method only consider a local window, whereas a global method take into account of a whole image region flow distribution. In the coarse resolution, a larger local window can be used to further improve its flow estimation accuracy and keeping its efficiency.

A well-defined Markov Random Field formulation is long established to solve for the flow field as in Eq. (1):

$$E(u,v) = E_{data} + E_{smooth} \tag{2}$$

This energy formulation for the flow field in two-dimensional image spatial domain $(u,v)$ consist of data term, $E_{data}$ and smoothness term $E_{smooth}$. The data term usually consists of brightness difference as well as gradient different. As the input image pair is usually sensitive to photometric variation, brightness constancy constrain is often violated. Hence, the data term is often coupled with a gradient constancy constraint as in Eq. (3). Gradient constancy constraint in Eq. (3) is useful to model translational movement. Brightness constancy constraint in Eq. (1) is more suitable for a complex motion model.

$$\nabla I_t(x,y) = \nabla I_{t+1}(x+u, y+v) \tag{3}$$

The smoothness term using a $L_2$ norm is in Eq. (4). It penalizes the variation in the flow smoothness in the image domain.



Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

$$E_{smooth}(u,v) = \iint_{\Omega} \nabla u^2 + \nabla v^2 \tag{4}$$

This energy formulation in Eq. (2) is non-linear in terms of $u$ and $v$. [5] have proposed to make a linearized approximation and solved the model using the iterative numerical method.

## II. RELATED WORKS

### *A. Local Method: Pixel-based and Featured-based Matching*

Various optical flow algorithms have been proposed in the past decades. Both local and global methods have extensive related works. In a local method, they can be categorized into sparse matching and dense matching depending on the flow estimation density. A local method is further split into pixel-based matching and feature-based matching. In a pixel-based matching, the data term usually consists of brightness constancy constraint in Eq. (1) and gradient constancy constraint in Eq. (3). In a feature-based matching, the difference in the feature descriptors such as SIFT [6], SURF [7], Daisy [8] are penalized. Feature-based matching algorithms are robust to scale and rotation distortion, thanks to use of histogram of gradient. The descriptor of a local region is summarised into a histogram and quantized using gradient strength and orientation. A pixel gradient in a local window may shift up to a few pixels and still having the same histogram. Feature-based matching initially consists of sparse feature matching at a few distinctive feature locations. Further algorithms interpolate these sparse matching into dense matching based on idea that nearby similar pixel appearance share similar flow. SiftFlow [9] estimate a dense flow field by accelerating the descriptor computation for every pixel in an image.

### *B. Local Method: Adaptive Weight Filter*

In flow estimation algorithm, a local method works by aggregating the cost function in Eq. (5) incurred by all pixels in a local filter as in the assumption that they share similar flow [6].

$$E_{data} = \sum\nolimits_{(x,y)\in k} w_{x,y}(I_t(x,y) - I_{t+1}(x+u, y+v)) \tag{5}$$

This assumption is made to solve the initial underdetermined problem in Eq. (1). This cost function is also known as data term, $E_{data}$ in Eq. (2) as they only consider the local window flow distribution. Initially, all the pixels in a filter, $k$ share the same weight in the cost function. However, the brightness constancy constraint is violated in object boundary region as there is also a flow discontinuity. Hence, a variety of edge-aware filter design are proposed to account for the flow discontinuity along the object boundary. Edge-aware filter assigns an adaptive weight, $w_{x,y}$ to each pixel $(x,y)$ depending on the color and spatial distance $(i,j)$ between each pixel with the center pixel $(x_c, y_c)$ as in Eq. (6). This is also the concept of Bilateral Filter [10] which calculate the filter weight based on two distinct color and spatial distance. The sigma, $\sigma_r$ and $\sigma_d$ in Eq. (6) denotes the smoothing parameter for color and spatial distance respectively. Hence, a higher weight is assigned for a pixel $(x,y)$ which is more similar to the center pixel $(x_c, y_c)$ in both color and spatial distance. Guided Filter [11] further accelerate the filter formulation in Eq. (6) to speed up the filter calculation time to a constant time $O(1)$, independent of kernel size.

$$w_{x,y} = \exp\left( -\frac{\left\| I_t(x_c, y_c) - I(x_c+i, y_c+j) \right\|^2}{2\sigma_r^2} - \frac{(i^2+j^2)}{2\sigma_d^2} \right) \tag{6}$$

### *C. Global Method: Variational Optical Flow*

A variational optical flow is often used to model a global energy model [2, 5]. A prior assumption of the flow field model is usually made in variational formulation. A piecewise-smooth and parametric motion model [12] are often assumed to be the estimated flow field distribution model. Other works [13] include smoothness term between estimated flow with the featured-based matching. It penalizing the variation between the estimated flow with the feature-based matching. It favours a flow field solution which is more coherent with the flow estimated by feature-based matching. Hence, [13] resembles the feature-based matching in their robustness to the geometrics distortion such as rotation and scale changes. [14] adopted a piecewise rigid regularization into the variational model.

### *D. Convolutional Neural Network for Optical Flow*

Recent success of deep learning in speech recognition, object detection and object recognition have led way to adopt the Convolutional Neural Network (CNN) into classical optical flow estimation. LeNet [15] first introducing the concept of CNN which is built by interleaving between convolution and max-pooling layer. AlexNet [16] won the 2012 ImagetNet Large-Scale Visual Recognition Challenge and the use of CNN concept in deep learning has increased tremendously since then. AlexNet contribution includes introducing the rectified linear units (ReLUs) nonlinear activation function as a ReLU is much faster than typical hyperbolic tangent function. AlexNet also used data augmentation techniques to increase the size of the training set for training a much larger network. AlexNet adopted graphics processing units (GPUs) which allows the training of a larger datasets and bigger images. VGG [17] built a deeper network by reducing the size of the filter. VGG further validates the use of a deeper network better capture the hierarchical feature representation through their classification accuracy. GoogleNet [18] introduces a Inception framework into the network based on the intuition of multiscale processing to further improve the classification accuracy. The Inception [18] works by concatenating the filter of multiple size and increasing the network width and depth for a better feature representation. ResNet [19] pushed the idea of a very deep network by proposing to train a residual network. It solved the gradient vanishing problem which usually occurs in the backpropagation optimization for training a very deep training. As the gradient backpropagate deeper in a network, its value vanishes and results in higher training and testing errors.

The success of deep learning in object detection and recognition tasks encourage its application in other domain and optical flow. The deep learning network automate the process of previously handcraft feature extraction and description. DeepFlow [20] is one of the earliest work on adopting deep network concept in optical flow estimation. DeepFlow interleaving between convolution and max-pooling layers in a multi-stage architecture. A recent work [21] adopted deep learning network and a strong prior in the motion field estimation. [21] has shown result outperform current state-of-art on the KITTI benchmark dataset [22].

FlowNet 2.0 [23] further demonstrate the concept of end-to-end learning of optical flow.

## III. Proposed Method

We propose an optical flow framework to further refine the original flow estimation. We adopt two state-of-art optical flow algorithms and refine them with three different strategies. DeepFlow [20] and Pyramid Lucas Kanade sparse-to-dense flow interpolation are two optical flow algorithms adopted in the experiment. Both algorithms have open source code in OpenCV library. Two optical flow algorithms are chosen as they estimate optical flow using different techniques. DeepFlow adopted a deep convolutional multi-stage architecture concept. Pyramid Lucas Kanade sparse-to-dense interpolation compute optical flow using typical sparse feature-based matching and it is further interpolated into dense flow. Two different optical flow refinement algorithms which are variational refinement and guided filter have tested. Variational refinement is chosen as it is typically adopted to minimize a global optical flow cost function. A typical variational flow adopted a zero initialization for solving its global cost function using iterative SOR algorithms. The flow candidates highly likely trap in a local minima using the zero initialization. Here, we found out that using a better initialization scheme leads to a better flow estimation. We used the flow estimation from DeepFlow and Pyramid Lucas Kanade sparse-to-dense algorithms as our initial flow estimate. Guided filtering is often adopted as the post-processing to the estimated flow. The flow field from natural image is usually piecewise smooth along object boundary. An edge-aware filter such as Guided Filter can further refine and adaptively smooth the optical flow field. We have tested the algorithms on the Mikolajczyk dataset [24]. The dataset [24] include the homography transformation matrix between all image pairs. The ground truth flow value can be computed using the given homography matrix. We evaluated the flow accuracy by marking the estimated flow is correct if the flow is within a 15 pixels radius from the ground truth flow.

## IV. Results

Figure 1 shows the visual flow estimation in colour code and its accuracy map for different estimation algorithms. Figure 2 shows the colour code used to illustrate the optical flow magnitude and orientation. The visual flow illustrates that the flow field with refinement scheme have a smoother distribution. Both variational refinement and Guided Filter enforce a piecewise smooth refinement implicitly and explicitly. The last two columns in Fig. 1 shows the error map before and after the refinement. The black pixel denote wrong estimate and white pixel denote correct estimate. Notice the error maps before refinement is noisy. Both refinement schemes remove most of the noisy pixels especially those surrounding by good pixels. Table I shows the flow estimation accuracy achieved with different flow algorithms before and after the refinement. Variational refinement shows improvement in both *tree* and *bike* image pairs for both DeepFlow and sparse-to-dense flow. Guided filter post-processing improves marginally as compared to variational refinement.

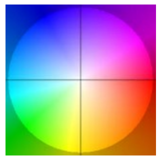

Fig. 2. The colour code for optical flow strength and orientation.

Table I. Showing accuracy for different algorithms.

| | ***DeepFlow with Variational Refinement*** | | ***DeepFlow with Guided Filter*** | | ***S2D with Variational Refinement*** | | ***S2D with Guided Filter*** | |
|---|---|---|---|---|---|---|---|---|
| | ***Bef.*** | ***Aft.*** | ***Bef.*** | ***Aft.*** | ***Bef.*** | ***Aft.*** | ***Bef.*** | ***Aft.*** |
| Tree | *0.9574* | *0.9641* | *0.9574* | *0.9590* | *0.8977* | *0.9203* | *0.8977* | *0.9034* |
| Bike | *0.9943* | *0.9953* | *0.9943* | *0.9947* | *0.9612* | *0.9748* | *0.9612* | *0.9514* |
| Boat | 0.9160 | 0.9081 | 0.9160 | 0.9156 | 0.5659 | 0.5380 | 0.5659 | 0.4959 |

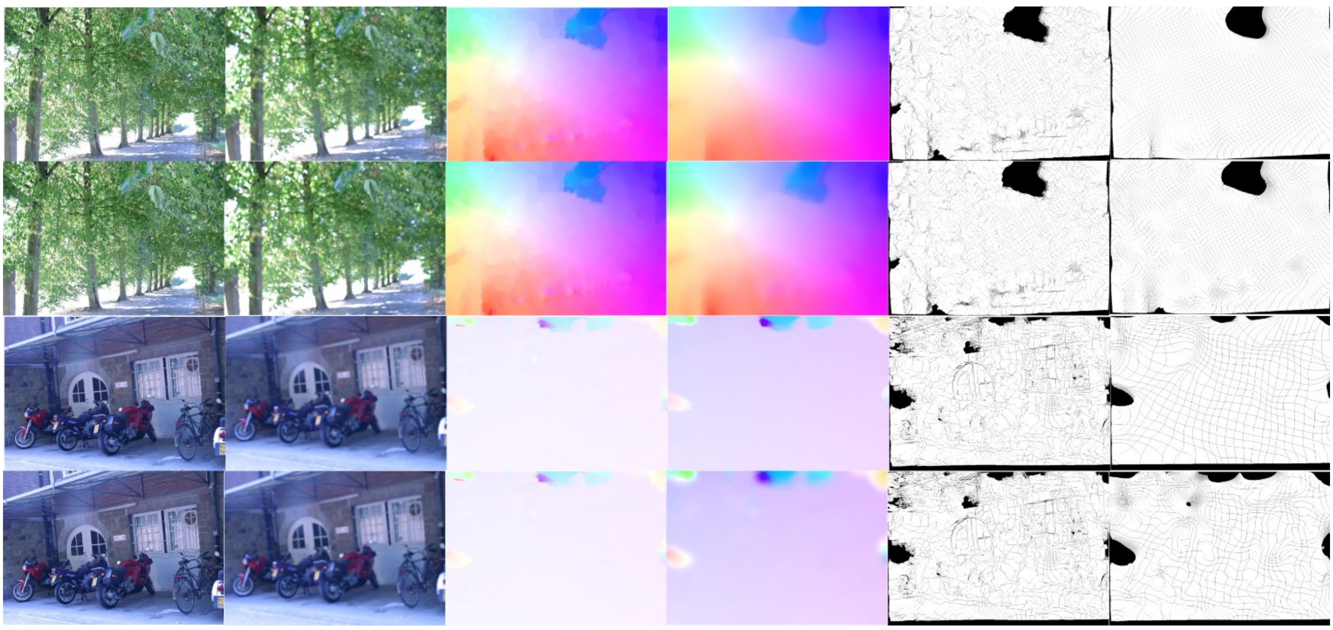

Fig. 1. The input image pair, the flow estimation in colour code before refinement, flow estimation after refinement, as well as error image before and after the refinement (from left to right). Figure illustrates different optical flow and refinement strategies: DeepFlow with variational refinement, DeepFlow with guided filter refinement, Pyramid Lucas Kanade Sparse-to-dense interpolation with variational refinement and Pyramid Lucas Kanade Sparse-to-dense interpolation with guided filter refinement (from top to bottom).

## V. CONCLUSION

In conclusion, a refinement scheme specifically variational refinement further improve the optical estimation accuracy. In variational refinement, an initial estimate is required for its iterative SOR solver. State-of-art optical flow estimates provide a better initial than zero movement. Further work includes adopts an end-to-end learning framework in optical flow task.

## REFERENCES

[1] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with An Application to Stereo Vision,", in *Proc. of the 7th Int. Joint Conf. on Artificial Intelligence*, vol. 2, pp. 674-679, 1981.

[2] B. K. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185-203, 1981.

[3] C. Zach, T. Pock and H. Bischof, "A Duality based Approach for Realtime TV-L 1 Optical Flow," in *Joint Pattern Recognition Symp.*, Springer, Berlin, pp. 214-223, 2007.

[4] J. Y. Bouguet, "Pyramidal Implementation of The Affine Lucas Kanade Feature Tracker Description of The Algorithm," *Intel Corporation*, vol. 5, no. 1-10, 4, 2001.

[5] T. Brox, A. Bruhn, N. Papenberg and J. Weickert, "High Accuracy Optical Flow Estimation based on A Theory for Warping," in *European Conf. on Computer Vision*, Springer, Berlin, pp. 25-36 2004.

[6] D. G. Lowe, "Object Recognition from Local Scale-invariant Features," in *Proc. of the Int. Conf. on Computer Vision*, vol. 99, no. 2, pp. 1150-1157, 1999.

[7] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.

[8] E. Tola, V. Lepetit and P. Fua, "Daisy: An Efficient Dense Descriptor Applied to Wide-baseline Stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815-830, 2009.

[9] C. Liu, J. Yuen and A. Torralba, "Sift flow: Dense Correspondence Across Scenes and Its Applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978-994, 2010.

[10] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Proc. of the Int. Conf. on Computer Vision,* vol. 98, no. 1, pp. 839-846, 1998.

[11] K. He, J. Sun and X. Tang, "Guided Image Filtering," in *European Conf. on Computer Vision*, pp. 1-14, Springer, Berlin, 2010.

[12] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-smooth Flow Fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75-104, 1996.

[13] T. Brox and J. Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500-513, 2010.

[14] C. Vogel, K. Schindler and S. Roth, "Piecewise Rigid Scene Flow," in *Proc. of the IEEE Int. Conf. on Computer Vision*, pp. 1377-1384, 2013.

[15] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[16] A. Krizhevsky, L. Sutskever and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," *Computer Vision and Pattern Recognition*, preprint arXiv:1409.1556, 2014.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov and A. Rabinovich, "Going Deeper with Convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[20] P. Weinzaepfel, J. Revaud, Z. Harchaoui and C. Schmid, "DeepFlow: Large Displacement Optical Flow with Deep Matching," in *Proc. of the IEEE Int. Conf. on Computer Vision,* pp. 1385-1392, 2013.

[21] W. C. Ma, S. Wang, R. Hu, Y. Xiong and R. Urtasun, "Deep Rigid Instance Scene Flow," *Computer Vision and Pattern Recognition,* preprint arXiv:1904.08913, 2019.

[22] M. Menze and A. Geiger, "Object Scene Flow for Autonomous Vehicles," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3061-3070, 2015.

[23] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy and T. Brox, "Flownet 2.0: Evolution of Optical Flow Estimation with Deep Networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2462-2470, 2017.

[24] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky and L. Van Gool, "A Comparison of Affine Region Detectors," *Int. J. of Computer Vision*, vol. 65, no. 1-2, pp. 43-72, 2005.