A multi-aperture optical flow estimation method for an artificial compound eye

Sidong Wu [a]   Gexiang Zhang [a,1]   Ferrante Neri [b]   Ming Zhu [c]   Tao Jiang [c] and Klaus-Dieter Kuhnert [d]

[a] *School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, P.R. China*

[b] *Centre for Computational Intelligence, School of Computer Science and Informatics, De Montfort University, The Gateway House, LE1 9BH, Leicester, United Kingdom*

[c] *Control Engineering College, Chengdu University of Information Technology, Chengdu 610225, P.R. China*

[d] *Institute of Real Time Learning Systems, University of Siegen, Siegen D-57076, Germany*

Abstract. An artificial compound eye (ACE) is a bio-inspired vision sensor which mimics a natural compound eye (typical of insects). This artificial eye is able to visualize large fields of the outside world through multi-aperture. Due to its functioning, the ACE is subject to optical flow, that is an apparent motion of the object visualized by the eye. This paper proposes a method to estimate the optical flow based on capturing multiple images (multi-aperture).

In this method, based on descriptors-based initial optical flows, a unified global energy function is presented to incorporate the information of multi-aperture and simultaneously recover the optical flows of multi-aperture. The energy function imposes a compound flow fields consistency assumption along with the brightness constancy and piecewise smoothness assumptions. This formula efficiently binds the flow field in time and space, and further enables view-consistent optical flow estimation. Experimental results on real and synthetic data demonstrate that the proposed method recovers view-consistent optical flows crossed multi-aperture and performs better than other optical flow methods on the multi-aperture images.

Keywords. optical flow, artificial compound eye, multi-aperture

1. Introduction

Modern robotics imposes the solution of complex tasks such as industrial operations [27,22,56], speech recognition [18], machine-human interaction [30] and navigation [47,21,23]. Within this context, machine vision tasks are very challenging and fundamental, see [16,26].

A powerful machine vision tool is the artificial compound eye, see [50], which allows a broad vision of the scene. The natural compound eyes widely exist in insects and crustaceans, and one of the mer-

its of natural compound eyes is its sensitivity to motion objects. The outstanding property is contributed by both its high temporal resolution and the structure with multiple ommatidia (singular: ommatidium, the basic unit of compound eyes that provides the picture element to brain). In general, the motion object is captured by a small number of ommatidia, and therefore, the pressure on a processor system can be very low and the processing speed can be very fast since only a small amount of data is needed to be processed. As a bionic system that hopes to inherit the merits of natural compound eyes, the artificial compound eye (ACE) is increasingly investigated [6,44,50]. However, few researches have addressed the practical use of ACE cameras [50], such as motion estimation. Although lots of literatures based on the 2-dimensional motion estimation method called *optical flow* [17] and the 3-dimensional motion estimation method called *scene*

flow [45] have been reported in recent years, the motion estimation method of ACE has scarcely been studied. There are mainly two limitations about the optical flow estimation of artificial compound eye. On one hand, ACEs have multiple apertures, while most of which contains only one pixel in each aperture. The image resolution obtained by ACE is limited by the number of apertures [12], and the interval between apertures are much larger than the pixel interval in traditional camera. The relatively low resolution and large aperture interval heavily restricts the accuracy of optical flow result, in addition, it also limits the use of optical flow information to other high level computer vision tasks, such as semantic segmentation, object recognition *et al.*, due to the sparse sampling of the scenario. This problem can be addressed by integrating a pixel array in each aperture. Therefore, an aperture of ACE captures a small image, and the image of ACE is composed of an image array. Then the image resolution of ACE with the same aperture is greatly improved. However, the image array captured by ACE incurs the second problem, since the ACE often has a small size [50], the image of an aperture only captures a small region of the scenario, therefore, estimating the optical flow on each aperture will suffer from the aperture problem [9]. In this paper, we mainly address the second problem of optical flow in ACEs based on a specific artificial compound eye (electronic cluster eye, eCley).

The eCley, inspired by a wasp parasite called *Xenos Peckii*, is widely researched to get super-resolution of imagery [6]. In this work, the eCley has a lenslet array of 17×13 apertures, and the size of the optics module is $6.8mm \times 5.2mm \times 1.4mm$ (Length $\times$ Width $\times$ Height). This means that the height of the optics module of eCley is about one third of a single-aperture optics module with the similar resolution. With such a compact size, the eCley can be widely applied to the fields such as bio-medical imaging, document analysis, fingerprint identification, micro-robot navigation. However, since the eCley was proposed several years ago, only a few researchers have adopted this kind of camera in the area of computer vision. This is partially because the optical axes between adjacent apertures of eCley have an small offset to obtain a larger field of view (FOV), which leads to the oblique incidence of the marginal aperture. As shown

in Fig. 1, a chessboard captured by eCley is parallel to the eCley, and the image captured by the central aperture (image in the cyan box) has less distortion, while the images captured by the marginal apertures suffer from the oblique distortion (images in the red and yellow box). The oblique distortion will lead to a complex process for computer vision tasks [52]. Therefore, standard vision algorithms cannot transfer to eCley directly to yield accurate results. Besides, the biggest challenge for using the eCley is that the small size and small FOV of aperture image resulting in little context to support the corresponding field inferencing.
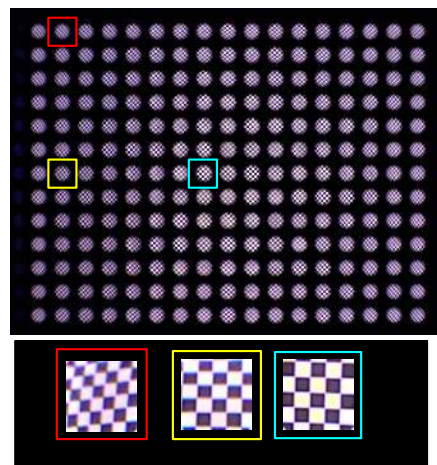

Figure 1. The captured image of eCley. Top, the image array of eCley; Bottom, the cropped image of the corresponding color box area.

The image caputered by eCley is an image array, and adjacent aperture images are partially overlapped. The multi-aperture structure will cause stereo problem. The depth estimation method for eCley can refer to [52,51]. Due to the short focal length and baseline of eCley, the stereo nature of aperture images is negligible when the objects locate larger than 86mm away from the eCley. Therefore, this paper assumes that the moving objects are far enough, and the motion estimation problem with eCley camera becomes multi-aperture optical flow problem. Howerver, traditional optical flow methods are based on single aperture image sequences. The small FOV of aperture image will lead to the inaccurate optical flow result, especially for the border of image, and inconsistent optical

flow between adjacent aperture images. For example, four adjacent aperture images at time t (left) and time $t + 1$ (right) are shown in Fig. 2(b), and the color coding in [2] for the visualization of optical flow is shown in Fig. 2(a), and the optical flow of each aperture image sequences is estimated independently based on the DeepFlow method [33,49], which are shown in the right of Fig. 2(c). The optical flow of bottom-right aperture image based on DeepFlow method can not be accurately estimated due to the aperture problem. Besides, the four aperture images capture the same motion boundary, but the optical flows at the motion boundary are less consistent.
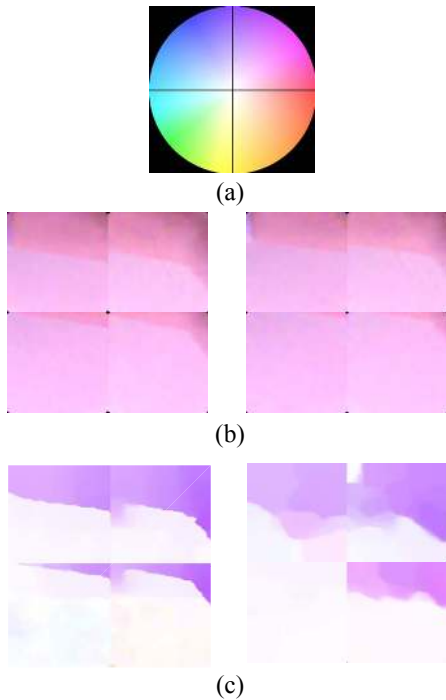

(a)

(b)

(c)

Figure 2. Optical flow results of eCley aperture images. (a) the color coding of the flow vectors; (b) the aperture images sequences, left: image at time t, right: image at time t+1; (c) optical flow. left: the proposed method, right: DeepFlow [33,49].

To address the optical flow problem of eCley, we first introduce an oblique distortion rectification method to rectify the image array and estimate the parallax between adjacent aperture images. Then, based on the rectified image array sequences, the Deep-Matching method [33] is used to obtained the initial

optical flow. While the initial optical flow estimation of an aperture image based on DeepMatching method uses only the contexts of corresponding aperture image sequences. The refinement method based on traditional variational method leads to less accurate results due to the aperture problem, see the right image of Fig. 2(c). Therefore, in order to obtain accurate optical flow, a multi-aperture optical flow method is proposed to refine the optical flow fields. In this method, a unified variational framework for all images is proposed to simultaneously refine the optical flows of all images. In contrast to traditional variational optical flow methods using brightness constancy and piecewise smoothness assumptions, our model combines the information of multi-aperture and further imposes a compound flow field consistency assumption that combines the corresponding flow fields of adjacent images. Consequently, the multi-aperture images information is used to define a global energy function. By solving the associated Euler-Lagrange equations and following the coarse-to-fine warping strategy, the minimization of the nonconvex energy function is obtained.

Experiments are based on synthetic and real data, and the results demonstrate that our method achieves more accurate and consistent performances in comparison with those optical flow methods based on single aperture.

The rest of the paper is organized as follows. The related work is presented in Section 2. Section 3 describes our multi-aperture optical flow method. Experiments are presented in Section 4. Finally, conclusions are followed in Section 5.

2. Related work

Since ACEs are bio-inspired instruments, the bio-inspired model of optical flow computation and the use of optical flow signal to control the movement are relatively important aspects for mimicking the function of their biological prototype. Franceschini *et al.* [13] researched the visual behavior and neural networks of the airborne insects with compound eyes, and designed an artificial compound eye with an array of elementary motion detectors (EMD) to avoid obstacles. Inspired by insects, Zufferey *et al.* [59] proposed

a simple control strategy based on optical flow to avoid obstacles. Pericet-Camara *et al.* [31] designed a lightweight artificial elementary eye to extract the local optical flow fields. Bračun *et al.* [3] used the functional subnetwork approach to model the neural system of ACE for visual motion recognition. Although above researches focused on the motion detections of ACEs and further handling method based on motion information, a basic unit of the ACE only contains an EMD, which is quite different from the eCley used in this paper. The basic imaging unit (the aperture) of eCley contains multiple pixels that will form an image of scenario in the FOV. Therefore, an aperture image of eCley is able to capture much more information (such as color and environmental context) than the EMD in above ACEs, and is a local image captured the scenario. The most related camera to eCley is the light field (LF) camera, which puts a microlenses array behind the main lens. Due to the ability to capture much richer information than traditional camera, LF camera is widely researched in the field of computer vision, such as depth estimation [58,7,19], motion deblurring [41,25], material recognition [46], super-resolution imaging [11], scene flow estimation [24]. Since the captured light field information mainly comes from the microlenses structure, the eCley can also be considered as a kind of LF camera that do not have the main lens. The widely used LF cameras capture multiple aperture images with small offset. However, the adjacent aperture images of eCley have much larger offset. For a pixel in an aperture image, the large offset leads to the corresponding pixels can only be found in several neighboring aperture images. Therefore, some methods based on LF image [20], such as epipolar plane image (EPI) method, angular patches method, can not yield good performance.

The bio-inspired model for the velocity estimation of local image has been widely researched [15, 40,38,9]. These models mimic the neural models existed in cortical areas called primary visual cortex (V1) and medio-temporal area (MT) by using a two layers feed-forward model. The V1 layers uses a bank of oriented filters (i.e. the Gabor filters or Gaussian derivatives [9]) to obtain the responses of the velocity and the direction of velocity. Then, the responses of V1 layer are pooled and transfered to the MT layers, fi-

nally, the final responses are acquired by transforming the pooled responses through a non-linear function. To compute the optical flow of large moved object, the multi-scale approach is adopted. Although the two layers model is likely to the model of the visual system, the obtained optical flow is worse than the results by using modern computational method (i.e. the variational method). The possible reasons may be due to the fact that the visual system has much more complicated procedure and can use much larger range of scenario to infer the contextual information. On the contrary, the convolutional neural network (CNN) using much more layers to extract the local features and global abstract descriptors greatly improved the accuracy of optical flow [10,14,1,8]. While the CNN-based method for optical flow needs numerous training data to learn the hyper-parameters, and the numerous training data of ACE images are currently unobtainable.

Except for the CNN-based method, the variational method is another widely researched optical flow method. After 35 years' development, current variational methods achieve more accurate results. Brox *et al.* derived a variational formulation to improve the optical flow result by imposing a coarse-to-fine warping technique for large displacements and using two nested fixed point iteration strategy to optimize the global energy function [4,29]. Zach *et al.* improved the Horn-Schunck model by imposing a robust L^1 data term and total variation (TV) regularization [55]. The energy function is minimized by alternating optimization strategy, see also [39,34,35,36,37]. Subsequently, Wedel *et al.* further improved the $TV - L^1$ optical flow algorithm by performing a structure-texture decomposition of the images and integrating a median filter into the numerical scheme [48]. Sun *et al.* gave a thorough analysis of what has made recent advances possible [42,43]. A systemic analysis of the energy function, the optimization, and modern implementation practices was drawn. Consequently, a method called "*Classic + NL*" was proposed to further improve the accuracy. The above methods were performed based on a multi-scale variational framework for large displacements. Although the multi-scale strategy is the key step for variational framework to estimate the large displacement, the fine motion structure with a large displacement cannot always be correctly estimated. Recently, several algorithms were

proposed to handle this problem by going beyond the variational framework and incorporating additional feature corresponding information [5,53,49,32]. Unfortunately, due to the small size and small FOV of aperture image, estimating the optical flow of aperture image independently often leads to inaccurate results caused by aperture problem. Since an object near the border of an aperture image will be close to the image center of one of the neighboring aperture, incorporating the neighbor aperture image to address the aperture problem can be considered to be a possible way to improve the accuracy of optical flow results.

3. Multi-aperture optical flow method

Our goal is to simultaneously recover the optical flows of an image array captured by a multi-aperture camera - eCley. However, as shown in Fig. 1, the oblique incidence of marginal aperture leads to oblique distortion, which will result in the parallaxes of adjacent aperture images inconsistent. Therefore, in this Section, we first introduce a method to rectify the image array and estimate the parallaxes in Section 3.1. Then the optical flow method is given in Section 3.2.

3.1. Preprocessing of eCley image array

To address the distortion caused by oblique incidence, this section introduces the rectification method for eCley to rectify all the aperture images simultaneously. Since the eCley is an array of optical channels, it can be considered as multi-apertures camera and the incident angle of marginal aperture is oblique. The parameters of the marginal apertures obtained by traditional calibration method [57] will not be the accurate parameters due to the oblique incidence. Therefore, based on the fixed system parameters and incident angles, we propose to deduce the rectification parameters of other apertures from the central aperture's. The rest of this section gives the detail of our rectification method.

As the same as calibration method, we use a chessboard as reference pattern. The chessboard is parallel to the eCley lens and put far enough to ensure that the disparity is negligible. The captured chessboard image is shown in Fig. 1, and each aperture image captures a part of the chessboard. Since the aperture image is a small size and small FOV image and the optical axis of central aperture is perpendicular to the imaging sensor plane, the image of central aperture is considered as undistorted image. Therefore, the size of black and white block of the central aperture image can be used as a reference for the rest of aperture to rectify those images.

To obtain the reference block size in the central aperture image, the corners of each block need to be extracted. In this paper, a canny edge detector is applied to obtain the edges, and then a Least Square (LS) fitting is performed over these edges to obtain multiple straight lines. Then the corner points can be extracted from the intersection points of these straight lines. Consequently, the average size of these white and black blocks is used as the reference block size. Since all of the black and white blocks in the chessboard we used have the same size, the sizes of blocks in the undistorted images will be the same as the blocks' in the central aperture image. Therefore, for an aperture image, the block size and corner points can be obtained by using the same procedure as the central aperture's. Assume that the corner point closest to the aperture image center is chosen as a reference point, then each corner point can be assigned an undistorted coordinate according to the block size of central aperture image. Base on the original coordinates and the undistorted coordinates of corner points, the Random Sample Consensus (RANSAC) method can be used to estimate the rectification parameters. Consequently, the aperture image can be rectified by using those transformation parameters.

Due to the unfixed reference point, the parallax between rectified adjacent images is required to be re-estimated. For a rectified aperture image, the undistorted corner points can be used to further estimate the translation parameters. Since there is an inherent parallax between adjacent aperture images (see [28] for more detail about eCley), a corner point has a roughly corresponding point at the neighboring image, then the exact corresponding point is searched around the corresponding point. Afterward, the translation parameters are obtained by averaging offsets of these corner points. For each aperture image (except the marginal aperture), 4-neighboring aperture images are used to obtain 4-pair translation parameters, which

can be considered as the inherent parallax between the rectified adjacent images.

Fig. 3 shows the aperture images and merged image of the chessboard before and after rectification. The oblique distortion is largely reduced in the rectified aperture images, and the merged image based on the re-estimated parallax has much less aliasing than the merged image with original aperture images. While the rectification needs to be an interpolation processes, the marginal aperture images after rectification is blurrier than the central image, and the merged image with rectified aperture images is blurry at the image border.
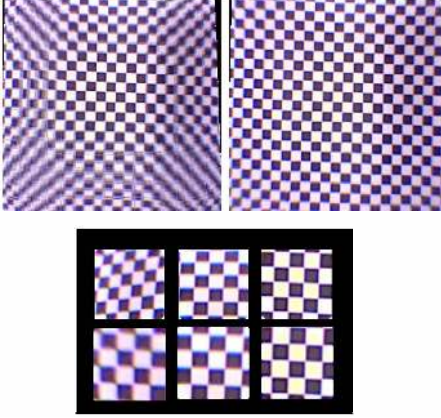

Figure 3. Top-left: the merged image with original aperture images; Top-right: the merged image with rectified aperture images. Bottom: the comparison of three aperture images, the first row: original aperture images of Fig. 1; the second row: corresponding rectified aperture images.

3.2. Optical flow method

3.2.1. System parameters and notations

Consider an array of $M \times N$ apertures, $\{A_i\}$, where i is a two-dimensional vector that indicates the position of the aperture. Let $I_i(x,y,t) : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$, be the rectified image sequence in aperture A_i. Let $\mathbf{x} := (x,y,t)^\top$ be the coordinate of pixel at time t, and $\mathbf{w} := (u,v,1)^\top$ be the displacement vector between an image at time t and another image at time $t + 1$. The optical flow $(u,v)^\top$ is the objective to estimate.

3.2.2. Optical flow initialization

In order to handle large displacement in optical flow, lots of literatures blend a matching approach with a variational method [5,53,49,32], and largely boost the optical flow performance. This paper also takes advantage of the matching approach to obtain the initial optical flow. Since the *DeepMatching* method is a quasi-dense matching method tailored for optical flow problem, for each aperture image sequence $I_i(x,y,t)$ and $I_i(x,y,t+1)$, the DeepMatching method as [49] is adopted to obtain the initial optical flow $\mathbf{w}_i^{init}$ of aperture image. Due to the small FOV of the aperture, the captured image frequently shows a textureless or little texture pattern, which leads to the initial optical flow less accuracy. To refine the initial optical flow, we propose a variational optical flow model to incorporate the adjacent aperture optical flow to improve the result.

3.2.3. The variational optical flow model

Based on the variational formulation of optical flow, the total energy function we aim to minimize is

$$E(u,v) = \sum_i E_i(u,v), \tag{1}$$

where $E_i(u,v)$ is the energy function of aperture A_i, and is a weighted sum of three terms,

$$E_i(u,v) = E_{Data}^i + \alpha E_{Smooth}^i + \beta E_{Compound}^i, \tag{2}$$

where E_{Data}^i is data term, E_{Smooth}^i is smooth term, $E_{Compound}^i$ is compound term that encourages the flow of adjacent images to be similar. α and β are weight parameters accordingly. If $\beta = 0$, $E_i(u,v)$ becomes the energy function of traditional optical flow model.

E_{Data}^i imposes the brightness constancy and gradient constancy assumption, and is obtained by

$$E_{Data}^i = \int_\Omega \Psi(|I_i(\mathbf{x}+\mathbf{w}) - I_i(\mathbf{x})|^2 + \gamma |\nabla I_i(\mathbf{x}+\mathbf{w}) - \nabla I_i(\mathbf{x})|^2) \mathbf{dx}, \tag{3}$$

where $\nabla = (\partial_x, \partial_y)^\top$ denotes the spatial gradient. The function Ψ is a penalty function. As illustrated in

[42,43], among three penalty functions, the quadratic penalty $\Psi(x^2) = x^2$, the Charbonnier penalty $\Psi(x^2) = \sqrt{x^2 + \varepsilon^2}$ and the Lorentzian penalty $\Psi(x^2) = \log(1 + \frac{x^2}{2\delta^2})$, the Charbonnier penalty performs the best. So the Charbonnier penalty is chosen in this paper and ε is a small positive constant to make sure the function is convex.

The smooth term E_{Smooth}^i imposes a piecewise smoothness assumption to deal with the ambiguities of low texture regions and give a smooth flow field. It is obtained by penalizing the total variation of flow field,

$$E_{Smooth}^i = \int_\Omega \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2)\mathbf{dx}, \qquad (4)$$

where $\nabla_3 = (\partial_x, \partial_y, \partial_t)^\top$. If only two sequence images are available, it will be replaced by the spatial gradient ∇ due to $\partial_t = 1$.

By using a modern optimization method, traditional optical flow model with the two terms mentioned above can yield an accurate result [4]. While in our eCley images, the image captured by an aperture only spans a small FOV and has a small size, which may lead to less accurate results due to the lack of texture information in aperture image. As shown in Fig. 2(c), the optical flows of adjacent aperture images are not consistent by using DeepFlow method, especially in the motion boundary. Fortunately, adjacent images are partially overlapping and can provide spatial correspondence information to improve the consistence optical flow results. As for the corresponding areas of the border area of aperture image in adjacent images, at least one of them will be close to the center of corresponding image. Therefore, the moving object in corresponding area will have less chance to move out of the aperture image and more likely to have an accurate flow boundary. Thus, this paper further imposes a compound term $E_{Compound}^i$ to improve the optical flow. The left image of Fig. 2(c) shows that the proposed compound term gives more consistent and accurate optical flow result. the compound term $E_{Compound}^i$ assumes that the corresponding pixels of adjacent images have the same flow field, and is expressed as

$$E_{Compound}^i = \int_\Omega \frac{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot \Psi(|\mathbf{w} - \mathbf{w}_j|^2)}{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j)}\mathbf{dx},$$

$$(5)$$

where $Ne(i)$ indicates the 4-nearest neighboring apertures of aperture A_i; $\mathbf{x}_j$ indicates the corresponding pixel of $\mathbf{x}$ in aperture A_j, and $\mathbf{w}_j$ is its displacement vector. $\delta(\mathbf{x}_j)$ is a Dirac function which is 1 if $\mathbf{x}_j$ exists, and is 0 otherwise. $\Psi(|\mathbf{w} - \mathbf{w}_j|^2)$ measures the difference between the flow of pixel $\mathbf{x}$ and the corresponding pixel $\mathbf{x}_j$.

Since pixel $\mathbf{x}$ in aperture A_i generally has at least two corresponding pixels among $Ne(i)$. The flow vector of $\mathbf{x}$ tends to be the mean of displacement vectors of corresponding pixels by using an uniform weight to all corresponding pixels. If all corresponding pixels use the same weight, the error flow vector of corresponding pixels located at the border of adjacent image will force the flow vector of pixel $\mathbf{x}$ to tend to the error flow vector. To reduce the effect of corresponding border pixels, a weight function $g(\mathbf{x}_j)$ is added. The pixel closer to the border of image will have a smaller weight, and the weight of pixel $\mathbf{x}_j$ only relates to its position in the corresponding image. Therefore, we use a Gaussian function to model the relative importance of the position of pixel.

3.2.4. Minimization

Based on the preprocessing method in 3.1, the parallax between aperture images is obtained, so for a pixel in aperture A_i, the corresponding pixels $\mathbf{x}_j$ are known. Although a new compound term is added, the same penalty function is used for compound term so that the total energy function is differentiable. The energy function of each aperture, according to the calculus of variations, a minimizer of eq.(2) must fulfill the associated Euler-Lagrange equations with homogeneous Neumann boundary conditions. For better readability we follow the abbreviations defined in [4]

$$
\begin{aligned}
I_x &:= \partial_x I(\mathbf{x}+\mathbf{w}), \\
I_y &:= \partial_y I(\mathbf{x}+\mathbf{w}), \\
I_z &:= I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x}), \\
I_{xx} &:= \partial_{xx} I(\mathbf{x}+\mathbf{w}), \\
I_{xy} &:= \partial_{xy} I(\mathbf{x}+\mathbf{w}), \\
I_{yy} &:= \partial_{yy} I(\mathbf{x}+\mathbf{w}), \\
I_{xz} &:= \partial_x I(\mathbf{x}+\mathbf{w}) - \partial_x I(\mathbf{x}), \\
I_{yz} &:= \partial_y I(\mathbf{x}+\mathbf{w}) - \partial_y I(\mathbf{x}),
\end{aligned}
\tag{6}
$$

where t is replaced by z. Since the energy function of each aperture is the same, i is omitted for simplicity.

The Euler-Lagrange equations of eq.(2) is expressed as

$$
\begin{aligned}
&\Psi'(I_z^2 + \gamma(I_{xz}^2 + I_{yz}^2)) \cdot (I_x I_z + \gamma(I_{xx}I_{xz} + I_{xy}I_{yz})) \\
&+ \beta \frac{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot \Psi'(|\mathbf{w}-\mathbf{w}_j|^2)(u - u_j)}{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j)} \\
&- \alpha\, \mathbf{div}(\Psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2)\nabla_3 u) = 0, \\
&\Psi'(I_z^2 + \gamma(I_{xz}^2 + I_{yz}^2)) \cdot (I_y I_z + \gamma(I_{yy}I_{yz} + I_{xy}I_{xz})) \\
&+ \beta \frac{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot \Psi'(|\mathbf{w}-\mathbf{w}_j|^2)(v - v_j)}{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j)} \\
&- \alpha\, \mathbf{div}(\Psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2)\nabla_3 v) = 0.
\end{aligned}
\tag{7}
$$

Since the initial optical flow based on Deep-Matching still suffers from the aperture problem, The variational refinement is based on a standard coarse-to-fine warping scheme. After a pyramid of images is constructed, the flow is estimated from the coarsest level to the finest level, and at each level, a fixed point iteration method is used to compute $\mathbf{w}$. Let $\mathbf{w}^k = (u^k, v^k, 1)^\top$, $k = 0, 1, ...$, be the displacement vector at iteration k. The coarsest level $\mathbf{w}^0$ is initialized based on the DeepMatching result $\mathbf{w}^{init}$, and the finer level is initialized by using the final result of previous level. At each step $k+1$, $\mathbf{w}^{k+1}$ will be the solution of

$$
\begin{aligned}
&\Psi'((I_z^{k+1})^2 + \gamma((I_{xz}^{k+1})^2 + (I_{yz}^{k+1})^2)) \\
&\cdot (I_x^k I_z^{k+1} + \gamma(I_{xx}^k I_{xz}^{k+1} + I_{xy}^k I_{yz}^{k+1})) + \\
&\beta \frac{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot \Psi'(|\mathbf{w}^{k+1}-\mathbf{w}_j^{k+1}|^2)(u^{k+1} - u_j^{k+1})}{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j)} \\
&- \alpha\, \mathbf{div}(\Psi'(|\nabla_3 u^{k+1}|^2 + |\nabla_3 v^{k+1}|^2)\nabla_3 u^{k+1}) = 0, \\
&\Psi'((I_z^{k+1})^2 + \gamma((I_{xz}^{k+1})^2 + (I_{yz}^{k+1})^2)) \\
&\cdot (I_y^k I_z^{k+1} + \gamma(I_{yy}^k I_{yz}^{k+1} + I_{xy}^k I_{xz}^{k+1})) + \\
&\beta \frac{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot \Psi'(|\mathbf{w}^{k+1}-\mathbf{w}_j^{k+1}|^2)(v^{k+1} - v_j^{k+1})}{\sum\limits_{j\in Ne(i)} \delta(\mathbf{x}_j)} \\
&- \alpha\, \mathbf{div}(\Psi'(|\nabla_3 u^{k+1}|^2 + |\nabla_3 v^{k+1}|^2)\nabla_3 v^{k+1}) = 0.
\end{aligned}
\tag{8}
$$

We assume $u^{k+1} = u^k + \triangle u^k$, $v^{k+1} = v^k + \triangle v^k$, and use the same approximate expression of I_*^{k+1} as [4]

$$
\begin{aligned}
I_z^{k+1} &\approx I_z^k + I_x^k \triangle u^k + I_y^k \triangle v^k, \\
I_{xz}^{k+1} &\approx I_{xz}^k + I_{xx}^k \triangle u^k + I_{xy}^k \triangle v^k, \\
I_{yz}^{k+1} &\approx I_{yz}^k + I_{xy}^k \triangle u^k + I_{yy}^k \triangle v^k.
\end{aligned}
\tag{9}
$$

Then let

$$
\begin{aligned}
(\Psi')_{Data}^k &= \Psi'((I_z^k + I_x^k \triangle u^k + I_y^k \triangle v^k)^2 \\
&+ \gamma((I_{xz}^k + I_{xx}^k \triangle u^k + I_{xy}^k \triangle v^k)^2 + (I_{yz}^k + I_{xy}^k \triangle u^k + I_{yy}^k \triangle v^k)^2)), \\
(\Psi')_{Smooth}^k &= \Psi'(|\nabla_3(u^k + \triangle u^k)|^2 + |\nabla_3(v^k + \triangle v^k)|^2), \\
(\Psi')_{Compound}^k &= \Psi'(|(\mathbf{w}^k + \triangle\mathbf{w}^k) - (\mathbf{w}_j^k + \triangle\mathbf{w}_j^k)|^2).
\end{aligned}
\tag{10}
$$

With eqs. (9) (10), the first equation in eq. (8) can be written as

$$(\Psi')^k_{Data} \cdot \left(I^k_x(I^k_z + I^k_x \triangle u^k + I^k_y \triangle v^k) + \right.$$

$$\left. \gamma(I^k_{xx}(I^k_{xz} + I^k_{xx} \triangle u^k + I^k_{xy} \triangle v^k) + I^k_{xy}(I^k_{yz} + I^k_{xy} \triangle u^k + I^k_{yy} \triangle v^k)^2) \right) +$$

$$\beta \frac{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot (\Psi')^k_{Compound} \cdot ((u^k + \triangle u^k) - (u^k_j + \triangle u^k_j))}{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j)}$$

$$- \alpha \, \mathbf{div} \left((\Psi')^k_{Smooth} \cdot \nabla_3(u^k + \triangle u^k) \right) = 0,$$

$$(11)$$

and the second equation can be expressed in a similar way. In order to estimate the increment of displacement vector at each iteration, we follow Brox *et al* [4] and use a second, inner, fixed point iteration. We initialize the inner iteration with $\triangle u^{k,0} = 0$, $\triangle v^{k,0} = 0$. Then at each step $l + 1$, the system of equations in increment of displacement vector $\triangle u^{k,l+1}$, $\triangle v^{k,l+1}$ are

$$(\Psi')^{k,l}_{Data} \cdot \left(I^k_x(I^k_z + I^k_x \triangle u^{k,l+1} + I^k_y \triangle v^{k,l+1}) \right.$$

$$+ \gamma(I^k_{xx}(I^k_{xz} + I^k_{xx} \triangle u^{k,l+1}$$

$$\left. + I^k_{xy} \triangle v^{k,l+1}) + I^k_{xy}(I^k_{yz} + I^k_{xy} \triangle u^{k,l+1} + I^k_{yy} \triangle v^{k,l+1})) \right) +$$

$$\beta \frac{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j) \cdot g(\mathbf{x}_j) \cdot (\Psi')^{k,l}_{Compound} \cdot ((u^k + \triangle u^{k,l+1}) - (u^k_j + \triangle u^{k,l+1}_j))}{\sum\limits_{j \in Ne(i)} \delta(\mathbf{x}_j)}$$

$$- \alpha \, \mathbf{div} \left((\Psi')^{k,l}_{Smooth} \cdot \nabla_3(u^k + \triangle u^{k,l+1}) \right) = 0.$$

$$(12)$$

For the first equation of eq.(8), here we further assume that $\triangle u^{k,l+1}_j = 0$, and we use the u^{k-1}_j to replace u^k_j so that there is not any new unknown variable added. Then the final linear system can be solved by successive overrelaxation (SOR) method [54].

The minimization is performed from the coarse level to the fine level, and therefore, the original image resolution is the finest level. At each image level, the outer loop is iterated k times. In each outer loop, the inner loop iterates l times to estimate the increment of displacement vector. After each outer loop iteration we further use a weighted median filter (WMF) with size of 5×5 to remove outliers and preserve motion boundary [42,43]. The pseudo code of our method is shown in **Algorithm** 1, and the function $f(\mathbf{w}^k_r)$ is to

---

Algorithm 1 Pseduo code of the proposed optical flow method

Input: two frame of image set $I_i(t), I_i(t+1)$, $i \in [1,M] \times [1,N]$.
Output: displacement vector $\mathbf{w}$
1: rectify the eCley image (Section 3.1)
2: initial optical flow estimation (Section 3.2.2)
3: variational refine method initialization
4: **for** image level $r = 1 : maxlevel$ **do**
5:      **for** iter $k = 1 : 3$ **do**
6:          **for** each $I_{im}(t)$ and corresponding $I_{im}(t + 1)$ **do**
7:              $\Delta\mathbf{w}^k_r = \mathbf{0}$
8:              $\Delta\mathbf{w}^k_r$ estimation (eq.(12))
9:              $\mathbf{w}^{k+1}_r = \mathbf{w}^k_r + \Delta\mathbf{w}^k_r$
10:            $\mathbf{w}^{k+1}_r = \mathbf{WMF}(\mathbf{w}^{k+1}_r)$
11:      $\mathbf{w}^1_{r+1} = f(\mathbf{w}^k_r)$

---

transfer the displacement vector $\mathbf{w}^k_r$ to next finer scale as an initialization.

4. Experimental results

This section presents the experimental evaluation of the proposed method. In this paper, all experiments are performed on a laptop with an Intel Core i5-3210M CPU clocked at 2.5GHz and 4GB RAM. To evaluate the efficiency of the proposed method, four optical flow methods, DeepFlow [33], Classic+NL [43], LDOF [5] and EpicFlow [32], are used for comparison. Classsic+NL method is a baseline variational optical flow estimation method. DeepFlow, EpicFlow and LDOF used dense descriptors matching and interpolation to estimate the initial flow, and then used variational method to optimize the optical flow. In the proposed method, the initial optical flow is obtained as the same as DeepFlow. The accurate descriptor matching method and interpolation method can largely improve the initial flow result, and further achieve more accurate optical flow after optimization. Among the optical flow results submitted to Sintel dataset based on variational method (with or without descriptor matching), EpicFlow ranks 17th, DeepFlow ranks 20th, Classic+NL ranks 43th, and LDOF ranks 50th. The parameters of DeepFlow, Classic+NL, LDOF, EpicFlow are kept the same as in the litera-

tures. The evaluation of the results is based on Middlebury datasets, Sintel datasets and real data captured by eCley, and the performances of benchmark datasets with ground truth are evaluated with the standard metrics, the average angular error (AAE) and average endpoint error (EPE). All the visualized optical flow image are using the color coding method as [2].

In our method, all parameters at the initial optical flow estimation stage are kept the same as in [33]. At the variational refinement stage, since the most of motions in the Middlebury dataset are small, while the Sintel dataset contains large motions, therefore, the downsampling factor in coarse-to-fine step of Middlebury dataset and Sintel dataset are 0.8 and 0.5, respectively. For the outer iteration at each image level k, Fig. 4 gives the AAE and EPE of four images and their averages under different k. In order to better visualize the results, the AAE and EPE at different iterations are compared with the initial flow's (iteration = 0), and the ratio of AAE and EPE are shown in the figure. As shown in the figure, after 1 iteration, both the EPE and AAE decrease significantly thanks to the coarse-to-fine scheme. As the iteration increase, most AAE and EPE decrease gradually and then turn to increases. When $k = 3$, the average AAE and EPE are the smallest. Therefore, we use $k = 3$ in our experiments. The inner iteration based on SOR method $l = 100$. Fig. 5 - Fig. 8 gives the AAE and EPE of Middlebury data and Sintel data under different α and β, respectively. The data are divided into image array as described in Section 4.1. We find that good results of Sintel data are obtained at $\alpha = 0.5$, $\beta = 1$. For Middlebury data, $\alpha = 0.9$ has the best result, and the smallest AAE obtained at $\beta = 1.2$ but smallest EPE at $\beta = 0.8$, in order to balance the AAE and EPE, we choose the median value with $\beta = 1$ for the Middlebury data.

4.1. Results on benchmark datasets

Middlebury dataset The evaluation on Middlebury Benchmark is performed on 7 pairs of frames, which have ground truth optical flow. In order to construct an image array like the images captured by eCley, each frame is divided into a set of partially overlapping images. More concretely, the *RubberWhale*, *Hydrangea* and *Dimetrodon* sequence with resolution of 584×388 pixels are divided into 7×11 sub-images with reso-
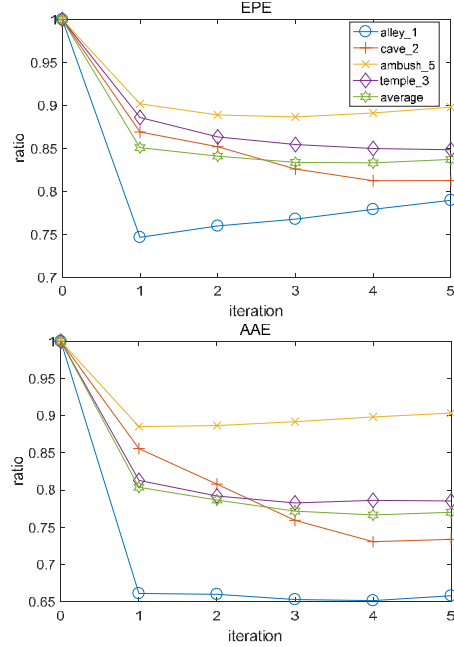

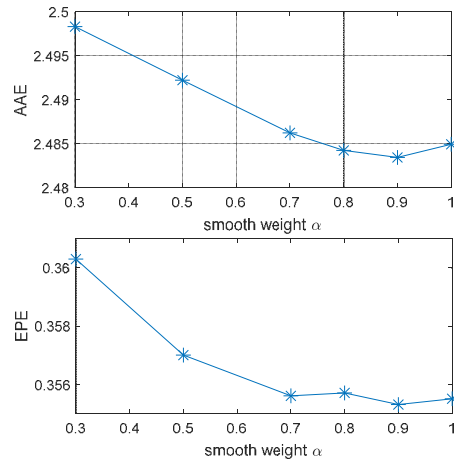**Figure 4.** The AAE and EPE ratio under different k.

Figure 5. The AAE and EPE of Middlebury data under different α.

lution of 96×96 pixels, the *Grove2*, *Grove3*, *Urban2* and *Urban3* sequence with resolution of 640×480 pixels are divided into 5×7 sub-images with resolution of 160×160 pixels, and the overlap between neighboring images is half FOV of the sub-image. The ground truth optical flow images are divided in the same way. The sub-image pair is extracted from the
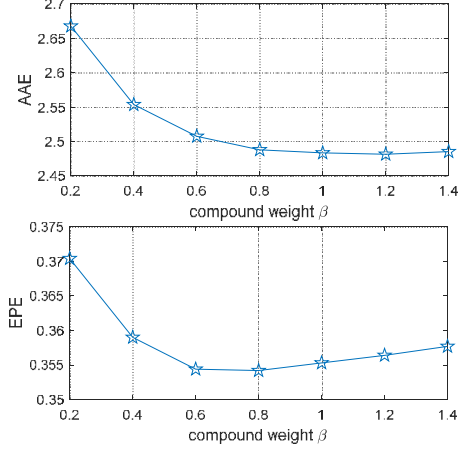
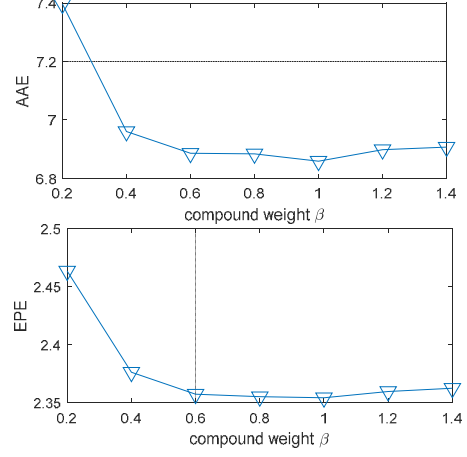**Figure 6.** The AAE and EPE of Middlebury data under different β.

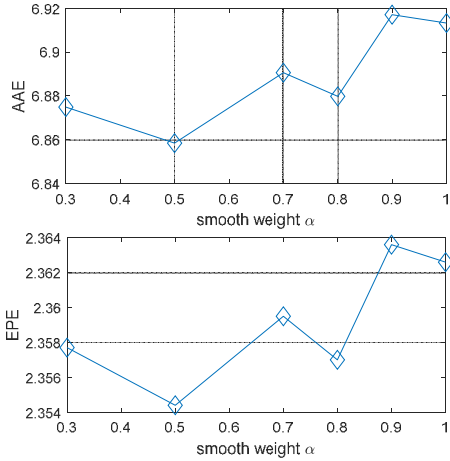**Figure 8.** The AAE and EPE of Sintel data under different β.

Figure 7. The AAE and EPE of Sintel data under different α.

original image pair at the same position but in different frame, and is a local image pair of original image pair. Since the ground truth describes the displacement of each pixel of current frame at the following frame, the displacement of each pixel in sub-image will keep the same as the original image. Therefore, the optical flow extracted from the same position can be considered as the ground truth of each sub-image. The original image, ground truth optical flow, and the divided image array and ground truth array are shown in Fig. 9. Table 1 gives AAE and EPE of five optical flow methods, and the visualized results are shown in Fig. 10. The percentage in the bracket shown in Table 2 indi-
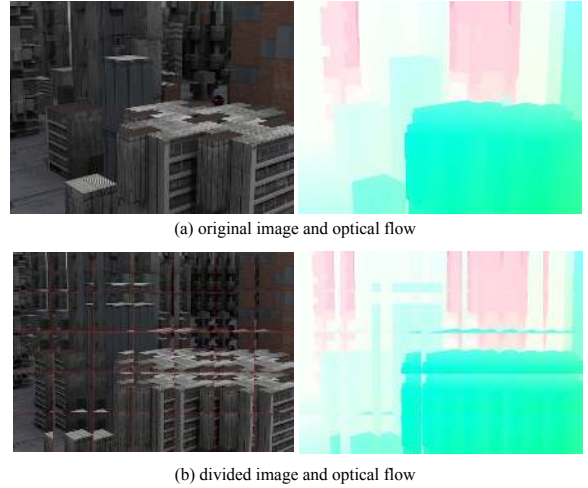

(a) original image and optical flow

(b) divided image and optical flow

Figure 9. (a) left: the original Urban2 image; right: the original optical flow ground truth. (b) left: the divided image array; right: the divided optical flow ground truth.

cates the improvement rate of our method, in which the negative percentage means our method is worse than the corresponding method. Our method achieves the best result among the five methods except for the *Dimetrodon* sequence.

Sintel dataset The Sintel dataset and ground truth is divided in the similar way. The resolution of Sintel image is 1024×436 pixels. Due to the large motion existed in Sintel image, we use a relatively high resolution (200×200) of sub-image to avoid the ob-
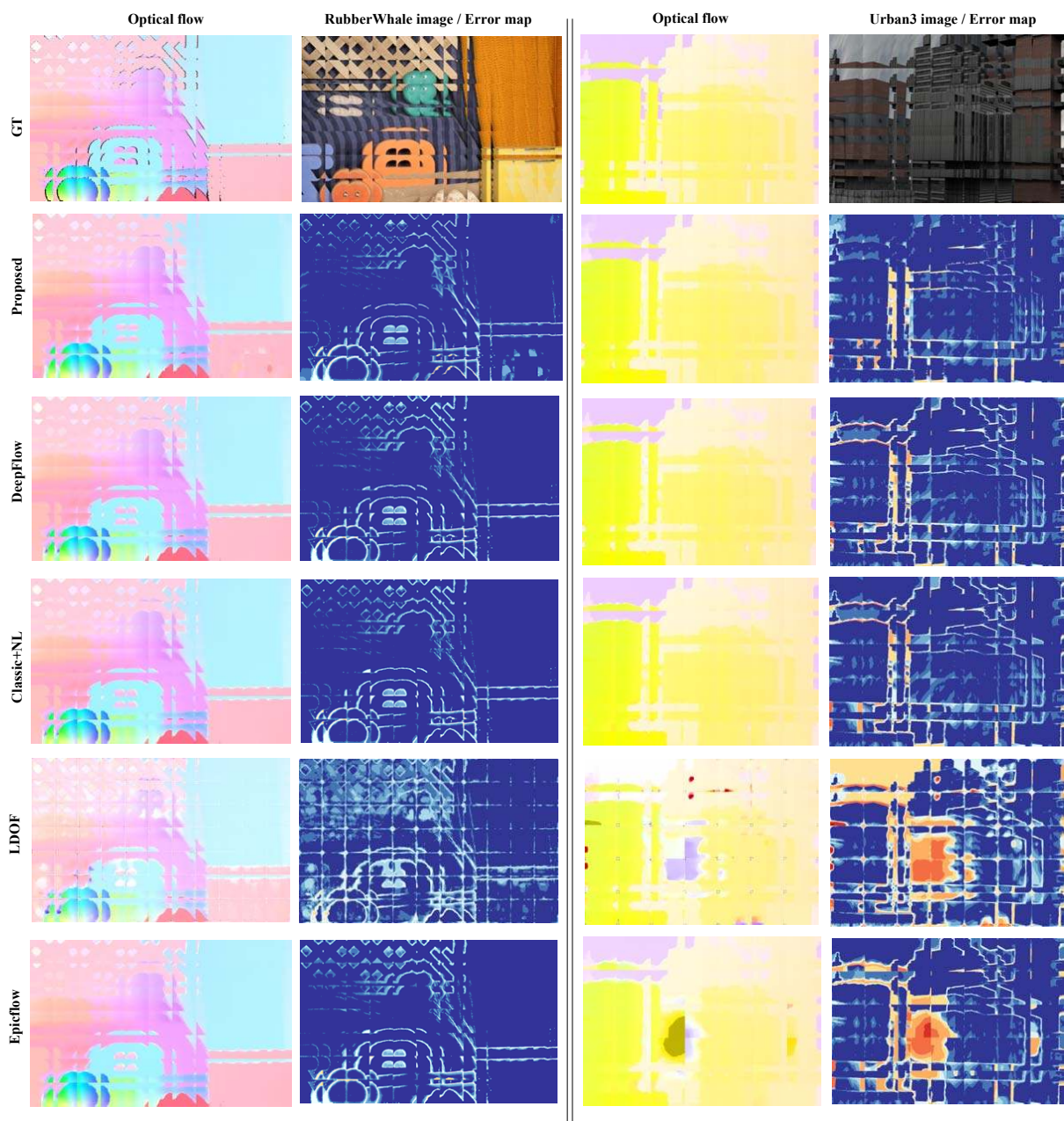
Figure 10. Visualized optical flow results on Middleburry datasets. In the error map, the blue color means no error or tiny error, white means small error, the orange color indicates large error.

Table 1. AAE/EPE for test sequences from Middlebury benchmark

Sequences	Classic+NL	DeepFlow	EpicFlow	LDOF	Proposed
RubberWhale	3.2144/ 0.0996	3.2442/ 0.1014	4.0227/ 0.1232	7.5509/ 0.2346	**2.8994/ 0.0947**
	(10.86%/ 5.17%)	(11.89%/ 7.07%)	(38.74%/ 30.10%)	(160.43%/ 147.73%)	
Hydrangea	2.1954/ 0.1809	2.6967/ 0.2027	2.3398/ 0.2003	3.6114/ 0.3680	**2.1344/ 0.1743**
	(2.86%/ 3.79%)	(26.34%/ 16.29%)	(9.62%/ 14.92%)	(69.20%/ 111.13%)	
Dimetrodon	1.7508/ 0.0909	**1.7448/ 0.0905**	2.0073/ 0.1016	2.4334/ 0.2610	1.8672/ 0.0969
	(-6.23%/ -6.19%)	(-6.55%/ -6.6%)	(7.5%/ 4.85%)	(30.32%/ 169.35%)	
Urban2	2.9435/ 0.7381	2.7157/ 0.3780	3.5239/ 0.5181	5.8456/ 1.5918	**2.5436/ 0.3301**
	(15.72%/ 123.6%)	(6.77%/ 14.51%)	(38.53%/ 56.95%)	(129.82%/ 382.22%)	
Urban3	3.5065/ 0.4410	3.4933/ 0.4251	6.5589/ 1.2668	18.4243/ 1.9045	**2.2718/ 0.3189**
	(54.35%/ 38.29%)	(53.77%/ 33.3%)	(188.71%/ 297.24%)	(711.0%/ 495.53%)	
Grove2	2.1397/ 0.1432	2.1276/ 0.1435	2.3944/ 0.1587	3.3487/ 0.2753	**1.4409/ 0.0981**
	(32.66%/ 45.87%)	(47.65%/ 46.28%)	(66.17%/ 61.77%)	(132.40%/ 180.63%)	
Grove3	5.8726/ 0.6077	5.8431/ 0.6001	6.7210/ 0.7004	8.7605/ 0.9313	**5.0618/ 0.4937**
	(16.02%/ 23.09%)	(15.44%/ 21.58%)	(32.78%/ 41.87%)	(73.07%/ 88.63%)	

ject moving outside the sub-image. The image array contains 3×9 sub-images. Table 2 shows the AAE and EPE of 16 Sintel image sequences. Our method performs the best for almost all the test images. Fig. 11 gives the visualized results of divided image arrays, and Fig. 12 shows the sub-image results of image array for visualized comparison. In Fig. 11, the Classic+NL and LDOF are much worse than the other three methods. Our method yields the similar performance to DeepFlow method and both of them outperform the EpicFlow method. As shown in Fig. 12, for the motion boundary near the image border, our method is able to correctly estimate the motion boundary and performs better than DeepFlow method.

In addition, the compared methods are used to estimate the optical flow based on single image sequence. Therefore, Table 3 gives the results on the original benchmark data in comparison with our result on multi-aperture images. The results indicate that the proposed method for small sub-image array achieves an approximate optical flow accuracy to these methods on the original image sequences, although the aperture problem of sub-image array is more severe than the original image.

4.2. Results on real data

In this section, we used the real-world sequences captured by eCley. the eCley image used in this section contains 13×13 apertures. Each aperture image after

rectification has a resolution of 101×101 pixels, and the parameters of our method on eCley image keep the same as in Sintel dataset. The first sequences (Fig. 13) involves the vertical motion of an arm and a hand, in a static scene. As shown in Fig. 13, Classic+NL method is able to estimate the motion boundary, but the corresponding flow fields across adjacent aperture images do not keep consistent; DeepFlow, LDOF and EpicFlow use descriptor matching to estimate initial optical flow and have more consistent results, but they can not accurately estimate the motion boundary; Our method is able to estimate the motion boundary and keep the flow fields consistent between adjacent images. In addition, EpicFlow uses the edges information for interpolation, the optical flows of textureless sub-images are much worse than our method and the other three methods. The second sequences in Fig. 14 show complex motions of a person, where the body and head have a horizontal motion, and two hands are vertically moved. Also, the Classic+NL method is less consistent. DeepFlow and LDOF fail to estimate the motion boundary. EpicFlow yields worse result at textureless sub-images. Our method performs best.

The computing time of the variational step in the proposed method takes 170 seconds for Sintel data, 125 seconds for Middlebury data, and 213 seconds for eCley images. The DeepFlow method takes 140, 103 and 194 seconds for Sintel data, Middlebury data and eCley images, respectively. The Classic+NL method takes 240, 177, 421 seconds, respectively. The LDOF
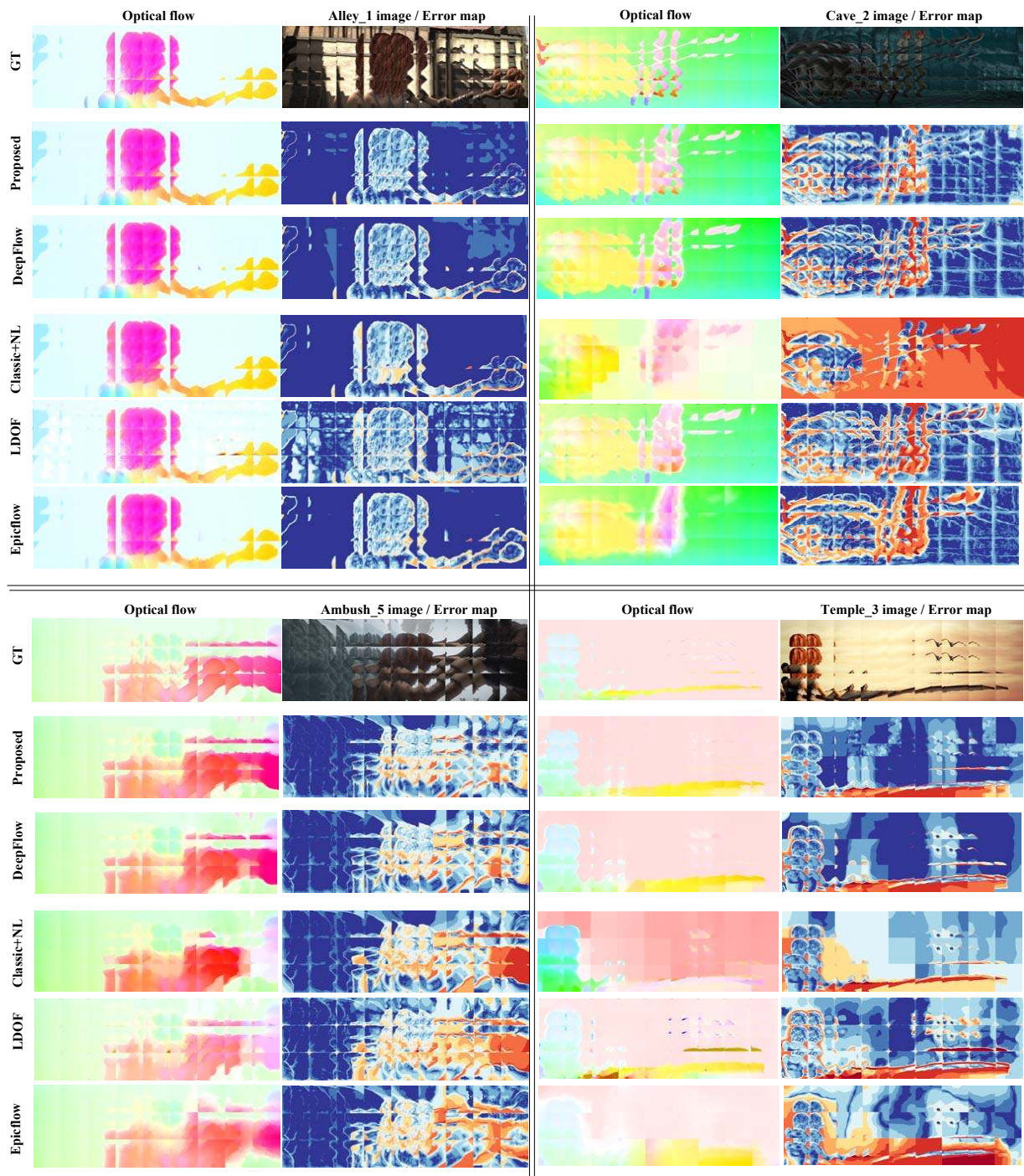
Figure 11. Visualized optical flow results on Sintel datasets. In the error map, the blue color means no error or tiny error, white means small error, the orange color indicates large error.

Table 2. AAE/EPE for test sequences from Sintel benchmark

Sequences	Classic+NL	DeepFlow	EpicFlow	LDOF	Proposed
alley_1	4.8839/ 0.5602	5.2970/ 0.4916	5.8962/ 0.5762	12.8830/ 0.8826	**3.4515/ 0.3710**
	(41.5%/ 51%)	(53.47%/ 32.51%)	(70.83%/ 55.31%)	(273.26%/ 137.90%)	
alley_2	4.1721/ 0.4771	3.4958/0.4010	4.6241/ 0.5438	4.9992/ 0.6666	**2.3877/ 0.3206**
	(74.73%/ 48.81%)	(46.41%/ 25.08%)	(93.66%/ 69.62%)	(109.37%/ 107.92%)	
ambush_4	30.8977/ 16.5084	18.8441/ 9.9561	21.3644/ 12.6051	24.7671/16.8604	**16.9577/ 9.1512**
	(82.2%/ 80.4%)	(11.12%/ 8.79%)	(25.99%/ 37.74%)	(46.05%/ 84.24%)	
ambush_5	14.1738/ 3.3918	9.0546/ 1.7866	14.2703/ 2.7259	13.8827/ 3.4023	**8.2569/ 1.6359**
	(71.66%/ 107.33%)	(9.66%/ 9.21%)	(72.83%/ 66.63%)	(68.13%/ 105.71%)	
ambush_7	17.0295/ 6.1172	**11.3141**/ 2.3059	16.1809/ 3.5922	20.1161/ 3.7040	12.7842/ **2.1459**
	(33.21%/ 185.07%)	(-11.5%/ 7.46%)	(26.57%/ 67.40%)	(57.35%/ 72.61%)	
bamboo_1	4.7622/ 0.4509	4.8079/ 0.4528	4.9010/ 0.4672	8.1417/ 0.7365	**4.2248/ 0.4180**
	(12.72%/ 7.87%)	(13.8%/ 8.33%)	(16.01%/ 11.77%)	(92.71%/ 76.20%)	
bamboo_2	11.3918/ 1.7387	8.6703/ 0.9452	10.0012/ 1.2423	19.5096/ 1.5828	**7.6564/ 0.8378**
	(48.79%/ 107.53%)	(13.24%/ 12.82%)	(30.63%/ 48.28%)	(154.81%/ 88.92%)	
bandage_1	8.0451/ 1.6240	6.2041/ 1.2877	8.1358/ 1.5168	9.8101/ 2.5097	**5.6050/ 1.2075**
	(43.53%/ 34.49%)	(10.69%/ 6.64%)	(45.15%/ 25.61%)	(75.02%/ 107.84%)	
cave_2	41.9416/ 16.0512	9.0333/ 2.8502	15.5628/ 4.1770	13.0757/ 3.9949	**7.8904/ 2.5342**
	(506.69%/ 579.39%)	(30.67%/ 20.64%)	(97.24%/ 64.83%)	(65.72%/ 57.64%)	
cave_4	22.2026/ 6.3129	14.0963/ 3.9122	21.2991/ 5.4596	17.9773/ 4.9500	**12.7640/ 3.6907**
	(73.95%/ 71.05%)	(10.44%/ 6.0%)	(66.87%/ 47.93%)	(40.84%/ 34.12%)	
market_2	12.9066/ 1.7816	12.0870/ 1.2845	15.9892/ 1.7816	21.6570/2.0481	**10.4974/ 1.1384**
	(22.95%/ 56.50%)	(15.14%/ 12.83%)	(52.32%/ 56.50%)	(106.31%/ 79.91%)	
market_5	21.2936/ 14.6889	7.5622/ 2.8192	9.6230/ 3.4391	11.2901/ 5.4965	**6.9999/ 2.5895**
	(204.2%/ 467.25%)	(8.03%/ 8.87%)	(37.47%/ 32.81%)	(61.29%/ 112.26%)	
market_6	25.7618/ 10.3904	18.9207/ 6.1782	22.6330/ 7.5181	24.7129/ 9.2926	**17.6978/ 6.0227**
	(45.57%/ 72.52%)	(6.91%/ 2.58%)	(27.89%/ 24.83%)	(39.64%/ 54.29%)	
mountain_1	**5.6542**/ 0.736	6.2371/ **0.5722**	8.6418/ 0.6092	20.6351/ 1.3487	7.0734/ 0.7204
	(-20.06%/ 2.16%)	(-11.82%/ -20.57%)	(22.17%/ -15.44%)	(191.73%/ 87.22%)	
temple_2	30.7928/ 24.9824	13.8948/ 11.8773	18.3791/ 15.3767	19.6556/ 18.5054	**11.7069/ 11.0138**
	(163.03%/ 126.83%)	(18.69%/ 7.84%)	(56.99%/ 39.61%)	(67.90%/ 68.02%)	
temple_3	16.9331/ 3.6524	12.4207/ 3.4254	25.2147/ 6.9696	16.6264/ 7.5460	**9.6350/ 2.9133**
	(75.75%/ 25.37%)	(28.91%/ 17.58%)	(161.70%/ 139.23%)	(72.56%/ 159.02%)	

method takes 317, 244, 648 seconds and EpicFlow method takes 123, 88 and 160 seconds. In addition, although our method imposes a compound term to improve the result, at each iteration, the aperture image only needs the optical flow fields of current aperture image and adjacent images from the last iteration. Therefore, the estimation of optical flow of one aperture image does not need the information from this iteration and can be accelerated by using multithreading processors.

From the experimental results on Middlebury datasets, Sintel datasets and the real data, the proposed method has a better performance on the sub-image array structure. But there are also some failure cases, i.e. the proposed method can only take care of the motion within an aperture image, if the motion crosses the aperture images, then the proposed method will fail to obtained the accurate flow. As shown in Fig 15, the bottom-left image sequences cannot estimate the accurate optical flow due to the across images motions. Since the large motion that across images will lead to the lack of matching point for the moving object, therefore, future work will focus on the fast moving objects that across multiple images.
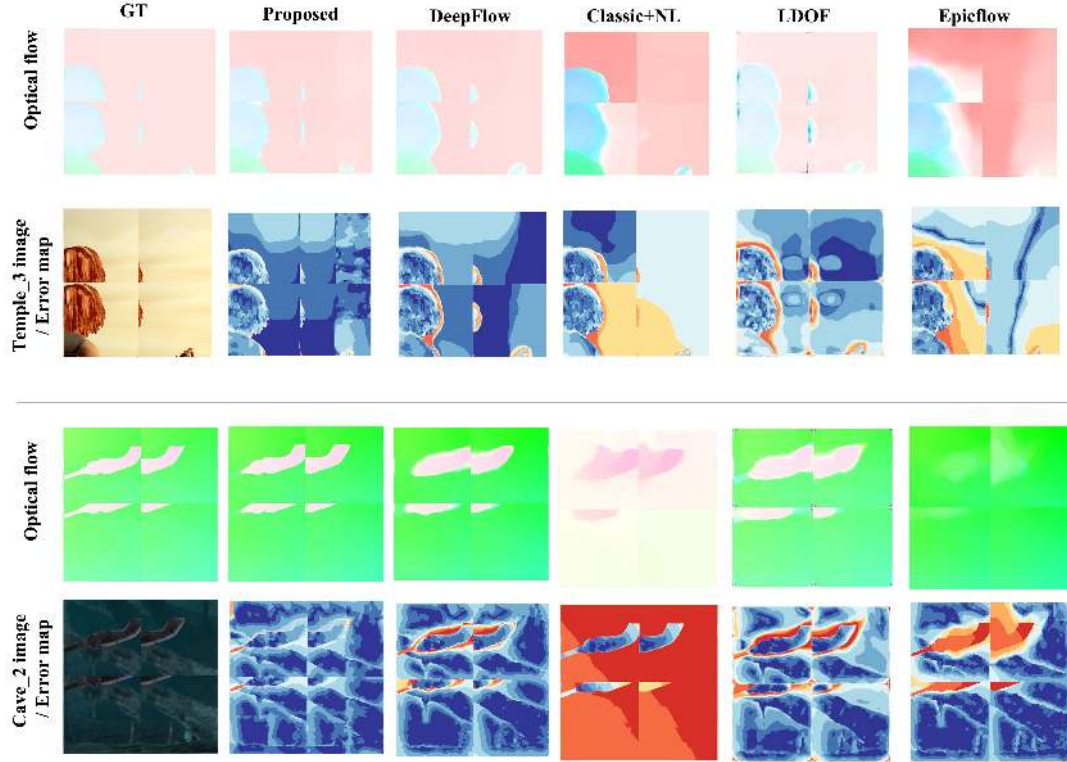
Figure 12. The optical flow results of sub-image array.

Table 3. AAE/EPE for data based on original images

| Sequences	Classic+NL	DeepFlow	EpicFlow	LDOF	Proposed
RubberWhale	3.0539/ **0.0921**	3.0679/ 0.0925	2.9382/ 0.0959	3.8738/ 1.3901	**2.8994**/ 0.0947
Hydrangea	1.8994/ 0.1520	1.8895/ 0.1507	**1.8235/ 0.1464**	2.7091/ 0.3102	2.1344/ 0.1743
Dimetrodon	1.4704/ 0.0771	**1.4699/ 0.0769**	1.4823/ 0.0821	2.0311/ 1.1304	1.8672/ 0.0969
Urban2	**2.2361**/ 0.3012	2.2407/ **0.2881**	2.3349/ 0.2990	6.1663/ 0.7509	2.5436/ 0.3301
Urban3	3.3396/ 0.3734	3.3770/ 0.3758	2.9892/ 0.3594	3.8125/ 0.4972	**2.2718/ 0.3189**
Grove2	2.0853/ 0.1388	2.1270/ 0.1417	1.7491/ 0.1097	2.3481/ 0.1596	**1.4409/ 0.0981**
Grove3	5.5149/ 0.5825	5.4848/ 0.5804	5.3204/ 0.5389	6.3925/ 0.6837	**5.0618/ 0.4937**
alley_1	3.8725/ 0.4002	3.9862/ 0.4054	3.7482/ 0.3897	4.2193/ 0.5203	**3.4515/ 0.3710**
ambush_5	8.5670/ 2.4261	**7.6437/ 1.3808**	7.9523/1.6944	9.3815/ 2.8471	8.2569/ 1.6359
cave_2	10.9686/ 3.1578	7.4004/ 2.3027	7.2018/ **2.2027**	10.8731/ 3.2010	**6.9132**/ 2.3626
temple_2	13.6418/ 12.3415	9.1151/ 6.0763	**8.9470/ 5.9871**	14.0981/ 12.5981	11.7069/ 11.0138
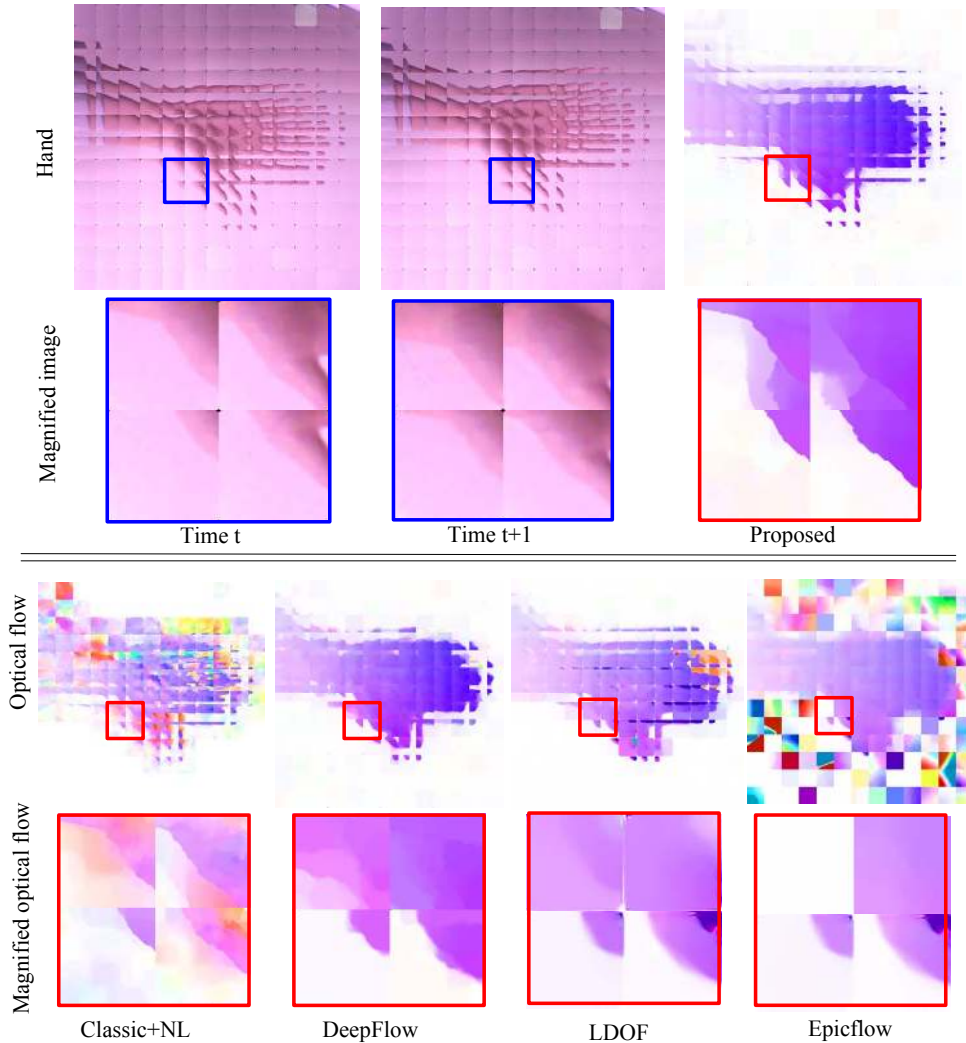
Figure 13. Results comparison of the real data captured by eCley (hand).

5. Conclusion

In this paper, we proposed a multi-aperture optical flow method for ACE - eCley. we first introduce an image rectification method for eCley. Based on the multi-aperture configuration, all apertures are used to simultaneously recover the optical flow of multi-aperture and keep the flow consistent across apertures. In addition, the method based on a variational framework, which imposes a compound flow field consistency assumption along with the brightness constancy and piecewise smoothness assumptions. The corresponding information from multi-apertures adds supplementary constrains that reduce ambiguities and improve stability. Our results on Middlebury benchmark, Sintel benchmark and real data from eCley demonstrate the validity of the proposed method.
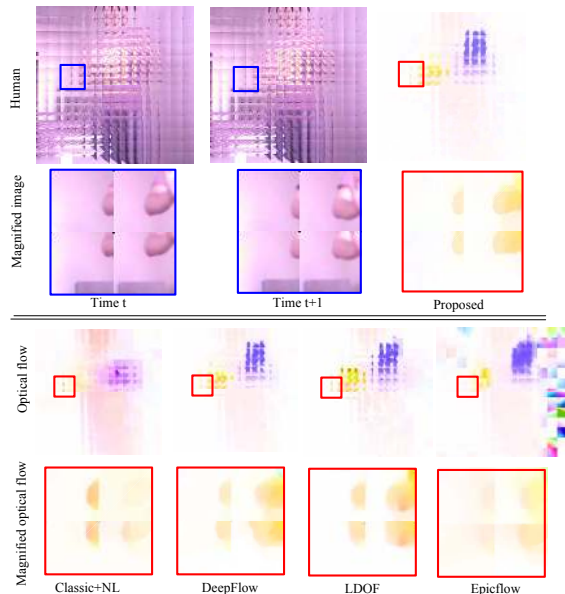
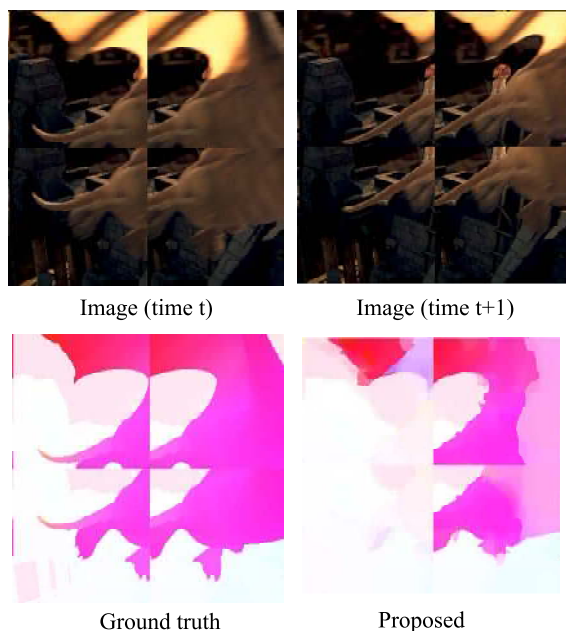**Figure 14.** Results comparison of the real data captured by eCley (human).

Figure 15. The failure situation.

References

[1] C. Bailer, K. Varanasi, and D. Stricker, "CNN-Based patch matching for optical flow with thresholded hinge embedding loss," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2710–2719.

[2] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[3] D. Bračun, N. S. Szczecinski, G. Škulj, A. J. Hunt, and R. D. Quinn, "Artificial compound eye and synthetic neural system for motion recognition," in *Biomimetic and Biohybrid Systems*. Springer International Publishing, 2018, pp. 52–63.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proceedings of the European Conference on Computer Vision*. Springer, 2004, pp. 25–36.

[5] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011.

[6] A. Brückner, J. Duparré, R. Leitel, P. Dannberg, A. Bräuer, and A. Tünnermann, "Thin wafer-level camera lenses inspired by insect compound eyes," *Optics Express*, vol. 18, no. 24, pp. 24 379–24 394, 2010.

[7] J. Chen, J. Hou, Y. Ni, and L.-P. Chau, "Accurate light field depth estimation with superpixel regularization over partially occluded regions," *IEEE Transactions on Image Processing*, 2018.

[8] J. Cheng, Y. H. Tsai, S. Wang, and M. H. Yang, "Segflow: Joint learning for video object segmentation and optical flow," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 686–695.

[9] M. Chessa, S. P. Sabatini, and F. Solari, "A systematic analysis of a V1-MT neural model for motion estimation," *Neurocomputing*, vol. 173, pp. 1811–1823, 2016.

[10] A. Dosovitskiy, P. Fischery, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2758–2766.

[11] R. A. Farrugia, C. Galea, and C. Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1058–1071, 2017.

[12] D. Floreano, R. Pericetcamara, S. Viollet, F. Ruffier, A. Bruckner, R. Leitel, W. Buss, M. Menouni, F. Expert, R. Juston *et al.*, "Miniature curved artificial compound eyes." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 23, pp. 9267–9272, 2013.

[13] N. Franceschini, J. M. Pichon, and C. Blanes, "From insect vision to robot vision," *Philosophical Transactions of the Royal Society B*, vol. 337, no. 1281, pp. 283–294, 1992.

[14] D. Gadot and L. Wolf, "Patchbatch: A batch augmented loss for optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4236–4245.

[15] N. M. Grzywacz and A. L. Yuille, "A model for the estimate of local image velocity by cells in the visual cortex," *Proceedings of the Royal Society B*, vol. 239, no. 1295, pp. 129–161, 1990.

[16] C. Hernández, J. Bermejo-Alonso, and R. Sanz, "A self-adaptation framework based on functional knowledge for augmented autonomy in robots," *Integrated Computer-Aided Engineering*, vol. 25, no. 2, pp. 157–172, 2018.

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

[18] S. Iliya and F. Neri, "Towards artificial speech therapy: A neural system for impaired speech segmentation," *Int. J. Neural Syst.*, vol. 26, no. 6, pp. 1–16, 2016.

[19] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Depth from a light field image with learning-based matching costs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[20] O. Johannsen, K. Honauer, B. Goldluecke, A. Alperovich, F. Battisti, Y. Bok, M. Brizzi, M. Carli, G. Choe, M. Diebold *et al.*, "A taxonomy and evaluation of dense light field depth estimation algorithms." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1795–1812.

[21] A. M. Kabir, J. D. Langsfeld, K. N. Kaipa, and S. K. Gupta, "Identifying optimal trajectory parameters in robotic finishing operations using minimum number of physical experiments," *Integrated Computer-Aided Engineering*, vol. 25, no. 2, pp. 111–135, 2018.

[22] Z. Kootbally, C. Schlenoff, B. Antonishek, F. M. Proctor, T. R. Kramer, W. Harrison, A. Downs, and S. K. Gupta, "Enabling robot agility in manufacturing kitting applications," *Integrated Computer-Aided Engineering*, vol. 25, no. 2, pp. 193–212, 2018.

[23] Z. Lu, X. Chen, X. Zhang, K. Tong, and P. Zhou, "Real-time control of an exoskeleton hand robot with myoelectric pattern recognition," *Int. J. Neural Syst.*, vol. 27, no. 5, pp. 1–11, 2017.

[24] S. Ma, B. M. Smith, and M. Gupta, "3d scene flow from 4d light field gradients," in *Proceedings of the European Confer-

ence on Computer Vision*, 2018, pp. 666–681.

[25] M. Mahesh Mohan and A. Rajagopalan, "Divide and conquer for full-resolution light field deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6421–6429.

[26] F. Martínez, E. Pissaloux, and A. Carbone, "Towards activity recognition from eye-movements using contextual temporal learning," *Integrated Computer-Aided Engineering*, vol. 24, no. 1, pp. 1–16, 2017.

[27] F. Neri and E. Mininno, "Memetic compact differential evolution for cartesian robot control," *IEEE Comp. Int. Mag.*, vol. 5, no. 2, pp. 54–65, 2010.

[28] A. Oberdörster, A. Brückner, F. Wippermann, and A. Bräuer, "Correcting distortion and braiding of micro-images from multi-aperture imaging systems," in *Proc. SPIE*, 2011, p. 78750B.

[29] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 141–158, 2006.

[30] S. Pellegrinelli and N. Pedrocchi, "Estimation of robot execution time for close proximity human-robot collaboration," *Integrated Computer-Aided Engineering*, vol. 25, no. 1, pp. 81–96, 2018.

[31] R. Pericet-Camara, M. K. Dobrzynski, R. Juston, S. Viollet, R. Leitel, H. A. Mallot, and D. Floreano, "An artificial elementary eye with optic flow detection and compositional properties," *Journal of the Royal Society Interface*, vol. 12, no. 109, p. 20150414, 2015.

[32] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Epicflow: Edge-preserving interpolation of correspondences for optical flow," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1164–1172.

[33] ——, "Deepmatching: Hierarchical deformable dense matching," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 300–323, 2016.

[34] D. Rodrigues, J. P. Papa, and H. Adeli, "Meta-heuristic multi- and many-objective optimization techniques for solution of machine learning problems," *Expert Systems*, vol. 34, no. 6, 2017.

[35] S. Rostami and F. Neri, "Covariance matrix adaptation pareto archived evolution strategy with hypervolume-sorted adaptive grid algorithm," *Integrated Computer-Aided Engineering*, vol. 23, no. 4, pp. 313–329, 2016.

[36] ——, "A fast hypervolume driven selection mechanism for many-objective optimisation problems," *Swarm and Evolutionary Computation*, vol. 34, pp. 50 – 67, 2017.

[37] S. Rostami, F. Neri, and M. G. Epitropakis, "Progressive preference articulation for decision making in multi-objective optimisation problems," *Integrated Computer-Aided Engineering*, vol. 24, no. 4, pp. 315–335, 2017.

[38] N. C. Rust, V. Mante, E. P. Simoncelli, and J. A. Movshon, "How MT cells analyze the motion of visual patterns," *Nature Neuroscience*, vol. 9, no. 11, pp. 1421–1431, 2006.

[39] N. H. Siddique and H. Adeli, "Hybrid harmony search algo-

rithms," *International Journal on Artificial Intelligence Tools*, vol. 24, no. 6, 2015.

[40] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vision Research*, vol. 38, no. 5, pp. 743–761, 1998.

[41] P. P. Srinivasan, R. Ng, and R. Ramamoorthi, "Light field blind motion deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 2, 2017, p. 3.

[42] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2432–2439.

[43] ——, "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 115–137, 2014.

[44] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by bound optics (TOMBO): concept and experimental verification," *Applied optics*, vol. 40, no. 11, pp. 1806–1813, 2001.

[45] S. Vedula, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 475–480, 2005.

[46] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4d light-field dataset and cnn architectures for material recognition," in *Proceedings of the European Conference on Computer Vision*. Springer, 2016, pp. 121–138.

[47] X. Wang, G. Zhang, F. Neri, T. Jiang, J. Zhao, M. Gheorghe, F. Ipate, and R. Lefticaru, "Design and implementation of membrane controllers for trajectory tracking of nonholonomic wheeled mobile robots," *Integrated Computer-Aided Engineering*, vol. 23, no. 1, pp. 15–30, 2016.

[48] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*. Springer, 2009, pp. 23–45.

[49] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392.

[50] S. Wu, T. Jiang, G. Zhang, B. Schoenemann, F. Neri, M. Zhu, C. Bu, J. Han, and K.-D. Kuhnert, "Artificial compound eye: a survey of the state-of-the-art," *Artificial Intelligence Review*, vol. 48, no. 4, pp. 573–603, 2017.

[51] S. Wu, G. Zhang, T. Jiang, M. Zhu, K. Fu, H. Rong, K. Xian, H. Song, and K.-D. Kuhnert, "Multi-aperture stereo reconstruction for artificial compound eye with cross image belief propagation," *Applied Optics*, vol. 57, no. 7, pp. B160–B169, 2018.

[52] S. Wu, G. Zhang, M. Zhu, T. Jiang, and F. Neri, "Geometry based three-dimensional image processing method for electronic cluster eye," *Integrated Computer-Aided Engineering*,

vol. 25, no. 3, pp. 213–228, 2018.

[53] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1744–1757, 2012.

[54] D. Young, "Iterative methods for solving partial difference equations of elliptic type," *Transactions of the American Mathematical Society*, vol. 76, no. 1, pp. 92–111, 1954.

[55] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Pattern Recognition.DAGM 2007*. Springer, 2007, pp. 214–223.

[56] G. Zhang, H. Rong, F. Neri, and M. J. Pérez-Jiménez, "An optimization spiking neural P system for approximately solving combinatorial optimization problems," *Int. J. Neural Syst.*, vol. 24, no. 5, 2014.

[57] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[58] H. Zhu, Q. Wang, and J. Yu, "Occlusion-model guided antiocclusion depth estimation in light field," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 965–978, 2017.

[59] J.-C. Zufferey, A. Beyeler, and D. Floreano, *Flying Insects and Robots*. Springer Berlin Heidelberg, 2010, ch. 6, pp. 73–86.