MOTION FIELD DISCONTINUITY CLASSIFICATION FOR TENSOR-BASED OPTICAL FLOW ESTIMATION

Hai-Yun Wang and Kai-Kuang Ma †

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
† Email: ekkma@ntu.edu.sg

ABSTRACT

In this paper, a much more accurate classification scheme is proposed for structure tensor-based optical flow estimation to address the difficulties of interpreting the motion field discontinuities. The key novelties of this approach are: (1) *scale-adaptive spatio-temporal filter*, (2) *weighted structure tensor*, and (3) *confidence measurements.* Multiple motions of moving objects are matched by utilizing spatio-temporal Gaussian filter with adaptive scale selection, which is steered by the condition number. To capture the neighborhood structure of local discontinuities, weighting the structure tensors is attempted. A new normalization function is exploited to facilitate accurate thresholding for confidence measurements. Experimental results demonstrate that these three novelties together effectively contribute much improved performance on motion field discontinuity classification compared with that of existing methods.

1. INTRODUCTION

Optical flow provides an estimation of 2D representation of *apparent velocities* based on the pixel intensity values across a group of adjacent frames [1]. Recently, new optical flow computation schemes, such as *tensor-based* method [3], were proposed based on the *total least squares (TLS)* approach. Tensor-based method is implemented by constructing 3D structure tensor and performing eigen-space analysis. Motion vectors are then estimated according to the thresholding of eigenvalues that also provides the confidence measurements on the discontinuities of motion field.

The optical flow estimation is normally accurate on large, textured areas with uniform motion. For the pixels at the non-coherent spatial areas that involves different motions or very small moving areas, it is difficult to obtain accurate motion vectors due to motion discontinuities. To address this issue, in [2], the geometry of the hypersurface was used for motion detection based on the gradient of hypersurface for tracking moving patterns and detecting motion discontinuities. The discontinuity interpretation based on *3D structure tensor* method has also been investigated in [3][4][5], which will be discussed in Section 3.4.1. However, the results obtained from these methods left room for further improvement.

In this paper, a new approach to classify the motion field through scale-adaptive filtering and eigen-space analysis is developed. In our method, adaptive scale selection steered by the *condition number* is implemented for the spatio-temporal filter to match different motions of moving objects. The *weighted structure tensor* is proposed to capture more neighborhood relationship of the local pixels than the conventional structure tensor. By exploiting new normalization function for the local eigenvalues analysis, each motion vector is classified into one of four categories: *spatial homogenous region, edge, corner and optical flow discontinuity.*

The paper is outlined as follows. Section 2 introduces the basics of the 3D structure tensor. Section 3 describes the proposed scheme for indicating motion field discontinuities. Experimental results of our method and their comparison with other approaches are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. 3D STRUCTURE TENSOR

The *3D structure tensor* is an effective representation of the local orientation for video object motion [3]. The image sequence $I(\mathbf{x})$ is treated as a *volume* data where $\mathbf{x} = (x, y, t)^T$, x and y are the spatial components, and t is the temporal component. The *3D structure tensor* can be generated by:

$$\begin{aligned}\mathbf{J}(x,y,t) &= \begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix} = h(x,y,t) * (\nabla I \cdot \nabla I^T) \\ &= h(x,y,t) * \begin{bmatrix} I_x^2 & I_xI_y & I_xI_t \\ I_xI_y & I_y^2 & I_yI_t \\ I_xI_t & I_yI_t & I_t^2 \end{bmatrix}, \end{aligned} \tag{1}$$

where $\nabla := (\partial_x, \partial_y, \partial_t)$ denotes the spatio-temporal gradients, $h(x,y,t)$ is the spatio-temporal filter, and operator "$*$" performs the convolution. The eigenvalue analysis of the structure tensor corresponds to a total least-squares fitting of a local constant displacement of image intensities [3]. After performing eigenvalue decomposition of the 3×3 symmetric positive matrix $\mathbf{J}(x,y,t)$, the eigenvectors of $\mathbf{J}(x,y,t)$ give the dominant local orientations. The corresponding eigenvalues denote the local grayvalue variations along these directions, which can be used as the confidence measurement for optical flow estimation.

3. PROPOSED CLASSIFICATION SCHEME

To facilitate the accurate tensor-based optical flow estimation and segmentation, a new classification scheme is proposed as shown in Fig. 1 to address the problem of mis-

0-7803-7663-3/03/$17.00 ©2003 IEEE

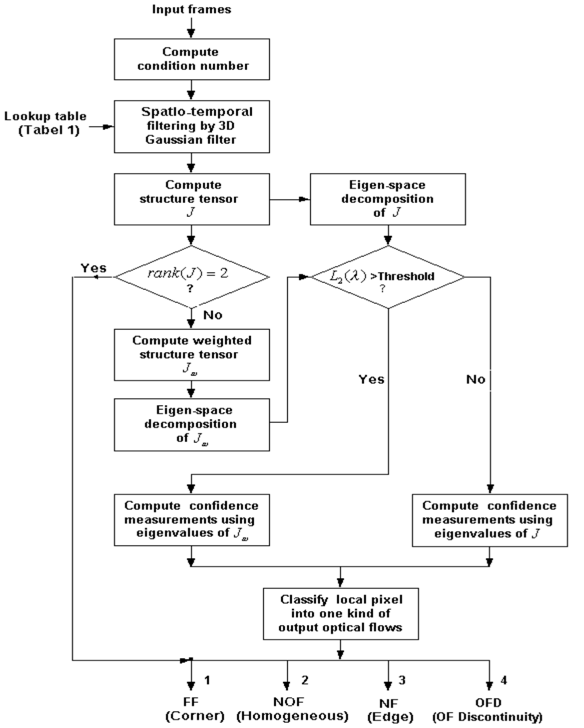

Input frames
Compute condition number
Lookup table (Tabel 1)
Spatlo-temporal filtering by 3D Gaussian filter
Compute structure tensor J
Eigen-space decomposition of J
Yes
$rank(J) = 2$?
$L_2(\lambda)$ >Threshold ?
No
Compute weighted structure tensor J_w
Yes
No
Eigen-space decomposition of J_w
Compute confidence measurements using eigenvalues of J_w
Compute confidence measurements using eigenvalues of J
Classify local pixel into one kind of output optical flows
1
2
3
4
FF (Corner)
NOF (Homogeneous)
NF (Edge)
OFD (OF Discontinuity)

Figure 1: Proposed methodology for motion field discontinuity classification.

Scale (Σ)	[1 1 0.5]	[2 2 1]	[3 3 1.5]	[4 4 2]	[5 5 2.5]
Spatial Window	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
Temporal Window	3	5	7	9	11

Table 1: Experimental scales and spatial windows for the spatio-temporal Gaussian filter, where the three values in Σ correspond to the scales on directions x, y and t, respectively.

classification of motion field discontinuities encountered in the previous works [3][4][5].

3.1. Adaptive scale selection for the spatio-temporal Gaussian filter

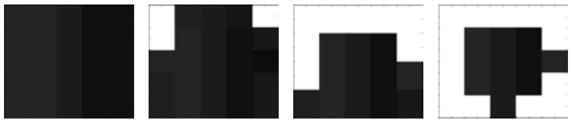
Fixed scale $\Sigma = [\sigma_x \quad \sigma_y \quad \sigma_t]$ was used in [3][5] for spatio-temporal Gaussian filter $h(x, y, t)$,

$$h(x,y,t) = \frac{1}{\sqrt{2\pi\sigma_x^2}\sqrt{2\pi\sigma_y^2}\sqrt{2\pi\sigma_t^2}} exp(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}). \quad (2)$$

Note that small scale size would not be able to match/capture the motion of a video object with large displacements, thus leading to unconnected object boundaries. On the other hand, exploiting large scale size for slow motions will reduce the effectiveness of localization and cause blurred motion discontinuities. Therefore, it is desirable to have *adaptive* scale for the spatio-temporal filter, rather than using *fixed* scale.

The spatio-temporal filter with variable scales is introduced in [4] by iterative symmetric Schur decomposition and re-composition. But the thresholds of its scale adaptation are determined experimentally without theoretical explanations.

Tensor-based optical flow computation is based on TLS fitting. Since numerical stability of the TLS solution can be indicated by *singular value decomposition* (SVD) of the local grayvalue variations, we exploit the *condition number* to guide the scale selection of the spatio-temporal Gaussian filter $h(x, y, t)$ based on the observation presented in Fig. 2, which shows the typical sub-regions of the input frames. The grayvalues of each sub-figure constitute one matrix, whose condition number is also illustrated.



(a) Homogeneous region CN=Infinite

(b) Region with corners CN=6.8229×10^3

(c) Region with edges CN=68.0100

(d) Region with more edges CN=5.8320

Figure 2: Some typical spatial sub-regions and their corresponding condition numbers (CN) computed from the matrix which is constituted by pixel grayvalues.

The *condition number* of local area I_Ω can be computed by means of *singular value decomposition* (SVD):

$$\text{Cond}(I_\Omega) = ||I_\Omega|| ||I_\Omega^{-1}|| = \frac{\sigma_{max}}{\sigma_{min}}. \quad (3)$$

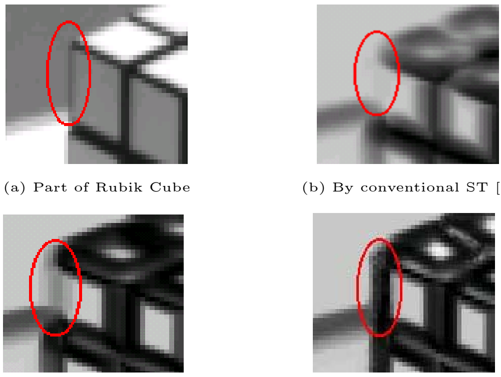
where Ω is the local area in the input frame which is constrained by the spatial scale (σ_x and σ_y) of the spatio-temporal filter, σ_{max} is the maximum singular value and σ_{min} is the minimum one. Note that the condition number of a singular matrix is infinite, and smaller condition number leads to more stable solution.

It can be further observed from Fig. 2 that the more homogeneous the area, the larger value the condition number. The reason for this phenomenon is that coherent grayvalues will cause high correlation in the matrix of I_Ω, thus the condition number is near to the infinity as shown in Fig. 2 (a). With the presence of corners and edges, the matrix correlation is decreased significantly, and the condition number becomes much smaller (see Figs. 2 (b)-(d)). Therefore, it is reasonable to use the condition number of the local intensities to steer the scale Σ of the spatio-temporal filter. In our experiments, the initialization of the scale Σ is set to be [1 1 0.5] and the spatio-temporal window size is $3 \times 3 \times 3$. The scale-size of the spatio-temporal filter should be extended recursively one by one to as referred to Table 1 until either the condition number are below 100 or the scale-size reaches to the maximum one ($11 \times 11 \times 11$).

3.2. Structure tensors computation

After the input frames are filtered by scale-adaptive spatio-temporal Gaussian filter which is steered by the condition number, the components of structure tensor **J** are computed using (1) for each pixel located within each sub-region. If $rank(\mathbf{J}) = 2$, it means that a distributed spatial brightness structure moves at constant velocity, and no aperture problem is presented within the sub-region [3]; thus, the estimated full optical flow is reliable.

Although the eigenvalues of **J** denote the local grayvalue variations along the dominant local orientations [3], conventional structure tensor may give wrong indicators for motion field discontinuities as shown in Fig. 3 (b), especially on the corners and edges as highlighted by the red circle where the grayvalue differences are less noticeable. The reason is that conventional structure tensor has no relationship between



(a) Part of Rubik Cube (b) By conventional ST [3]

(c) By the proposed *equal* weights (d) By the proposed *unequal* weights

Figure 3: Sub-figures (b)-(d) show the motion discontinuity classification results under different kinds of structure tensors (ST) computation.

① 0.1250	② 0.1250	③ 0.1250
④ 0.1250	*	⑤ 0.1250
⑥ 0.1250	⑦ 0.1250	⑧ 0.1250

(a) Proposed *equal* weights

① 0.1250	② 0.1250	③ 0.1875
④ 0.1875	*	⑤ 0.1250
⑥ 0.0625	⑦ 0.1250	⑧ 0.0625

(b) Proposed *unequal* weights

Figure 4: Neighborhood weighting for the weighted structure tensor.

each pixel and its neighborhood. In order to solve this problem, the *weighted structure tensor* is proposed in our scheme to provide local adaptation and defined as:

$$\mathbf{J}_\omega(x, y, t) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \omega_{ij} \mathbf{J}(x + i, y + j, t). \quad (4)$$

which is the sum of functions $\mathbf{J}(x, y, t)$ within an $N \times N$ window centered at (x, y, t), where N can be set from 3 to 7 and $N = 3$ is used in this paper. The weights ω_{ij} given in their corresponding positions are experimentally designed and represented in Fig. 4. The center pixel is marked by "*".

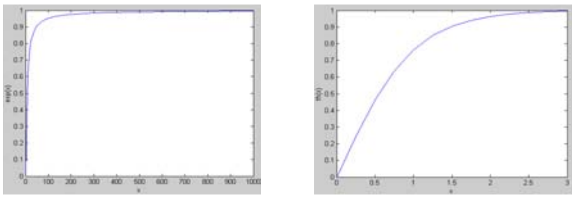
Classification results by using the weighted structure tensors are illustrated in Fig. 3 (c) and (d). It is quite obvious that Fig. 3 (d) achieves the best result.

3.3. Eigen-space analysis on the structure tensors

To estimate optical flow field based on 3D structure tensor, three eigenvalues λ_k (for $k = 1, 2, 3$) of the local pixel should be computed by eigen-space decomposition. Since the smallest eigenvalue points to the main spatio-temporal motion direction of local constant patches [3], the differences between the eigenvalues could be used to measure the reliability of optical flow estimation, they are also used as the indicators to distinguish different kinds of motion vectors, such as *full flow*, *normal flow*, and so on.

As mentioned in Section 3.2, if $rank(\mathbf{J}) \neq 2$, the weighted structure tensor $\mathbf{J}_\omega$ is computed to avoid mis-classification on motion discontinuities. However, in some cases, $\mathbf{J}_\omega$ needs not to be used if the distance between $\mathbf{J}$ and $\mathbf{J}_\omega$ in the eigen-space is very small, i.e.,

$$L_2(\lambda) = \sqrt{(\lambda_1 - \lambda_{\omega 1})^2 + (\lambda_2 - \lambda_{\omega 2})^2 + (\lambda_3 - \lambda_{\omega 3})^2}. \quad (5)$$



(a) Exponential function [5] (b) Hyperbolic tangent function (in our method)

$E(x) = e^{-5/x}$, for $x \geq 0$ $T(x) = \frac{e^{-x} - e^{x}}{e^{-x} + e^{x}}$, for $x \geq 0$

Figure 5: Illustration of different functions used in confidence measurements for classifying motion field discontinuities. Only positive part of the functions are illustrated because x used in mentioned approaches is the distance measurement of the eigenvalues.

where λ and λ_ω are the eigenvalues before and after weighting the structure tensors, and they are sorted in the descending order. The original structure tensor $\mathbf{J}$ should be kept for the following eigen-decomposition if $L_2(\lambda)$ is less than pre-determined threshold (e.g., 10), which means that $\mathbf{J}_\omega$ captures no more neighborhood information than $\mathbf{J}$.

3.4. Proposed criteria for motion discontinuity interpretation

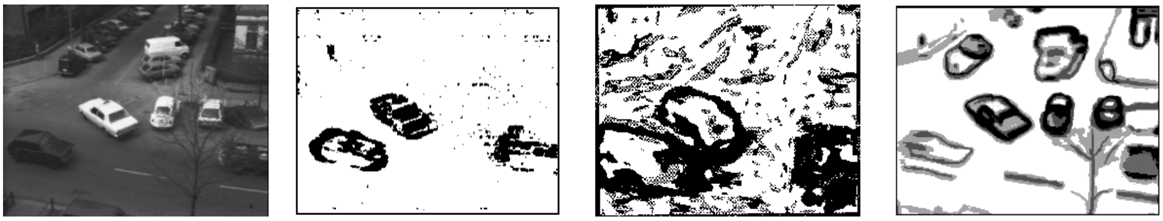
3.4.1. A briefing of previous works

Several confidence measurements have been proposed for classifying motion field discontinuities in [3][4][5]. As mentioned in Section 3.3, the most original criteria [3] were directly derived from the concept of tensor-based optical flow estimation by thresholding the eigenvalues of the structure tensor $\mathbf{J}$. Here, the thresholds are set to be zero from the theoretical point of view. But, for the real-world image sequences, it is impractical to threshold the eigenvalues by zero value due to certain noise level in the input sequences.

Therefore, the normalized confidence measurements are introduced [4][5]. The ratios of local eigenvalues are calculated using the fractional function $f(x_1, x_2) = x_1/x_2$. Some experimental thresholds are used to distinguish five kinds of motion discontinuities as proposed in [4] and shown in Fig. 6 (c). Since the fraction function has no convergence region (within $[-\infty, +\infty]$), there is no theoretical method to determine the variable scopes of the thresholds. Another normalization method is proposed in [5] by utilizing the exponential function $E(x) = e^{-C/x}$ as shown in Fig. 5 (a). It can be observed that the convergence region of $E(x)$ is within $[0, \quad 1000]$ for $x \geq 0$. In order to use the maximum value (i.e., 1) as the threshold for confidence measurements, the value of the normalization parameter C should be heuristically determined ($C = 5$ in Fig. 5 (a)). Thus, the motion field discontinuity interpretation is inaccurate due to the application-dependent parameter used in this approach.

3.4.2. Our proposed confidence measurements

In order to address the inaccurate thresholding in the previous works, new normalization function should be conducted, whose convergence region needs to be shorter than the uncertain scopes of confidence measurements. Therefore, we exploit *hyperbolic tangent* function $T(x) = \frac{e^{-x} - e^{x}}{e^{-x} + e^{x}}$, for $x \geq 0$, to generate reliable confidence measurements as shown in Fig. 5 (b). The convergence region of $T(x)$ is very short within $[0, \quad 3]$ as compared with those of the functions



(a) Taxi (9^{th} frame)

(b) Kühne *et al.*'s method [5]

(c) Middendorf *et al.*'s method [4]

(d) Our method

Figure 6: Motion field discontinuity interpretations on the 9^{th} frame of "Taxi" sequence. In sub-figure (b), black and white colors are used to represent optical flow and no flow fields, respectively. For representing more categories of motion vectors, five gray levels are used in sub-figure (c): dominant gradient direction-perpendicular (DGD-PP) in black, optical flow discontinuities (OFD) in dark gray, dominant gradient direction-optical flow (DGD-OF) in medium gray, neutral OF (NOF) in light gray and regular optical flow (ROF) in white; in sub-figure (d), four gray levels are used in our method: FF in black, NOF in darker gray, OFD in light gray and NF in white.

in Section 3.4.1; thus, it provides accurate thresholding for the value of $T(x)$, where the variant x is constituted by the eigenvalues of the local pixel. The belonging of the local pixel will be determined into one of four kinds of motion fields according to the minimum value 0 or maximum value 1 of $T(x)$, which are defined as:

- *Full flow (FF)*: If $rank(\mathbf{J}) = 2$ or $rank(\mathbf{J}_\omega) = 2$, or $T(x_1)$ is near to 0 (where $x_1 = |\lambda_2 - \lambda_3|$), this indicates that a structure containing grayvalue changes in two directions, i.e., corner area, and moves at a constant speed, and the real motion can be calculated in this case.
- *No optical flow (NOF)*: If the condition number is near to infinity (∞), or all three eigenvalues equal to zero, i.e., $T(x_2)$ is near to 0 (where $x_2 = |\lambda_1 - \lambda_3|$), this indicates a homogeneous local area; thus, no motion can be detected.
- *Normal flow (NF)*: If $T(x_{31})$ is near to 0 (where $x_{31} = |\lambda_3|/|\lambda_1 + \lambda_2 + \lambda_3|$), and $T(x_{32})$ is near to 1 (where $x_{32} = |\lambda_2 - \lambda_3|$), this indicates that the gray-value changes will happen in one direction only, i.e., edge area, and this is the well-known aperture problem.
- *Optical flow discontinuity (OFD)*: If all three eigenvalues are greater than zero, i.e., $T(x_4)$ is near to 0 (where $x_4 = |\lambda_3|/|\lambda_2|$), the local fitting of the optical flow model will fail, this indicates the discontinuity along the border of conjoint areas with different motions.

4. EXPERIMENTAL RESULTS

Simulation results on part of the ninth frame of "Rotating Rubik Cube" have been presented in Fig. 3. It is quite clear that the proposed unequal weights yields much better classification results on motion discontinuities (i.e., corners and edges) compared with that of conventional structure tensor approach [3].

Test sequence "Hamburg Taxi" is also chosen because the classification resulting from our approach can be compared with the ones provided in the previous methods as illustrated in Fig. 6. Sub-figure (a) shows the original frame. Experimental results from Kühne *et al.*'s method [5] are given in sub-figure (b), and note that some boundaries around the moving objects are both unconnected and inaccurate. Although results from Middendorf *et al.*'s method [4] demonstrated in sub-figure (c) have enclosed boundaries, they are not aligned with the actual edges of the objects. Significant improvement on boundary accuracy of video objects resulted from our method are illustrated in sub-figure (d). Furthermore, their backgrounds (i.e., those homogeneous regions as indicated by the white color) are much more clean than the ones from other approaches.

5. CONCLUSIONS

Owing to unsatisfied classification for tensor-based motion field discontinuity presented in the previous works, a much more accurate approach is proposed with three novelties as follows. First, adaptive scale selection steered by the condition number for the spatio-temporal filtering is provided to match multiple object motions in input image sequences. Second, more accurate motion discontinuity interpretation is obtained by implementing unequally weighted structure tensor. Third, a new normalization function is developed to facilitate accurate thresholding for confidence measurements. Experimental results show that the performance of our scheme is much better than that of some previous works on the aspect of improving the accuracy of motion field discontinuity classification.

Some aspects of the proposed method need to be further investigated. Optical flow estimation will be implemented using our scheme, as well as its comparison with the results of previous leading algorithms. Another work is to conduct *anisotropic* spatio-temporal filtering instead of the *isotropic* one. The anisotropic filters are capable of preserving sharp edges/local direction patterns by automatic scaling of local areas and iterative refinement of the filtering results. They are also less sensitive to the noise than isotropic filters; therefore, the problems of exploiting anisotropic filters into our scheme requires further investigation.

6. REFERENCES

[1] B. K. P. Horn and B. G. Schunk, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–204, 1981.

[2] C. Zetzsche and E. Barth, "Direct detection of flow discontinuities by 3D curvature operators," *Pattern Recognition Letters*, vol. 12, pp. 771-779, 1991.

[3] B. Jähne, H. Haussecker and P. Geissler, *Handbook of Computer Vision and Applications*, Academic Press, 1999.

[4] M. Middendorf and H.-H. Nagel, "Estimation and interpretation of discontinuities in optical flow fields," in *Proc. Eighth IEEE Int. Conf. Computer Vision (ICCV2001)*, vol. 1, pp. 178-183, 2001.

[5] G. Kuhne, J. Weickert, O. Schuster and S. Richter, "A tensor-driven active contour model for moving object segmentation," in *Proc. Int. Conf. Image Processing*, , vol. 2, pp. 73-76, Oct. 2001.