Selective Change-Driven image processing for high-speed motion estimation

Fernando Pardo, Xaro Benavent, José A. Boluda, Francisco Vegara
Departamento de Informática - Universitat de València
Avda. Vicente Andrés Estellés s/n, 46.100 Burjassot, Valencia, SPAIN
Phone: +34 963160426 Fax: +34 963160418 E-mail: Fernando.Pardo@uv.es

Keywords: Address Event Representation AER, asynchronous pixel processing, change-driven image processing.

Abstract - Biological visual systems are becoming the most interesting source for the improvement of artificial visual systems. A biologically inspired read-out and pixel processing strategy is presented. This strategy is based on Selective pixel Change-Driven (SCD) processing. Pixels are individually processed and read-out instead of the classical approach where the read-out and processing is based on complete frames. The most interesting pixel is read-out and processed at every short time interval. The response delay using this strategy is several orders of magnitude lower than current cameras while still keeping the same, or even tighter, bandwidth requirements.

1. INTRODUCTION

Most current image and video processing applications are based on the processing of a static image stream coming from a visual system. Each image in this stream is a snapshot of the environment taken at regular intervals. Biological visual systems work in a complete different way: each sensor cell (image pixel) sends its illumination information asynchronously, making the concept of "image" or "frame" useless. Another biological feature is the space-variant cell distribution in the image plane, which reduces the amount of visual data to be transmitted and processed. The artificial strategy based on a synchronous flow of space-uniform images has a history of few decades, while the biological visual system is the result of several million years of evolution, reaching an outstanding performance compared to most artificial visual systems. Current technology makes it difficult to fabricate a complex biological visual system, but some ideas coming from the biological system can be adapted, after some changes, to improve artificial visual systems.

The space-variant nature of most biological eyes has been extensively exploited in artificial visual systems for roughly two decades. Nevertheless, little attention has been focused on exploiting the asynchronous nature of biological visual system, probably because the advantages are not worth the effort of stop working with synchronous static images. It is shown in this paper, that independent pixel processing has advantages over frame processing; moreover, its implementation is feasible with current imaging technology, taking into account the limitations in bandwidth and processing power.

1.1 Biologically inspired visual sensors

The space-variant nature of visual acuity found in the human eye has been extensively studied in the past. Some artificial sensors have been designed to have a foveal log-polar pixel distribution resembling the eye of most primates [1]. Similar pixel distribution can be modeled from conventional cameras reducing the necessity of having such sensors to emulate space-variant biological systems. This is not the case when trying to implement independent pixel sampling and processing since conventional cameras are based on constant sampling of space and time. This is the reason why, despite the fact that research on asynchronous pixel processing is at the beginning, there are almost as many sensors exploiting this characteristic than space-variant image sensors.

The studies of the visual system neural activity show that some cells respond to illumination transients in time and space (ON/OFF cells) while others have a sustained response dependent on illumination levels. This behavior has been emulated in many biomorphic sensors.

The combined transient and sustained behavior has been implemented in [2], though this approach was "frame based" instead of asynchronous as the biological visual system. The cell activity (measured as spikes) dependency on light intensity has been emulated in some sensors [3][4][5] which also worked asynchronously like the human eye; in these sensors each pixel works independently and sends spikes with a period which inversely depends on light intensity. The available output bandwidth is allocated according to pixel output demand, favoring the brighter pixels and not the most interesting ones that are usually those with a high spatial and/or temporal transient.

Some of the most interesting sensors, from the movement analysis point of view, are those presented in [6][7]. In these sensors, the spike interval is dependent on change intensity, allowing the selection of interesting areas (those with larger change). The change event signaling depends on a Temporal Contrast Threshold, which is also found in most biological vision systems. This Contrast Sensitivity Threshold has been already successfully employed to improve some movement analysis algorithms [8].

It is usual to consider in movement analysis that the most interesting pixel to process is the one with larger transient change. There are also some image sensor implementations that not only detect image changes, but also yield the pixel presenting a larger change thanks to a Winner Take All circuit (WTA) [9]. These sensors [10][11][12] avoid the readout of the complete frame, allowing a speed-up in the process of movement detection and analysis.

1.2 Change-driven image processing

The fixed spatial-temporal sampling rate of standard cameras makes it difficult to exploit many of the advantages of biological vision systems. Nevertheless, there is a feature that can be exploited for artificial image processing: it is the change driven image processing. In biological systems, illumination variations drive the movement detection; while in typical artificial systems each frame is completely processed at every time interval, even if no changes have been produced at all. Processing only those pixels that have changed, above a certain threshold, can decrease the total data to be processed while still keeping accuracy in the image processing. The benefits of this change-driven processing have been already measured [8] and some sensors have been designed to take advantage of the data reduction (image compression) [13][14].

In this article we present a further image data reduction based on Selective Change Driven (SCD) processing. This technique is especially interesting for image capture at high rates (above the standard 25/30 fps) or custom high-rate asynchronous pixel based sensors.

2. SELECTIVE CHANGE-DRIVEN PROCESSING

The integration time (shutter) of current pixel technology can be as small as few microseconds. A single pixel could deliver visual information at a rate of tenths of kHz, which is several orders of magnitude above the current usual speed (25/30 fps). Nowadays, there are some cameras that reach this speed, and the limitation is not the integration time (shutter) but the bandwidth: a 10 kHz VGA (640x480) grey-level camera would deliver around 3 Gbytes/s which is very difficult to transmit and almost impossible to process. In fact, these cameras have their own internal storage and are intended to record few seconds to be processed afterwards. The Selective Change-Driven (SCD) processing dramatically reduces the bandwidth of this kind of cameras while still keeping the advantages of such high-speed acquisition rates.

In change-driven processing, only those pixels that change, above a threshold, from frame to frame are processed. This technique may speed-up a complex motion detection algorithm up to one order of magnitude [8], but this speed-up depends on the number of effective changes from frame to frame and the data reduction is far from being enough when high-speed cameras are in use.

A further step in data reduction is to process only those changes from frame to frame that can be considered interesting. Only one pixel could be processed for every frame, in the most limited bandwidth case. The best pixel to process at each time would be that which presented the higher illumination change since it means two things: first, a large change in intensity means a fast movement around that pixel, and second, it also means that there is an object edge around that pixel. Movement and edges (high spatial/temporal frequencies) are usually the most interesting parts of a scene. If more bandwidth is available, more than one pixel could be taken from frame to frame; in this case, a list of "most wanted" pixels is prepared based on their illumination transient.

Both, change-driven processing and selective driven-processing are based on pixel processing instead of frame processing. In fact, there is no "frame" or "image" since no snapshot is taken at any time. It resembles much the biological visual system behavior, since each rod or cone cell in the human retina sends its visual information in asynchronous way just in the moment it seems interesting to do so. ON/OFF cells in the retinal neural tissue send spike signals when some spatial/temporal illumination change has been detected. Biological systems take advantage of the slow asynchronous but parallel behavior of the neural tissues, while artificial systems may take advantage of the fast synchronous but serial processing.

Selective Change-Driven (SCD) image processing reduces by several orders of magnitude the amount of data to be processed. The question is to know whether this reduction can still offer advantages, not only by reducing bandwidth and processing requirements, but also for motion analysis. In other words, would this technique deliver more accurate motion estimation while still keeping the same bandwidth and processing power? The answer to this question is shown in the experiment section.

2.1 SCD camera architecture

The proposed SCD image capture strategy works as follows: every pixel may work synchronously or not, being completely independent from the others. Every pixel has an analog memory with the last read-out value. The absolute difference between the current and the stored value is compared among all pixels in the sensor; the pixel with higher difference is selected and its illumination level and address are read-out for processing. The read-out can take place at regular intervals or asynchronously depending on the light capture cell or voting circuit. The experiments shown in next sections use a complete synchronous strategy, since it can be adapted to existing high-speed cameras and there is no change in performance.

Though there are already some sensors with similar behavior, there is a fundamental difference with large impact in image processing: every pixel has a memory of the last read-out value, while the existing sensors just signal the most interesting pixel but do not store any value. The advantage of storing the last read-out value is that the election of the most interesting pixel is based on global change along time instead of just transient. For example, a sudden change in the image (say, switching a light on) would generate many events, but only few pixels will be processed during the transient. This problem does not appear if the pixel selection is based on the change compared to last stored read-out value, since every pixel that has changed will be processed sooner or later even long after the transient has occurred.

3. EXPERIMENTAL RESULTS

Before designing a camera or sensor able to deliver at least 100.000 pixels per second (each pixel as a result of a selection process) some experiments have been carried out to measure some motion estimation algorithm accuracy.

In the proposed experiment a standard camera is compared to a selective change-driven (SCD) camera; the condition is that both have the same resolution and both deliver the same amount of visual information and thus share the same bandwidth limitation.

Let suppose a constant resolution of 320x240 pixels for both cameras. Let the standard camera image rate be 25 fps, thus the total bandwidth and processing requirement is 1.92 Mbytes/s. A SCD camera with this bandwidth would capture and deliver one pixel every

520 ns. Any integration cell would yield a very poor image if such a short shutter (integration time) were chosen. Other approach would be using a continuous cell [1] that can be accessed at any time. First we are going to show the results just supposing that it is possible to implement a SCD camera able to capture illumination at 520 ns; with present technology it is not feasible (considering image quality) but it is the first step to measure the benefits of such an image capture strategy. Afterward, more realistic implementations will show that these benefits still hold.

To measure and compare such two cameras, a synthesized scene has been virtually created, since there is no way to emulate a SCD camera using standard cameras. The synthesized scene consists of a background in which a ball moves at very high speed (more than 1500 pixel/s). The image processing algorithm calculates the center of mass of the moving ball binarizing the image to isolate the object. The scene and ball trajectory are shown in Fig. 1. The ball takes 160 ms to complete its trip.

3.1 Same bandwidth and ultra high-speed shutter

In this first experiment, the SCD camera and the standard camera are compared supposing the same bandwidth. An ultra high-speed shutter is used for the SCD camera. Such a fast shutter (around 600 ns) is out of the reach of current technology when good image quality is of concern, but it is interesting as first approach to estimate the SCD maximum performance.

Fig. 1 shows the last image and the last position of the tracked ball along with its sinusoidal trajectory from left to right. The five circles in the trajectory show the tracked positions using a standard camera. With these five points is impossible to measure or even guess the original sinusoidal trajectory of the ball. The tracked positions using a SCD camera exactly match the sinusoidal trajectory and it is impossible to distinguish it in the figure, since the error is less than one pixel.

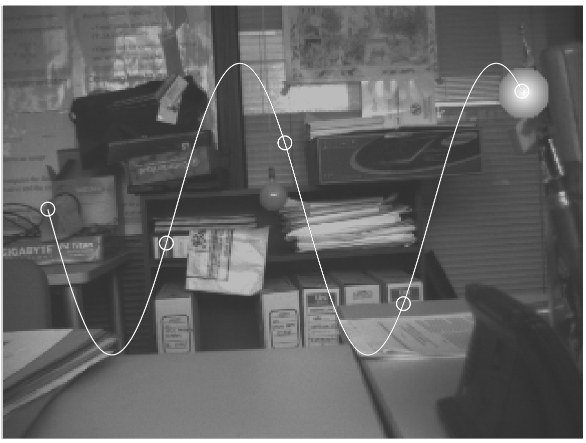

Fig. 1. Ball trajectory and five calculated positions using a standard camera

The SCD camera perfectly reproduces the ball trajectory, while the standard camera is not even close. But the main advantage is in fact other, since a new object position is calculated for every new pixel, so there is a short delay between the real and tracked ball positions. In contrast, this delay is of the order of milliseconds in a standard camera, since all the pixels must be read-out and processed; it takes almost one frame to complete the position calculation using a standard camera.

The exact time delay between tracked and real ball positions using a SCD camera can be calculated measuring the difference between the real and calculated position at every instant divided by the ball speed at that instant. The mean value of this delay has been calculated for the complete trajectory obtaining an average delay of 190 µs between the real and tracked trajectory. This short delay produces an error of roughly one pixel considering the real speed of the ball. The delay of a standard camera comes from the read-out time plus the processing time; a 20 ms delay (half frame) is an optimistic estimate for this time. An error of about 100 pixels is obtained for such a delay; this is one third of the image length and half of the trajectory shown in Fig. 1, thus it is useless for high-speed object tracking.

3.2 Same bandwidth and high-speed realistic shutter

In this case, the problem of having a shutter larger than the pixel read-out time is addressed. Current high-speed cameras can reach up to 10 Kfps or even more. The shutter or integration time of these cameras is in the order of tenths of microseconds. Let's suppose a shutter time of 500 µs while keeping the pixel read-out at 520 ns to maintain the same bandwidth.

The trajectory obtained with this camera exactly matches the real trajectory of the ball like in the first experiment. The only difference observed is the time delay between the calculated and real ball positions. This delay is about 650 µs and produces an error in the position estimation of roughly 2 or 3 pixels.

3.3 Less bandwidth and high-speed realistic shutter

The experiments presented show that a SCD camera yield enough information to accurately track a fast moving object with the same bandwidth of a standard camera, while the standard camera is not able to track such trajectories and has a large delay.

The SCD camera accuracy and delay is enough for fast movement analysis. But, would it be possible to reduce the required bandwidth while still keeping its accuracy? The last experiment (real shutter) has been repeated with a pixel read-out of 2.6 µs, which supposes one fifth of the bandwidth required by the standard camera and the preceding experiments.

The trajectory calculation of this experiment is the same as the others, so there is no accuracy missing while reducing bandwidth up to one fifth (for lower bandwidths, an error in the trajectory estimation appears). The delay between real position and calculated is again the only parameter that is different. The delay in this case is around 1.3 ms and produces and error in the position estimation of roughly 5 pixels.

3.4 Discussion

A standard camera working at 25 fps has shown useless features for high-speed object tracking as shown in Fig. 1; it is not able to estimate the object trajectory and the time for the image capture and computation produces a high position error, since the object travels a long distance during that time.

A camera working using the Selective Change-Driven (SCD) strategy accurately calculates the object trajectory at any time, even using different shutter speeds and with equal or even lower bandwidths than a standard camera. The differences found, among different configurations, come from the delay between the real and calculated object positions, which produce an error of the object position estimation. Fig. 2 shows the position estimation error for the three experiments along the experiment elapsed time. The bottom curve (around 1 pixel error) corresponds to the first experiment where an ideal unrealistic shutter has been used. The middle curve corresponds to a realistic implementation and it is worse, though still useful, since the error position estimation is around 2-4 pixels. The top curve corresponds to the last experiment where the bandwidth has been dramatically reduced to one fifth, and the error appears to still be small for high-speed object tracking. The error variability along time is due to the non-constant object speed (the sinusoidal speed is faster around cero crossing and slower at the valleys and mountains).

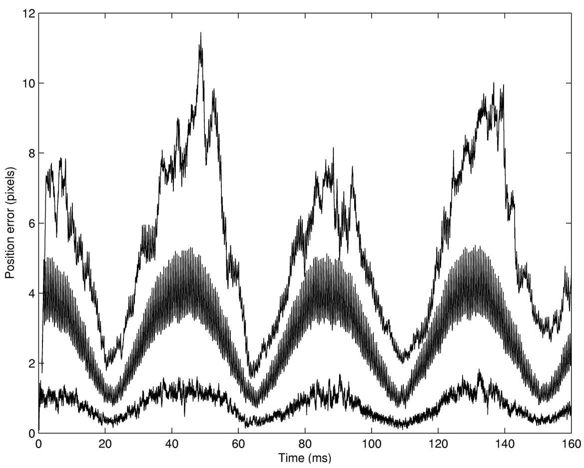

Fig. 2. Position estimation error due to the computation delay for the three experiments

4. CONCLUSION

A new biologically inspired strategy for pixel read-out and processing has been presented. This strategy allows the analysis of ultra-high speed movements while still keeping the same bandwidth requirements of conventional cameras. The simulated experiments have shown that it is possible to accurately measure movements even with less bandwidth than other cameras. The hardware requirements for the implementation of a SCD camera are affordable with present CMOS technology.

ACKNOWLEDGMENT

This work has been supported by the project UV-AE-20060242 of the University of Valencia and the project GV2005-184 of the Generalitat Valenciana.

REFERENCES

[1] F. Pardo, B. Dierickx and D. Scheffer, *Space-Variant Non-Orthogonal Structure CMOS Image Sensor Design*, IEEE Journal of Solid-State Circuits, vol. 33(6), p. 842-849, June 1998.

[2] S.Kameda, T.Yagi, *A Silicon Retina System That Calculates Direction Of Motion,* In Proc. IEEE Int. Symposium on Circuits and Systems, ISCAS'03, vol IV, p. 792-795, Bangkok, Thailand, May 2003.

[3] Eugenio Culurciello, Ralph Etienne-Cummings, *A biomorphic digital image sensor*, IEEE Journal of Solid-State Circuits, vol. 38(2), p. 281-294, February 2003.

[4] T. Teixeira, E. Culurciello, J. Hyuk Park, D. Lymberopoulos, A. BartonSweeney and A. Savvides, *AddressEvent Imagers for Sensor Networks: Evaluation and Modeling*, In Proc. Information Processing in Sensor Networks, IPSN'06, April 2006.

[5] Ralf M. Philipp and Ralph Etienne-Cummings, *Single-Chip Stereo Imager*, Analog Integrated Circuits and Signal Processing, vol. 39, p. 237-250, Kluwer, 2004.

[6] Charles M. Higgins and Christof Koch, *A Modular Multi-Chip Neuromorphic Architecture for Real-Time Visual Motion Processing*, Analog Integrated Circuits and Signal Processing, vol. 24, p. 195-211, Kluwer, 2000.

[7] Erhan Özalevli and Charles M. Higgins, *Reconfigurable Biologically Inspired Visual Motion Systems Using Modular Neuromorphic VLSI Chips*, IEEE Transactions on Circuits and Systems, vol. 52(1), January 2005.

[8] F. Pardo, J. A. Boluda, X. Benavent, J. Domingo, J. C. Sosa, *Circle detection and tracking speed-up based on change-driven image processing*, ICGST International Conference on Graphics, Vision and Image Processing. GVIP'05. pp. 131-136. Cairo, Egypt, December 2005.

[9] Matthias Oster and Shih-Chii Liu, *A Winner-take-all Spiking Network with Spiking Inputs*, In Proc. IEEE International Conference on Electronics, Circuits and Systems ICECS'2004, Tel Aviv, Israel, December 2004.

[10] P. Lichtsteiner, T. Delbruck and J. Kramer, S.C. *Improved On/Off Temporaly Differentiating Address-Event Imager,* In Proc. IEEE Int. Conference on Electronics, Circuits and Systems ICECS'2004, Tel Aviv, Israel, December 2004.

[11] Jörg Kramer, *An On/Off Transient Imager With Event-Driven, Asynchronous Read-Out*, In Proc. IEEE International Symposium on Circuits and Systems, ISCAS'02, vol II, p. 165-168, Phoenix, AZ, May 2002.

[12] Giacomo Indiveri, Philip Oswald and Jörg Kramer, *An Adaptive Visual Tracking Sensor With A Hysteretic Winner-Take-All Network*, In Proc. IEEE International Symposium on Circuits and Systems, ISCAS'02, vol II, p. 324-327, Phoenix, AZ, May 2002.

[13] K. Aizawa, Y. Egi, T. Hamamoto, M. Hatori, M. Abe, H. Maruyama, and H. Otake, *Computational image sensor for on sensor compression*. IEEE Transactions on Electron Devices, 44(10):1724-1730, October 1997.

[14] T. Hamamoto, R. Ooi, Y. Ohtsuka, and K. Aizawa, *Real-time image processing by using image compression sensor*. In International Conference on Image Processing, ICIP'99, vol 3, p. 935-939, October 1999.