MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

13—12 Determining Temporally Evaluated Optical Flow

Ryuzo Okada Yasuhiro Taniguchi Kazunori Onoguchi
Multimedia Laboratory, Corporate Research & Development Center, Toshiba Corp. *

Abstract

The existing methods to extract optical flow measure the distance that each pixel moves between a pair of successive image frames. It is however difficult to extract accurate optical flow when the motion in the image is very large or small. In order to deal with this problem, we propose *temporally evaluated optical flow*, which we compute by measuring the time required for each pixel to move a predefined distance using temporally consistent shift of each pixel in an image sequence captured at a high frequency. Experimental results show that the proposed method is robust and accurate compared to the typical methods.

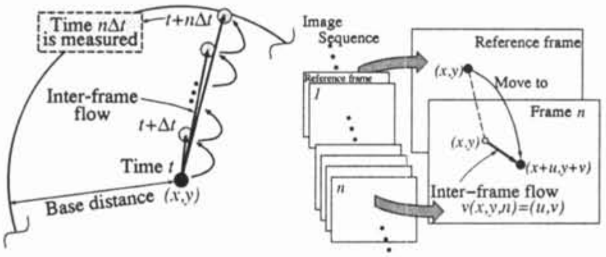
1 Introduction

Optical flow is usually measured in terms of the movement of each pixel between a pair of successive image frames[1]–[6]. Since the distance of the move is measured, we call the optical flow *distance evaluated optical flow*. As is well known, it is difficult to extract accurate optical flow when the motion in the image is large. When the motion is small, it is also difficult to detect accurate optical flow since there is almost no variance in the image, and then the flow vectors are detected to be zero.

If we use an image sequence captured at a sufficiently high frequency, all the flow vectors will appear to be almost zero. Although employment of such an image sequence eliminates the necessity of computing a large flow vector[7], the resulting accuracy would not be desirable since the problem of computing small flow vectors still remains.

To cope with the problem of computing small flow vectors, Imiya et. al. utilized the randomized sampling and voting process of the constraint equations of optical flow derived at several frame intervals[8]. Although the employment of a large frame interval make it possible to compute small flow vectors, it is difficult to select the suitable frame interval which is scene dependent.

In order to deal with the problems respecting the *distance evaluated optical flow*, we propose a novel method to extract optical flow using an image sequence captured at a high frequency, in which we measure the time taken for each pixel to move a



(a) Measuring the time (b) Inter-frame flow

Figure 1: Basic idea of *T-Flow*

predefined *base distance* (see Figure 1(a)). In contrast with *distance evaluated optical flow*, we call the extracted optical flow *temporally evaluated optical flow*, or simply *T-Flow*.

Mori et. al.[9] have independently applied the idea of measuring the time taken for each pixel to shift a fixed distance to depth extraction using a camera moving with a constant velocity vertical to the optical axis. However this method with one-dimensional search on the epipolar line cannot be applied to optical flow extraction requiring two-dimensional search.

It should be noted that *T-Flow* is not a simple extension of their method, and we further propose a technique to increase the robustness of optical flow extraction, in which we examine the reliability of an estimated movement of each pixel by the consistency as a trajectory.

2 Temporally evaluated optical flow

We first measure the movement of each pixel between a reference frame and a subsequent frame as a flow vector and call it an inter-frame flow (see Figure 1(b)). Then, we measure the time required for the magnitude of each inter-frame flow to reach the *base distance* using only reliable inter-frame flows, and finally compute the flow vector by dividing the *base distance* by the measured time.

In order to extract robust optical flow, it is important to accurately measure the time, which is realized on the basis of following two features.

1. We have only to compute the inter-frame flow whose magnitudes are close to the *base distance*. In other words, an appropriate setting of the *base distance* enables us to eliminate the necessity of computing large or small flow vectors.
2. The reliability of inter-frame flow is examined by the consistency of a time series of inter-frame

*Address: 1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan E-mail: ryuzo.okada@toshiba.co.jp

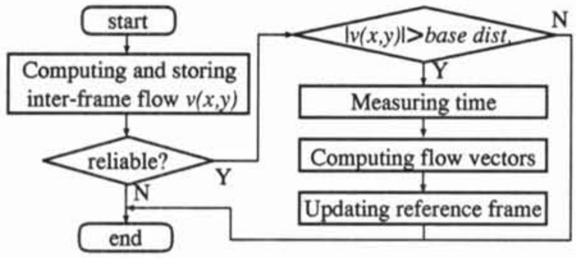


Figure 2: Procedure for each pixel

flows based on the characteristic of an image sequence captured at a high frequency. Namely, many image frames are captured within an interval short enough to approximate the motion of each pixel to be constant.

2.1 Procedure of extracting *T-Flow*

Initially the first image frame is set to be the reference frame for all the pixels, and the *base distance* is set to be a predefined value (described in section 2.1.1).

In each subsequent frame, the following procedures are performed for each pixel (see Figure 2). First, the inter-frame flow is computed between the reference frame and the current frame, and the data respecting the inter-frame flow are stored. Arbitrary method is available to compute the inter-frame flow and we use one of the most conventional methods, in which an image subregion corresponding to an reference subregion is searched for based on the sum of absolute difference (SAD). Next, the reliability of each inter-frame flow is examined using the stored data of inter-frame flows (described in section 2.1.2). If the inter-frame flow is reliable and its magnitude reaches the *base distance*, the time required for the magnitude of inter-frame flow to reach the *base distance* is measured(described in section 2.1.3), and the flow vectors are computed using the measured time(described in section 2.1.4). Furthermore the reference frame with respect to the concerning pixel is updated to be the current frame.

2.1.1 Determining the *base distance*

The *base distance* should be in a range where the inter-frame flows can be accurately computed and a time series of data respecting inter-frame flows can be obtained sufficiently before the magnitude of inter-frame flow reaches the *base distance*. Considering these factors, we set it to be 2 pixel.

2.1.2 Computing reliability

A definition of the reliability utilizing the direction of inter-frame flow is described in this section.

Let $s(x, y, n)$ denote the SAD of the inter-frame flow at a pixel, (x, y), in a image frame, n, which we call the mismatching score. The low mismatching score signifies that the inter-frame flow is reliably estimated. Before the magnitude of inter-frame flow reaches the *base distance*, D, we regard the inter-frame flow with the lowest mismatching score in the time series of inter-frame flows computed at the concerning pixel,(x, y), as the correctly estimated inter-frame flow $\boldsymbol{v}_r(x, y)$. When the magnitude of inter-frame flow is close to D, the inter-frame flow whose direction is close to that of $\boldsymbol{v}_r(x, y)$ is considered to have high reliability.

Thus, we define an indicator, $r(x, y, n)$, of reliability of the inter-frame flow, $\boldsymbol{v}_D(x, y, n)$, whose magnitude is close to D as an angle between the two vectors $\boldsymbol{v}_D(x, y, n)$ and $\boldsymbol{v}_r(x, y)$:

$$r(x, y, n) = \left| cos^{-1} \frac{\boldsymbol{v}_r(x, y) \cdot \boldsymbol{v}_D(x, y, n)}{|\boldsymbol{v}_r(x, y)||\boldsymbol{v}_D(x, y, n)|} \right|, \quad (1)$$

If $r(x, y, n)$ exceed a threshold, $\boldsymbol{v}_D(x, y, n)$ is determined to be unreliable and is not used in the following procedures.

2.1.3 Measuring time

Measuring the time is equivalent to counting the number of frames from the reference frame to the frame in which the magnitude of inter-frame flow reaches the *base distance* D. However, it is difficult to determine such a frame in an image sequence captured at a high frequency because the magnitude of inter-frame flow is close to D in many frames. We choose the frame, $k(x, y)$, in which the extracted inter-frame flow has the lowest mismatching score of the time series of inter-frame flows with the magnitude close to D at the concerning pixel (x, y):

$$k(x, y) = \underset{n}{arg} \underset{n \in N_2(x,y)}{min} s(x, y, n), \quad (2)$$

where $N_2(x, y)$ represents a set of frames in which the magnitudes of inter-frame flow are close to D.

2.1.4 Computing flow vectors

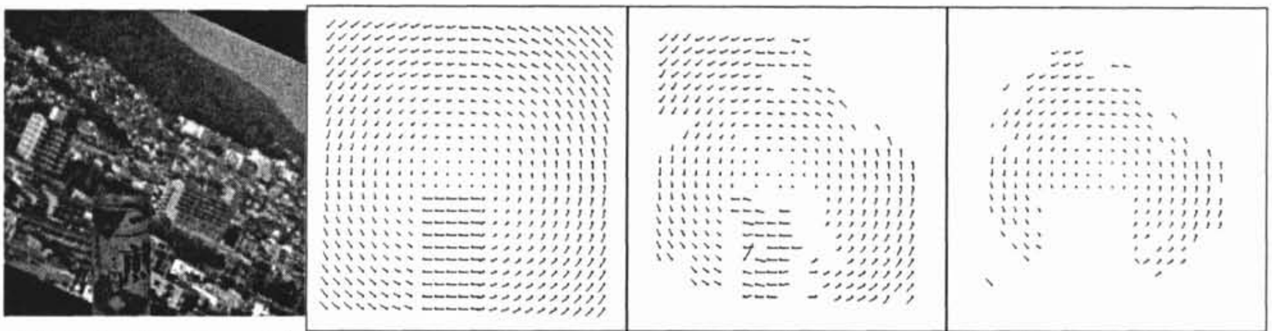
The magnitude of a flow vector is calculated by dividing the *base distance* by the measured number of frames, and the direction is determined to be that of the inter-frame flow in the frame $k(x, y)$.

However, the number of frames is measured by an integral number of frames. In order to increase the accuracy of time measurement, we fit a quadratic curve to a set of data $(n, s(x, y, n))$ $(n \in N_2(x, y))$, and the frame where the magnitude of inter-frame flow reaches the *base distance* is determined to the one minimizing the quadratic curve.

The direction is also determined by a pixel order using the SAD-based method without taking subpixel information into account. Likewise, we fit a quadratic curve to a set of data $(d, s(d; x, y, k(x, y)))$, where d represents a direction close to that of $\boldsymbol{v}_D(x, y, k(x, y))$ and $s(d; x, y, k(x, y)))$ is a SAD calculated for a direction d in the frame $k(x, y)$ at the concerning pixel (x, y), and the direction of inter-frame flow is determined to the one minimizing the quadratic curve.

3 Experimental Results

We have conducted experiments in order to compare the robustness and accuracy of *T-Flow*



(a) 300th frame($\sigma = 5.0$) (b) Correct flow (c) *T-Flow* (d) Phase-based method

Figure 3: Synthesized image and extracted optical flow (Flow vectors in (b) and (c) are magnified 30 times.)

	Method	no noise		noise $\sigma = 5.0$		noise $\sigma = 10.0$	
		error[deg.]	density[%]	error[deg.]	density[%]	error[deg.]	density[%]
1000 frame/sec.	*T-Flow*	0.48	73.2	0.54	74.4	0.71	76.3
	gradient[2]	1.61	65.2	2.08	65.2	2.93	65.7
	SAD[4]	4.11	73.1	7.35	79.4	13.87	88.7
	phase[5]	0.51	81.6	0.84	78.3	2.37	26.1
30 frame/sec.	gradient	1.96	40.4	1.98	40.6	2.87	56.0
	SAD	1.02	70.1	1.22	76.1	2.90	86.6
	phase	0.19	59.6	0.25	59.0	2.40	57.0

Table 1: Error and density of extracted optical flow

with those of *distance evaluated optical flow* using a synthesized image sequence and real image sequences. We have implemented three typical methods to compute *distance evaluated optical flow*, which are the gradient-based method[2], the SAD-based method[4], and the phase-based method[5].

Synthesized image sequence. The synthesized image sequence consists of real images of a can and those of a view of a mountain and a city (see Figure 3(a)). The former is the frontal object translating in the right direction and the latter is the background rotating around the center of image. The synthesized flow field is shown in Figure 3(b). There is a motion boundary between the frontal object and the background, and the brightness of frontal object changes by 0.1% of its original brightness in each frame. Furthermore, we add gaussian noise with standard deviation of 5 and 10 to the brightness as an image noise.

The error of flow vector is represented by an angle between the computed flow vector, $(u, v, \delta t)$, and the corresponding correct flow vector, $(u_c, v_c, \delta t)$, expressed in the spatiotemporal space, where δt represents a frame interval[6]. The error shown in Table 1 is the average of all the errors in an image.

In the experiment using the image sequence with 1000 frames/sec. (see the upper table of Table 1), the errors of *T-Flow* are the smallest of all the results with any image noises. Note that the results computed with the SAD-based method include large errors because the resolution of computed flow vectors is one pixel.

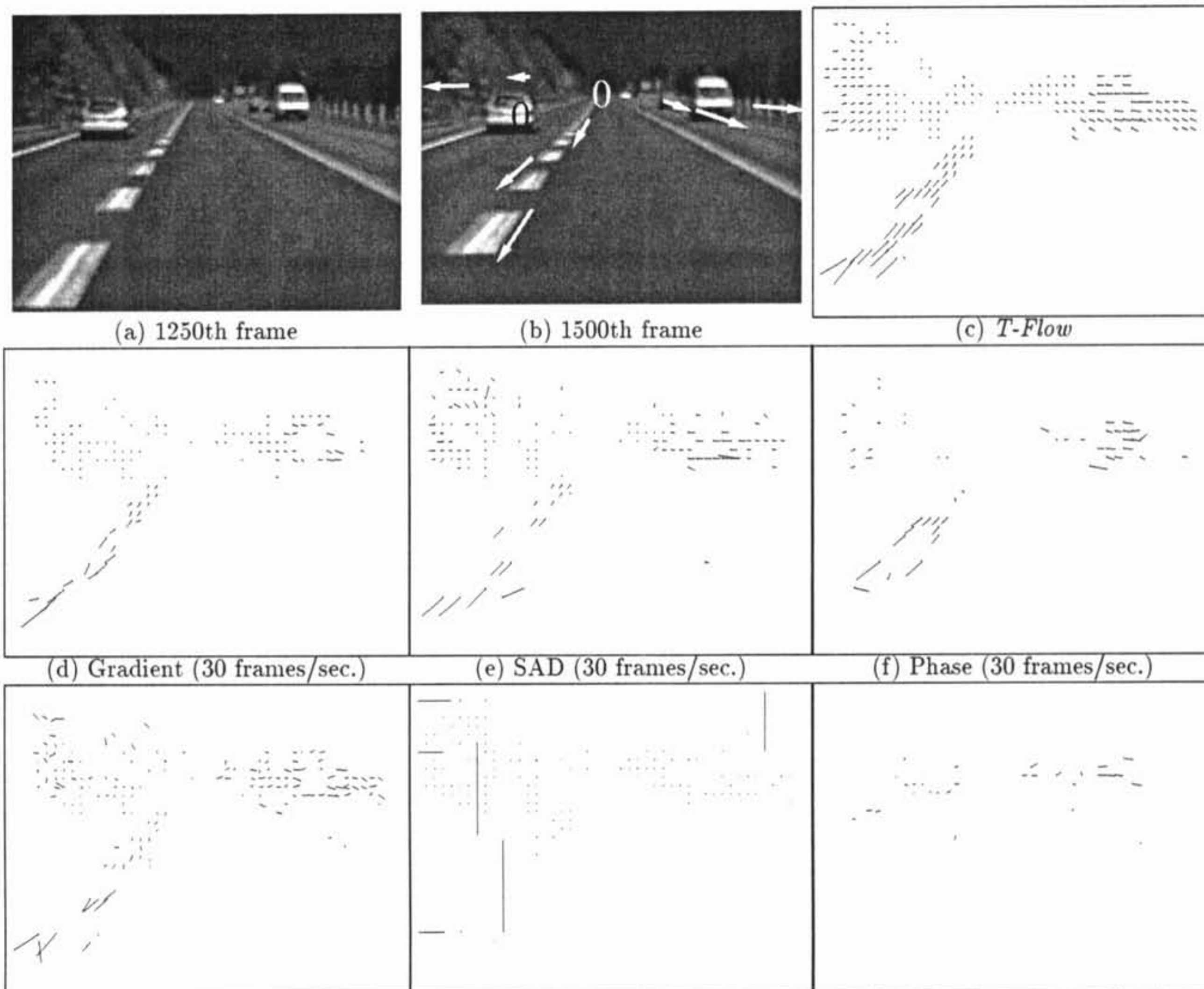
Next, we use the image sequence with 30 frames/sec. in order to simulate the video rate of NTSC for which the typical methods are designed. In this experiment (see the lower table of Table 1), the results computed with the phase-based method include the smallest error for the image sequence with the small image noise. However, large flow vectors, for example the flow vectors in the frontal object, are not calculated (see Figure 3(d)), and moreover the error for the image sequence with the large image noise is large.

The results shows that the proposed method is capable of accurately computing both small and large flow vectors, and also the most robust against the image noise.

Real image sequence. Figure 4(a) and (b) show real images captured at 1000 frames/sec. Computed *T-Flow* is shown in Figure 4(c), and the results of typical methods with frame rate of 30 frames/sec. and 1000 frames/sec. are shown in (d)–(f) and (g)–(i), respectively. *T-Flow* is apparently more accurate and robust than the results obtained by the typical methods.

4 Conclusions

In this paper, we proposed a method of extracting optical flow by measuring the time required for each pixel to move the predefined *base distance* using an image sequence captured at a high frequency, instead of measuring the distance that each pixel moves for a fixed frame interval. In this method, there are two advantages contributing to the robustness and the accuracy, which are (i) measuring the time eliminates the necessity of computing large or



(a) 1250th frame (b) 1500th frame (c) *T-Flow*

(d) Gradient (30 frames/sec.) (e) SAD (30 frames/sec.) (f) Phase (30 frames/sec.)

(g) Gradient (1000 frames/sec.) (h) SAD (1000 frames/sec.) (i) Phase (1000 frames/sec.)

Figure 4: Extracted optical flow using a real image sequence (The original images are shown in (a) and (b), and the arrows shown in (b) outline motions in the image and '0's represent no motion. The results shown in (d)–(f) and (g)–(i) are computed using the image sequence with 30 frames/sec. and 1000 frames/sec., respectively. The flow vectors shown in (c), (g)–(i) are magnified 75 times, and those of (d)–(f) are magnified 2.5 times.)

small flow vectors and (ii) only the temporally consistent motion is used for measuring the time based on the analysis of a time series of data respecting the movement of each pixel.

References

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

[2] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. DARPA Image Understanding Workshop*, pp. 121–130, 1981

[3] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *IJCV*, 2: pp. 283–310, 1989

[4] H. Inoue, T. Tachikawa, and M. Inaba, "Robust Vision System with a Correlation Chip for Real-time Tracking, Optical Flow and Depth Map Generation," *Proc. R & A*, pp. 1621–1626, 1992

[5] D.J. Fleet and A.D. Jepson, "Computation of Component Image Velocity from Local Phase Information," *IJCV*, 5:1 pp. 77–104, 1990

[6] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Systems and Experiment Performance of Optical Flow Techniques," *IJCV*, 12:1, pp. 43–77, 1994

[7] Y. Nakabo, M. Ishikawa, H. Toyoda and S. Mizuno, "1ms Column Parallel Vision System and Its Application of High Speed Target Tracking," *Proc. R & A*, pp. 650-655, 2000

[8] A. Imiya and K. Iwawaki, "Detection of Subpixel Flow Field," *Proc. IPSJ SIG-CVIM*, 2000-CVIM-121, pp. 17–24, 2000

[9] T. Mori and M. Ymamoto, "A Dynamic Depth Extraction Method," *Proc. ICCV*, pp. 672-676, 1990