Fast digital optical flow estimation based on EMD

Mickael Quelin
University of Wollongong, mq102@uow.edu.au

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Fast digital optical flow estimation based on EMD

Abstract

This paper presents an optical flow estimation technique based on the so called Elementary Motion Detector (EMD). The aim is to provide a fast but not necessarily very accurate system to be used for specific post processing purposes. This model uses a low complexity algorithm for detecting motion in four directions by identifying specific motion templates. By extending the motion templates to different scales of the input video, an Elementary Velocity Detector (EVD) is created. This one can save computation time by estimating different speeds in parallel. Information from the EVD outputs is then used to generate an estimate of the optical flow along edges and other intensity transitions. Experimental results are presented, which show the effectiveness of the proposed model. Furthermore, the robustness of the model can be enhanced by using pre-filtering.

Keywords

estimation, emd, fast, optical, digital, flow

Disciplines

Physical Sciences and Mathematics

FAST DIGITAL OPTICAL FLOW ESTIMATION BASED ON EMD

Mickael Quelin, Abdesselam Bouzerdoum, Son Lam Phung

ICT Research Institute
University of Wollongong, Northfields Avenue, NSW 2522 Wollongong, Australia

ABSTRACT

This paper presents an optical flow estimation technique based on the so called Elementary Motion Detector (EMD). The aim is to provide a fast but not necessarily very accurate system to be used for specific post processing purposes. This model uses a low complexity algorithm for detecting motion in four directions by identifying specific motion templates. By extending the motion templates to different scales of the input video, an Elementary Velocity Detector (EVD) is created. This one can save computation time by estimating different speeds in parallel. Information from the EVD outputs is then used to generate an estimate of the optical flow along edges and other intensity transitions. Experimental results are presented, which show the effectiveness of the proposed model. Furthermore, the robustness of the model can be enhanced by using pre-filtering.

Index Terms— insect vision, optical flow, elementary motion detector (EMD), elementary velocity detector (EVD), Horridge template model, Reichardt correlator

1. INTRODUCTION

Motion detection is an important cue for interpreting the visual field. Its most detailed representation is displayed with vectors as an optical flow. Recently, new evaluation methods have been proposed to compare and improve the latest computer vision optical flow estimators [1]. Numerous effective methods have been proposed using gradient-based and correlation models as well as token-based models [2]. However simple models, inspired from insect vision, have been suggested to lead to specific application models without need for accurate velocity outputs.

Insect-based vision systems have been investigated with great interest as their low complexity can contribute to the creation of fast and robust visual navigation systems. Today, the application of computer vision in real-time embedded systems is considered to be a promising technology, in part because of its low hardware cost.

Since 1956, when Reichardt correlator was proposed [3], many models have been designed to exploit or improve this Elementary Motion Detector (EMD). Its limitation to estimate velocity of motion have been discussed in [4] and else-

where. More recent models also successfully improved its performance for velocity estimation [5, 6]. Simple application models using EMDs can be found in, e.g., odometers or moving object detectors [7, 8]. The aim often is to lead to autonomous robot navigation with key functions such as collision avoidance. A good example is Harrison's radially oriented EMD architecture [9]. Another original solution towards visual field interpretation is Tianguang et al. receptive fields [10], which demonstrates the high efficiency of optical flow estimation for post processing. Generally these designs have been successfully implemented, often in real-time, in a confined environment. Their main limitation is lack of adaptation of the EMD response to changes in contrast and luminance and the computational bottleneck, if a higher resolution image or frame rate is utilized.

Proposed in the early 1990's, the template model [11] simplifies the Reichardt model in order to be implemented with digital hardware. We use this model for our design in order to offer an alternative fast technique to Reichardt correlator to estimate motion velocity.

In this paper, we propose an optical flow estimator based on insect vision principles, a model which is suited for a digital hardware architecture. First, a brief review of the template model is presented, then its extension to 2-D video is discussed. Section 3 presents the Elementary Velocity Detector (EVD) principle using a new template model that we call velocity template. The experimental results of the estimator in different scenarios, using pre-filtering, are presented in section 4. Finally this paper concludes with a summary and future work.

2. TWO DIMENSIONAL TEMPLATE MODEL

The template model can be seen as a digital version of the Reichardt EMD. It is an empirical model which extracts from intensity changes (jumps), an indicator of directional motion. Figure 1 presents the EMD of the template model where the visual field is spatially and temporally sampled. The variable $\Delta\varphi$ indicates the distance or spatial period between the photoreceptors. A filter produces information on changes in the input which is sampled by a switch. A threshold function then splits it into 3 states: -1 if the intensity is decreasing, +1 if it is increasing and 0 otherwise. The current states of the two

channels are transmitted to the bottom row of the 2-by-2 output matrix and the previous states (defined by a delay z^{-1}) are transmitted to the first row of the matrix.
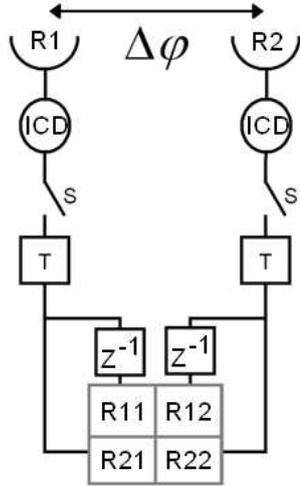

Fig. 1. Template model scheme. R1 and R2: Photoreceptors, ICD: Intensity Change Detection, T: Threshold function, S: Switch, z^{-1}: Delay function, $\Delta\varphi$: Spatial period.

Over the 81 possible templates of size 2-by-2, 8 of them indicate the existence of motion to a specific direction (from one receptor to another). Information about the polarity of the edge transition as well as the start and end of a phenomenon are also coded (for a list of templates see [12]).

Using the template model to process 2D frames of a video input requires the definition of the spatiotemporal sampling of the EMD and the extension of the one-dimensional (1D) direction sensitivity to 2D direction sensitivity. In order to retain all the details that the input signal provides, it is logical to match the input spatiotemporal resolution to the EMD. Therefore, we choose $\Delta\varphi$ as the distance between two pixels and delays z^{-1} as the frame rate of the video. The intensity change detection is, in our design, implemented as a high-pass (HP) filter by simply subtracting the previous intensity input from the current one, and thus the sign of the filter output codes the evolution of the intensity.

To detect all kinds of motion in the video, it is often proposed to combine two EMD architectures in an orthogonal fashion, such as sensing in the x and y directions, so that motion in all directions can be reconstructed. Jun et al. [8] and Tianguang et al. [10] proposed two 2D EMDs using a 4-pixel X or 3-pixel Γ shape, respectively. The association of motion to the output they proposed is not completely consistent in terms of location. Here, we propose another 5-pixel + shape EMD which has the advantage of linking motions to the starting point which is an actual pixel position in the input frame (see Fig. 2).

Finally, the threshold value has a significant influence on the output and its optimum value depends on changes in contrast and luminance of the input signal. In addition, the ve-
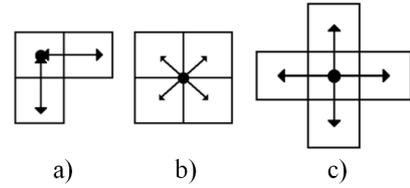
Fig. 2. EMD architectures. The dots represent location to where motions detected are linked. a) Γ, b) X, c) +.

locity sensitivity of the Reichardt model is lost here by digitizing the output into 3 states. Moreover, the spatial distance between associated templates, which code the beginning and ending of a transition, depends on both the spatial length of the transition and the velocity of the motion. Therefore, it cannot be used in the estimation of the velocity without extra information.

3. ELEMENTARY VELOCITY DETECTOR (EVD)

Use of the template model for velocity estimation has been realized in the past by tracking templates and linking them over space and time [12]. Such a method gives an accurate estimation but as the tracking process depends on the searching for templates step and on the number of features to track, the algorithm processing time will partially depend on the input nature. Because the proposed system aims to be implemented in real-time, we need a solution which uses a defined time to process any input frame.

3.1. Template for Velocity

To be able to estimate velocity, the matching of at least two templates in space and time needs to be achieved. In order to remove any tracking process or search space, we have designed a new family of velocity dependent templates by linking two motion templates in space and time as described in Table 1. To achieve such a match the elementary detector, shown in Fig. 1 needs to be revised by implementing an extra photoreceptor and an extra delay. This feature can be implemented in our four directional detector by using a 9 pixels + shape. Those velocity templates however only match motion templates over one pixel space difference and one frame temporal difference. In other words, this architecture is sensitive to an elementary velocity of one pixel per frame (1 ppf).

3.2. Multi-scale EVDs

In order to use the same EVD model to estimate other velocities, a multi-scale architecture can be set up (see Fig. 3). As mentioned before, to estimate 1 ppf velocity, a photoreceptor is defined as one pixel. However, to estimate higher odd velocities (3 ppf, 5 ppf etc.), we propose to average a section of pixels and to consider it as one receptor. Thus the actual

Table 1. List of velocity templates.

| Polarity	Motion to right		Motion to left	
	code	Template	code	Template
Dark-to-bright	A	$\begin{matrix} +1 & 0 & - \\ +1 & +1 & 0 \\ - & +1 & +1 \end{matrix}$	E	$\begin{matrix} - & 0 & +1 \\ 0 & +1 & +1 \\ +1 & +1 & - \end{matrix}$
	B	$\begin{matrix} +1 & +1 & - \\ 0 & +1 & +1 \\ - & 0 & +1 \end{matrix}$	F	$\begin{matrix} - & +1 & +1 \\ +1 & +1 & 0 \\ +1 & 0 & - \end{matrix}$
Bright-to-dark	C	$\begin{matrix} -1 & 0 & - \\ -1 & -1 & 0 \\ - & -1 & -1 \end{matrix}$	G	$\begin{matrix} - & 0 & -1 \\ 0 & -1 & -1 \\ -1 & -1 & - \end{matrix}$
	D	$\begin{matrix} -1 & -1 & - \\ 0 & -1 & -1 \\ - & 0 & -1 \end{matrix}$	H	$\begin{matrix} - & -1 & -1 \\ -1 & -1 & 0 \\ -1 & 0 & - \end{matrix}$

Fig. 3. Multi-scale architecture.

distance between the motion templates will be changed even though the velocity template to be detected stays the same. Therefore the velocity estimated is directly linked to the number of pixel defining a detector. Location of motion is then linked to the central pixel of the starting receptor (left or right depending on the template detected) and can be mapped into the input frame. Note that the averaging is necessary as the pixels in the center of the blocks are not adjacent and thus an edge transition might not be detected or mistakenly detected if the surrounded pixels are ignored.

To process a video input, a four directional EVD needs to be implemented at each pixel location. The velocity estimated will be defined by the scale of the photoreceptor. In order to generate an optical flow output, the video input needs to be processed by an EVD architecture working at different scales. Those scales can easily be implemented in parallel. At each pixel, one or more velocity values for horizontal and vertical directions can be linked. The priority here is given to the highest velocity.

4. RESULTS

Figures 4 and 5 present the results of the optical flow estimator using seven EVD scales (1 to 13 ppf). To make the sensor more robust to noise, input videos have been convolved with a Gaussian window (a typical size used was 11 by 11 with a σ of one sixth of the window size). This also improves the EVDs outputs by smoothing the edge transitions. The threshold of the EVDs is adjusted to be sufficiently large, to be sensitive to changes of intensity of moving edges and to avoid detecting noise (a typical value used was 5 for a 255 gray level input).

Figure 4 shows the results for a static camera video using a resolution of 120 by 160 and at 30 frames per second.
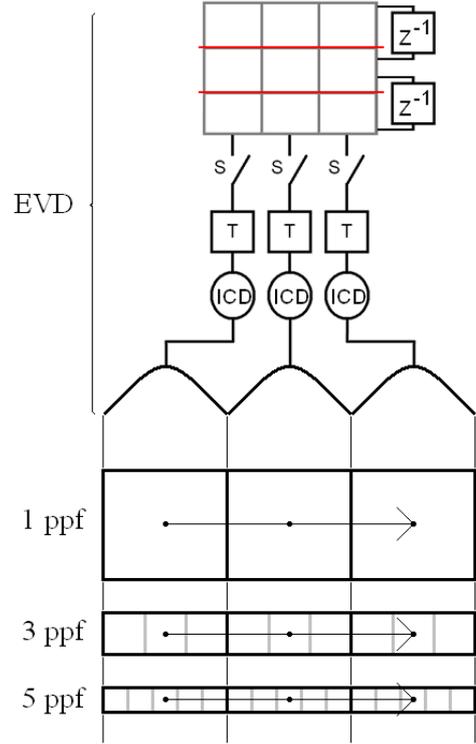
The black cup held by a hand is moved towards the camera. Results show that the association of horizontal and vertical velocities detected matches consistently with the direction of motion.

Figure 5 shows the results for a non-static camera video using a resolution of 256 by 256 at 30 frames per second. The camera is moving forward and is being overtaken by the car on the left. Details of the scenery such as white marking lines and road signs are also detected.

This model is not expected to provide precise dense optical flow outputs, therefore comparison with state of the art estimators has not been realized at this stage of the work. However results obtained using accelerating objects are realistic, with vectors length growing over time. This demonstrates that the system can be used for further postprocessing. Note that results are given with relatively low resolution and frame rate considering today video camera hardware.

5. CONCLUSIONS

In this paper, the template model has been extended to process video frames. The concept of the Elementary Velocity Detector using velocity templates has then been proposed. This one has been implemented on a multi-scale architecture and has been successfully tested. Our model estimates velocities quickly and in parallel. The simulations and experiments demonstrate that this elaborated model provides consistent and promising results. However, with more realistic
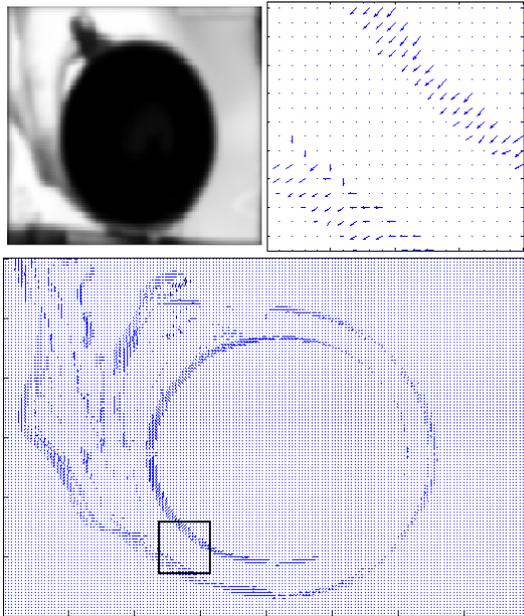
Fig. 4. Static camera results. top left: input frame, top right: zoom-in output frame, bottom: output frame.
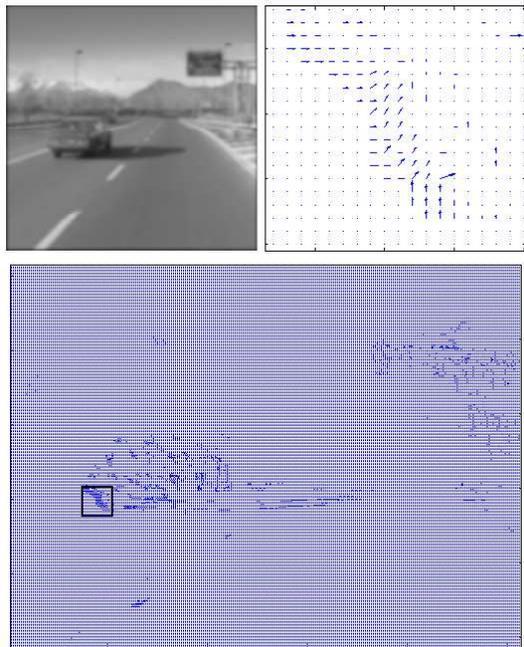
Fig. 5. Non-static camera results. top left: input frame, top right: zoom-in output frame, bottom: output frame.

video inputs, i.e. complex backgrounds or moving cameras, a further adaptation to changes in contrast and luminance would be required.

6. REFERENCES

[1] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," in *Proceedings of the IEEE Int. Conf. Computer Vision*, 2007, pp. 1–8.

[2] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog vlsi architectures for motion processing: from fundamental limits to system applications," in *Proc. IEEE*, 1996, vol. 84, pp. 969–987.

[3] B. Hassenstein and W. Reichardt, "Functional structure of a mechanism of perception of optical movement," in *Proc. Int. Cong. Cybern.*, Namur, 1956, pp. 797–801.

[4] Ron O. Dror, David C. O'Carroll, and Simon B. Laughlin, "Accuracy of velocity estimation by reichardt correlators," *J. Opt. Soc. Am. A*, vol. 18, no. 2, pp. 241–252, 2001.

[5] J. M. Zanker, M. V. Srinivasan, and M. Egelhaaf, "Speed tuning in elementary motion detectors of the correlation type," *Biol. Cybern.*, vol. 80, no. 2, pp. 109–116, 1999.

[6] Olena Riabinina and Andrew O. Philippides, "A model of visual detection of angular speed for bees," *J. Theor. Biol.*, vol. 257, no. 1, pp. 61–72, 2009.

[7] Fumiya Iida and Dimitrios Lambrinos, "Navigation in an autonomous flying robot by using a biologically inspired visual odometer," in *Proc. SPIE Sensor Fusion and Decentralized Control in Rob. Syst. III*, 2000, vol. 4196, pp. 86–97.

[8] Yin Jun, Li Dong-Guang, Fang Hui-Min, and Liu Zhi-Feng, "Moving objects detection by imitating biologic vision based on fly's eyes," in *Proc. ROBIO 2004 IEEE Int. Conf. Rob. and Biomim.*, 2004, pp. 763–766.

[9] R. R. Harrison, "A biologically inspired analog ic for visual collision detection," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 52, no. 11, pp. 2308–2318, 2005.

[10] Zhang Tianguang, Wu Haiyan, A. Borst, K. Kuhnlenz, and M. Buss, "An fpga implementation of insect-inspired motion detector for high-speed vision systems," in *Proc. IEEE Int. Conf. Rob. and Autom.*, 2008, pp. 335–340.

[11] G. A. Horridge, "A template theory to relate visual processing to digital circuitry," in *Proc. R. Soc. Lond.*, 1990, vol. B 239, pp. 17–33.

[12] X.T. Nguyen, A. Bouzerdoum, and R.E. Bogner, "Backward tracking of motion trajectories for velocity estimation," in *Proc. 1996 Australian and New Zealand Conf. Intelligent Information Systems*, 1996, pp. 338 – 341.