Embedded Contours Extraction for High-Speed Scene Dynamics Based on a Neuromorphic Temporal Contrast Vision Sensor

A. N. Belbachir, *Member*, *IEEE*, M. Hofstätter, N. Milosevic and P. Schön
smart systems Division
Austrian Research Centers GmbH - ARC
Tech Gate Vienna, Donau-City-Straße 1, 1220 Vienna, Austria
{ahmed.belbachir, michael.hofstaetter, nenad.milosevic, peter.schoen}@arcs.ac.at

Abstract

The paper presents a compact vision system for efficient contours extraction in high-speed applications. By exploiting the ultra high temporal resolution and the sparse representation of the sensor's data in reacting to scene dynamics, the system fosters efficient embedded computer vision for ultra high-speed applications. The results reported in this paper show the sensor output quality for a wide range of object velocity (5—40 m/s), and demonstrate the object data volume independence from the velocity as well as the steadiness of the object quality. The influence of object velocity on high-performance embedded computer vision is also discussed.

1. Introduction

The **A**ddress-**E**vents **R**epresentation (AER) [5] is the data space applied to the neuromorphic temporal contrast vision sensors referenced in [8][14]. These sensors have the property to only transmit the illumination changes in the visual scene such that activities are well captured in form of object edges. In contrast to traditional images collected by the frame-based sensors, the **A**ddress-**E**vents (AE) are asynchronously generated and consequently, they are not equidistantly sampled neither in space nor in time. AE correspond to scene dynamics and thus, they allow sparse representation of scene activities with a drastically reduced data volume compared to frame-based sensors.

This paper is concerned with the exploitation of the neuromorphic temporal contrast vision sensors for embedded computer vision in two aspects. The first aspect relates to the reduced data volume from scene activities monitoring due to the efficient representation of activities using on-chip pre-processing. Indeed, almost all computer vision tasks rely on efficient discontinuities (edges) representation as they consist of relevant scene information. Huge effort has to be invested to accurately extract this information such that important and costly processing resources have to be reserved for this task. Therefore, embedded computing has to face with this limitation for clock-based sensors, while the neuromorphic temporal contrast vision sensors support the discontinuities (edges) detection by means of the on-chip pre-processing.

The second benefit, from using the neuromorphic temporal contrast vision sensors, stands for the high temporal resolution. The temporal accuracy of the scene activities follows the object velocity and does not depend on a frame rate as in the case of clock-based sensors. AE are autonomously generated by the neuromorphic pixels upon illumination change, with ultra-high temporal precision, that is compatible with the change of the velocity. Furthermore, this paper shows the object steadiness in edge representation and consequent volume of AE along a wide range of object velocities, which make the sensor attractive for embedded computer vision. A neuromorphic-based sensor system equipped with a dual-line temporal contrast vision sensor [14] and the Blackfin BF-537 processor from Analog Device is used to evaluate the benefit of the neuromorphic technology for embedded computer vision.

The paper is structured as follows. In Section 2, a brief review of neuromorphic sensors and their characteristics is provided. Section 3 presents AER representation of the dual-line temporal contrast vision sensor for an object with velocities ranging from 5 – 40 m/s. An analysis of the advantages of the embedded computer vision using the dual-line sensor is given in Section 4. Section 5 discusses the advantages and challenges of an embedded computer vision system based on the dual-line temporal contrast sensor and potential improvements. A summary is given in Section 6.

2. Review of the Neuromorphic Sensors

To make this paper self-contained, a short overview of the neuromorphic technology is provided. Afterwards, the dual-line sensor characteristics, relevant to understand the experimental results, are summarized.

2.1. Neuromorphic Sensors

In the late 1980's Carver Mead [12] introduced the neuromorphic concept to describe VLSI systems containing analogue and asynchronous digital electronic circuits that mimic neural architectures present in biological nervous systems. This concept revolutionized the frontier of computing and neurobiology to such an extent that a new engineering discipline emerged, with the aim to design and build artificial neural systems, such as vision systems, auditory processors or autonomous, roving robots. The field is referred to as "neuromorphic engineering". The term neuromorphic has also been coined by Carver Mead in an attempt to name artificial systems that adopt the form of, or morph, neural systems.

In a groundbreaking invited paper about neuromorphic electronic systems [13], published 1990, Mead argues that the advantages of biological information-processing can be principally attributed to the use of elementary physical phenomena as computational primitives, and to the information representation by the relative values of analog signals, rather than by the absolute values of digital signals. He further argues that this approach requires adaptive techniques to correct for differences of nominally identical components, and that this adaptive capability naturally leads to systems that learn about their environment.

Vision models have been built in sensors like the one of Mahowald and Mead [10] [11], originally named the "silicon-retina" sensor. In succession, a large variety of diverse silicon-retina sensor designs have been carried out and reported, including gradient based sensors sensitive to static edges [6], temporal contrast vision sensors that are sensitive to relative light intensity changes [8], orientation selective spiking neurons devices [9] from Tobi Delbrück and its group in ETH Zürich and optical flow sensors [4] from Bernabe Linares-Barranco.

Silicon-retina sensors feature massively parallel pre-processing of the visual information in on-chip analogue circuits and are commonly characterized by high temporal resolution, wide dynamic range and low power consumption. Typically these sensors exploit a very efficient asynchronous, event-based encoding of the visual information that drastically reduces the data redundancy and optimizes the use of the transmission channel bandwidth which is called "Address-Event-Representation" (AER). The address-event communication channel itself is a model of the transmission of neural information in biological systems.

2.2. Dual-Line Sensor Characteristics

The dual-line sensor system comprises two boards, a sensor board and a DSP board. A sensor chip is located on the sensor board and a data processing unit, based on a Blackfin BF537 from Analog Device, is the core part of the DSP board. The processing unit contains embedded software for real-time analysis and interpretation of the sensor data.

The vision sensor chip [14] contains a dual-line arrangement of 2×256 temporal contrast pixels. The on-chip distance between the two lines is 250µm. This distance and the event generation time are used for the velocity estimation of objects crossing both lines. The chip includes a built-in synchronous arbiter, which is responsible for structuring the transfer of the events generated from the pixels to the data buffer. An on-chip electronic [7] assigns timestamps to the events during their generation on-chip at the pixel level with a minimal time resolution of 100ns. The output data stream of the sensor consists of 16-bit words of address events (AE), which contains the location (address) of the event generating pixel within the dual-lines, and the timestamps (TS) to represent the event occurrence time. The timestamp value is transmitted once for corresponding address events occurring within the timestamp interval (down to 100 ns). TS data and AE are always transferred together such that the timestamps are only annexed when AE are generated. If no AE are generated, no AE data are put out. The timestamp counter wrap-around is signaled in any case. This enables the processor to perform a timestamp value expansion to keep the absolute time differences. Figure 1 depicts an output data sequence example.

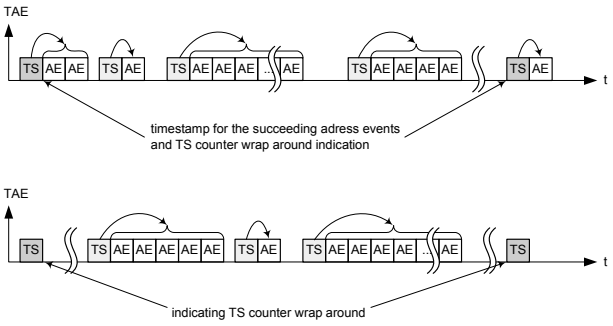


Figure 1: Sequence protocol of TAE data types

AE are subdivided in ON-event to illustrate the intensity increase and OFF- event to show the intensity decrease. To distinguish between TS, ON-event, OFF-event, a dedicated bit coding is used [7]. The sensor board also includes auxiliary electronics for the on-the-fly configuration of the sensor chip. This allows adapting the sensor to different scene conditions like varying illumination, object reflectance and speed.

The timestamps (event occurrence time) and the corresponding AE are transmitted to a FIFO on a 16-bit parallel bus. The FIFO is placed between the sensor and the DSP to cope with peaks of AE activity and is capable of handling up to 40 MHz memory access frequency.

An embedded computing unit based on the Blackfin BF-537 allows real-time interpretation and embedded computer vision. It is annexed to the FIFO via the parallel port. Hereby, new data are signaled to the computing unit by using an external interrupt request input. The interrupt service routine handles the date transfer from the parallel port to the data buffer which is located in the external memory. This data buffer is the base for the algorithmic processing. The computing unit has a maximum frequency of 600 MHz, 128 KB internal memory and 32 MB external SDRAM memory. This memory resource limits the processing capabilities of any high-resolution video system and may not feed the video processing needs.

3. Address-Events Representation with a Dual-Line Sensor

This Section presents the AE raw data resulting from an object moving at selected velocities from 5 to 40 m/s. Next, the resulting object's data volume has been evaluated for the different velocities. Finally, a comparison of the data rate between this neuromorphic sensor and clock-based line sensor is provided.

3.1. Address-Event Representation

The 2-D object representation with the dual-line sensor consists of the pixel index (y-axis) versus the time (x-axis). The x-axis represents the event generation time in units of the timestamp period, the y-axis is the pixel address (0-255). For this test, a 2-D object has been fixed on a rotating drum with tunable velocity from 1 to 40 m/s, and the corresponding AE data have been generated at selected velocities (5 – 40 m/s).

Figure 3 depicts an original object (a) and its corresponding AE on one sensor line for the object crossing the sensor field of view at several velocities ((b): 5 m/s, (c): 15 m/s, (d): 20 m/s, (e): 30 m/s and (f) 40 m/s) at a distance of about 18 cm. The magnification factor of the used lens has been estimated to about 5.7. The black dots represent the OFF-events while the white dots show the ON-events. The background (grey) does not correspond to any data. The timestamp period was configured to 5µs for all velocities.

The time has been scaled to retrieve the original 2-D object shape. The scale factor is the object velocity v as the distance x is calculated as follows:

$$x = v \cdot t \qquad (1)$$

where t is the event generation time. By using the prior knowledge about the distance between the sensor lines (250µm) and by correlating the data from both lines, the object velocity can be estimated to build the 2-D object form. The velocity estimation using the dual-line sensor has been performed with 1% accuracy as in [3].

Figure 3 (b)-(f) shows the object quality steadiness at different velocities. Figure 3(b) depicts the double edging effect mainly related to the common configuration of the refractory period, which is a tunable parameter to control the maximal event generation rate on pixel level, for all tests and velocities

3.2. Resulting Object's Data Volume at Different Velocities

The corresponding data volume per object has been evaluated and the results are shown in Figure 2. The x-axis shows the object velocity and the y-axis depicts the respectively generated number of AE per object. The same sensor configuration has been used for all tests. Objects with slow velocities show several spikes (AE) for coding one edge, due to the unified configured refractory period for all velocities. Figure 3(a) shows the effect of double

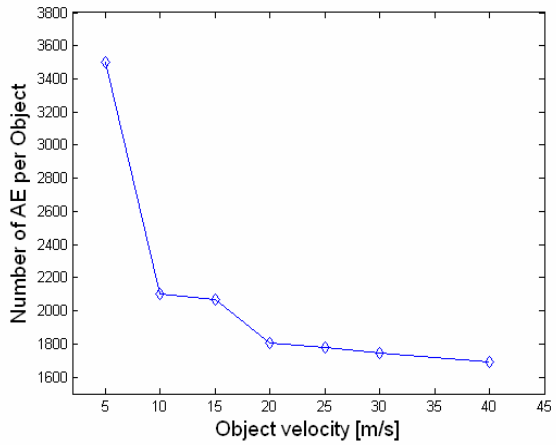


Figure 2: Number of Address-Events generated by the dual-line sensor for one object taken at different velocities

spiking at object speed of 5 m/s. Therefore, Figure 2 shows higher data volume for the same object at low velocities than at high velocities.

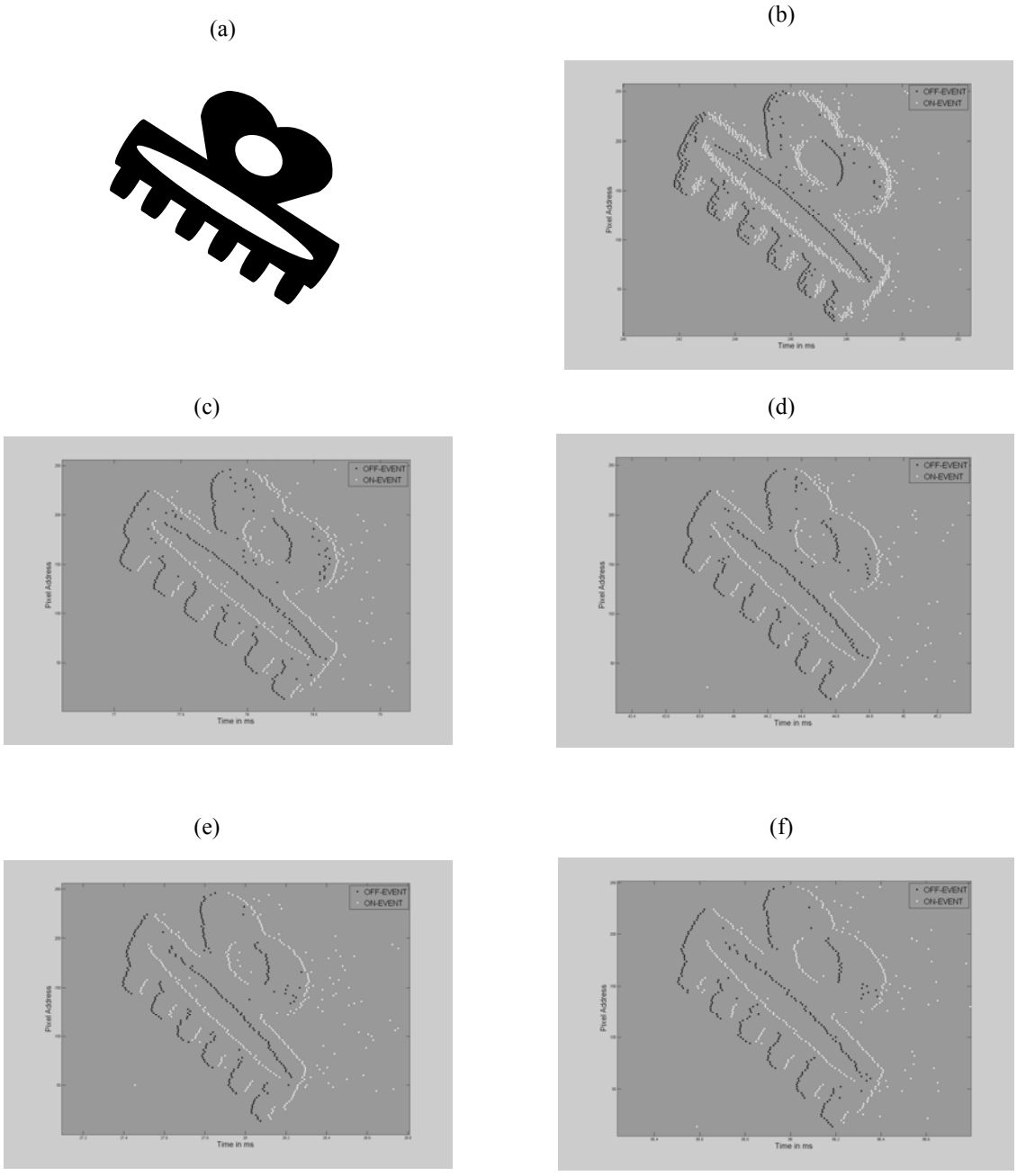


Figure 3: Original object (a) and its representation in AER space at velocities 5m/s (b), 15m/s (c), 20 m/s (d) 30m/s (e) and 40 m/s (f) using one sensor line of the neuromorphic temporal contrast dual-line vision sensor

3.3. Data Rates

Table 1 shows an example of the resulting data rates for objects crossing the sensor field of view at different speeds. For object length of 2.5 cm and distance between objects of about 7.5 cm, the dual-line temporal contrast vision sensor can scan 10 objects per meter. The corresponding number of objects/s for different velocities with the corresponding data rate is reported in Table1. One sensor line produces 2.6 Mbit/s for 50 objects per second and 10.3 Mbit/s for 400 objects per second.

Speed[m/s]	Number of objects/s	Data rate [Mbit/s]
5	50	2.6
10	100	3.2
15	150	4.7
20	200	5.5
25	250	6.7
30	300	7.9
40	400	10.3

Table 1: Representative data rates from one sensor line of the neuromorphic dual-line sensor for object length of 2.5 cm moving at different velocities and at with a repetition distance of 7.5 cm

The comparison with standard clock-based line sensor is not trivial because of the difference in the characteristics. The 5µs timestamp are comparable to a 200 kHz line rate however the AER with OFF and ON events cannot be compared to 1-bit resolution because the temporal contrast vision sensor is sensitive to illumination changes and thus supports a high range of grey values, without taking the absolute grey value into account. Indeed, AER consists of relative values for illumination change. For a fair comparison, an existing ultra high-speed clock-based line scanner in the market, running at 100 kHz line rate and having 8-bit pixel resolution, has been chosen. The data rates resulting from the comparison are presented in Table 1. The data rate D_c for the clock-based line scanner is calculated as follows:

$$D_c = 100000 \cdot 8 \cdot 256 = 195\ Mbit/s \qquad (2)$$

This comparison is illustrated in Figure 4. The x-axis represents the object velocity while the y-axis shows the resulting data rate in a logarithmic scale for the clock-based line scanner (dashed line) and one sensor line of the dual-line sensor (solid line with diamonds). It can be noticed the drastic reduction of the data volume and the efficient exploitation of the bandwidth using the neuromorphic dual-line sensor. The data volume almost increases linearly with the number of objects/s.

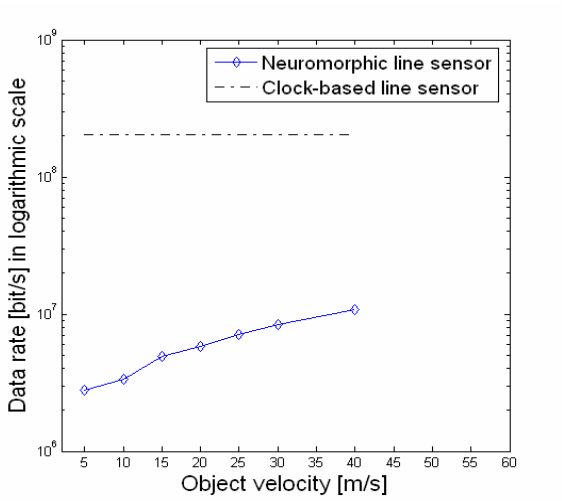


Figure 4: Comparison between a 256-pixel line neuromorphic sensor and a clock-based line sensor in terms of data rate vs. object velocity (b).

4. Embedded Object's Contours Extraction

This Section presents the initial steps for embedded computer vision, which include the data acquisition, object detection and preparation for interpretation. The computational effort for this step and the available results for further processing are assessed for the Blackfin BF-537 from Analog Device.

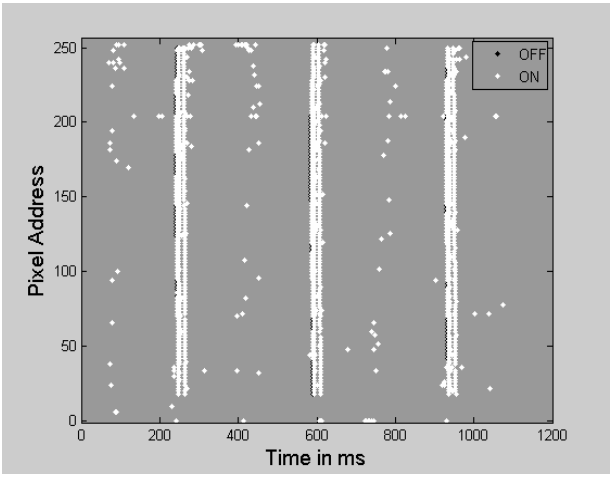
4.1. Object Detection

Clock-based sensors require additional systems to detect objects within the data stream. These systems can be mechanical (external trigger) or algorithmic, such that a software module has to be included for extracting the objects out of the data stream. The neuromorphic temporal contrast vision sensor has the advantage to implicitly detect the object as the data rate will drastically increase whenever this object crosses the sensor's field of view. The object motion causes illumination changes that implicitly trigger the generation of a bulk of AE. Therefore, the individual objects are detected in the data stream by monitoring the data rate in time slots under the assumption that the objects are sequentially crossing the sensor field of view. An object is detected when the event rate exceeds a first threshold (threshold 1 is set typically to 5 events per time slot).

When the event rate falls below second threshold (threshold 2 < threshold 1), the detection algorithm assumes the end of the object. The event rate is calculated for time slots of 1 ms. The identified cluster of events is collected and associated to one object. Figure 5 depicts 3

objects crossing the neuromorphic sensor field of view at 5 m/s (a) and their corresponding averaged events rate (b).

(a)



(b)

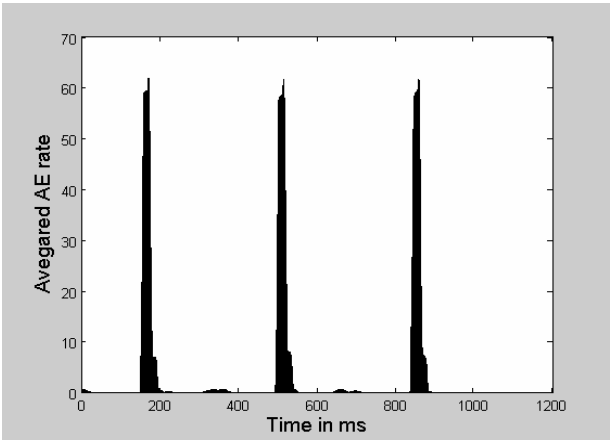


Figure 5: Detection of three objects (a) crossing sensor field of view at a velocity of 5 m/s using averaged AE rate

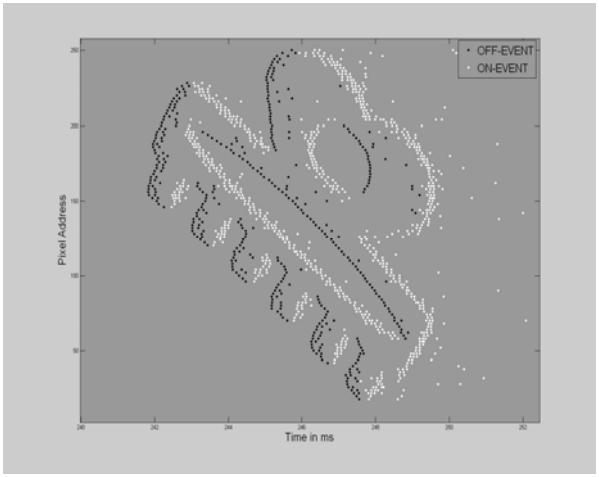
The averaging has been performed to reduce the effect of the salt-and-pepper noise (white dots) originating from the internal sensor noise as well as from fluctuations in the scene. The three clusters of events allow the detection of the objects and their localization in time (relative occurrence time in the x-axis).

4.2. Contours Extraction

For the extraction of the object contours, two processing steps are carried out: the edge thinning and outliers removal. The edge thinning is performed by emulating an additional event type (ON or OFF) dependent refractory period in the processing while outliers removal aims to suppress the isolated AE (salt and pepper noise).

The refractory period is the sensor parameter that affects the AE volume and depends on the object velocity. This parameter defines the duration in which the pixels are not allowed to generate a second event. It is beneficial in avoiding several spiking (AE) on the same intensity change information and thus avoiding double edges. For the same refractory period, several edges can be produced for slow objects while fast moving objects may generate individual edges due to the spike-rate coding of contrast edges.

(a)



(b)

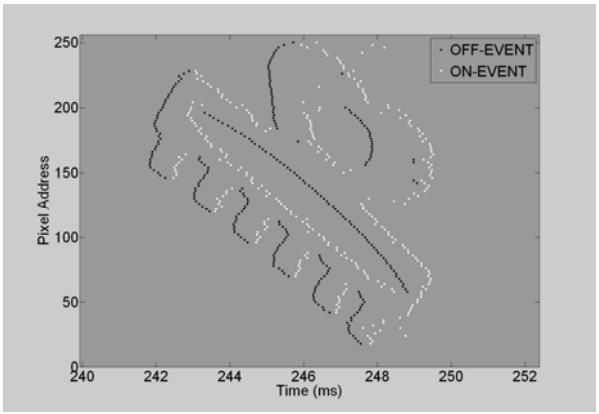
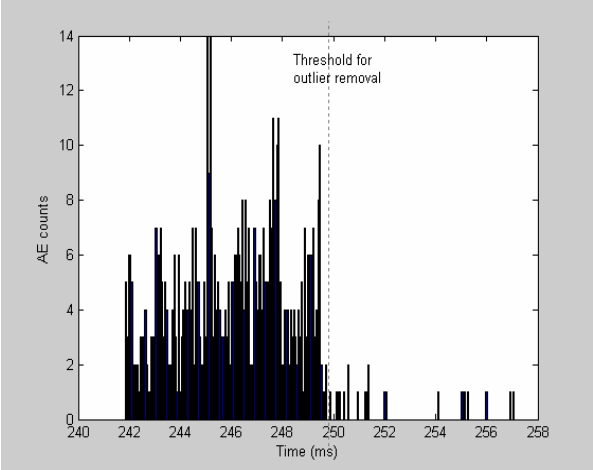


Figure 6: Edge thinning results

To keep the system configuration independent from the object velocity, a short refractory period has been used while an edge thinning module is included in the embedded processing to unify the object contours representation independently from its speed. This module is intended for extracting the object edges by removing

redundant events. Each pixel can deliver one to several events per time slot and contrast change.

(a)



(b)

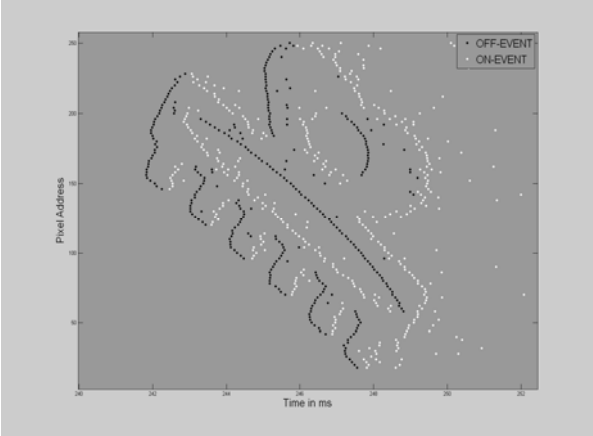


Figure 7: Outliers removal results

This algorithm reduces the data to one event per pixel and slot duration, thus only the first pixel event is kept per slot. The time slot is adjusted for compatible object contours in the velocity range of 1 – 40 m/s.

Figure 6(b) shows the edge thinning results for the object moving at 5 m/s (Figure 6(a)). It can be noticed that the first event per pixel has been kept while the next redundant spikes within a time slot have been removed. The resulting object is comparable to those in Figure 3 ((c) – (f))

The second step consists of removing the isolated AE, corresponding to salt and pepper noise. By building a histogram of the occurrence times of AE, many outliers can be detected. Figure 7(a) depicts this histogram for AE shown in Figure 6(b). We can see a bulk of AE between 242 and 250 ms and few isolated AE for time >250ms. These Individual AE are detected and removed by setting an appropriate threshold (if the number per timestamp = 1 then the corresponding AE will be removed). The result of the outlier removal is shown in Figure 7(b). We can notice the drastically reduced number of outliers compared to Figure 6(a) while preserving the main object contours.

4.3. Embedded Computing Performance

In order to evaluate the embedded computer vision and scene interpretation performance using the neuromorphic dual-line temporal contrast vision sensor, a quantitative estimation of the used resources in term of memory and processing effort for data acquisition, object detection and contours extraction. Table 2 provides this quantitative evaluation for the Blackfin BF-537 processor from Analog devices with 32 MB RAM and 600 MHz.

Speed [m/s]	Number of objects/s	Memory Usage /s [% from 32 Mbytes]	CPU Workload [% from 600 MHz]
5	50	1	5
10	100	1.3	8
15	150	1.9	11
20	200	2.2	16
25	250	2.7	27
30	300	3.1	34
40	400	4.1	58

Table 2. Quantitative evaluation of the relative resources usage using the dual-line sensor and the Blackfin BF-537 from Analog Device with 600 MHz and 32 MB memory

The memory usage is less than 4.1% of the total memory for up to 400 object/s (with velocities up to 40 m/s) such that the object has 25 mm length. A Clock-based line sensor with 256 pixels with 8 bit output and running at 100 kHz line rate would provide a permanent data volume of 24.4 MByte/s, that is 76% of the Blackfin overall memory. This memory usage is permanent and independent from the object speed and the number of objects per second. Therefore, the neuromorphic allows a reduction of the used memory space for embedded processing and storage by more than 20 times.

Furthermore, processing workload has been estimated to 58 % for 400 objects crossing the sensor field of view at a velocity of 40 m/s. This workload includes the data acquisition, the object detection and the contours extraction. The clock-based line sensor may saturate the processing at velocity ≥ 10 m/s because the object detection is not implicitly fostered by the sensor and

therefore, the corresponding algorithms have to be implemented in the embedded computing unit, which become computationally prohibitive at higher object velocities.

5. Discussions

The neuromorphic temporal contrast vision sensors are very attractive for efficient embedded computer vision especially for high-speed applications. They have the advantage to provide drastic reduction of the data volume such that further data interpretation can be achieved on low-cost embedded computing unit. Furthermore, the sensor provides high temporal resolution as the time information does not depend on a frame rate but on the scene dynamics and the object velocity. Moreover, the sensor is not sensitive to illumination conditions as it reacts to relative illumination changes. As a consequence object detection is implicitly supported by the sensor.

The typical sensor output data are the object edges. However, further embedded processing can be performed for contours extraction as the data may be contaminated by sensor internal noise (salt-and-pepper noise) or duplicated edges.

Another side effect of the sensor consists of non-sensitivity to object edges, which are parallel to the motion direction. These edges may not generate AE as they will not involve illumination change. An additional processing step can be included for the reconstruction of these edges.

6. Summary and Conclusions

In this paper, an embedded computer vision system for contours extraction of high-speed dynamics based on a neuromorphic temporal contrast vision dual-line sensor has been presented. By exploiting the sparse representation of the scene dynamics and by offering an ultra-high temporal resolution (at least 100ns), this system is very beneficial for ultra-high speed applications where traditional clock-based sensor systems are computationally prohibitive. The system has been demonstrated for contours extraction of objects moving at velocities up to 40 m/s. The quantitative evaluation has shown a reduction of the data rate by a factor of 20 for 400 objects per second moving at 40 m/s.

As perspective, advanced embedded computer vision dealing with object recognition will be investigated using approaches from the computational intelligence like the neural networks and support vector machines.

References

[1] D.H. Ballard and C.M. Brown, “Computer Vision,” Prentice-Hall Inc. Englewood Cliffs, New Jersey, 1982

[2] A.N. Belbachir, M. Litzenberger, C. Posch and P. Schön, “Real-Time Vision Using a Smart Sensor System,” in the International Symposium on Industrial Electronics, ISIE2007, Spain, June 2007.

[3] A.N. Belbachir, M. Hofstätter, K. M. Reisinger, N. Donath and P. Schön, “Object Velocity Estimation Based on Asynchronous Data from a Dual-Line Sensor System,” in the International Conference on Industrial Informatics, INDIN2007, Vienna, Austria, July 2007

[4] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona and B. Linares-Barranco, “A Spatial Contrast Retina with On-chip Calibration for Neuromorphic Spike-Based AER Vision Systems,” IEEE Trans. Circuits and Systems, Part-I: Regular Papers, vol. 54, No. 7, pp. 1444-1458, July 2007.

[5] E. Culurciello, R. Etienne-Cummings, K. Boahen, “Arbitrated address event representation digital image sensor,” ISSCC Dig. Tech. Papers, pp. 92–93, Feb. 2001.

[6] T. Delbrück, “Silicon Retina with Correlation-Based, Velocity-Tuned Pixels,” IEEE Transactions on Neural Networks, Vol. 4, No. 3, pp. 529-541, 1993.

[7] M. Hofstätter, A.N. Belbachir, E. Bodenstorfer and P. Schön, “Multiple Input Digital Arbiter with Timestamp Assignment for Asynchronous Sensor Arrays,” *IEEE ICECS06*, Nice, France, 2006

[8] P. Lichtsteiner, C. Posch and T. Delbrück, “A 128×128 120dB 15us Latency Asynchronous Temporal Contrast Vision Sensor,” IEEE Journal of Solid State Circuits, Vol. 43, pp. 566 - 576, Feb. 2008.

[9] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, T. Burg, and R. Douglas, “Orientation-selective aVLSI spiking neurons,” Neural Networks, Special Issue on Spiking Neurons in Neuroscience and Technology, 14(6/7), pgs 629-643, 2001.

[10] M. A. Mahowald and C. A. Mead, “A Silicon Model of Early Visual Processing,” Neural Networks, vol. 1, pp. 91-97, 1988.

[11] M. A. Mahowald and C. A. Mead, “Silicon Retina,” in C. A. Mead, Analog VLSI and Neural Systems. Reading, pp. 257-278, MA: Addison-Wesley, 1989.

[12] C.Mead, “Analog VLSI and Neural Systems,” Addison-Wesley, 1989.

[13] C. Mead, “Neuromorphic electronic systems,” Proceedings of the IEEE Volume 78, Issue 10 pp. 1629-1636, Oct 1990.

[14] C. Posch, M. Hofstätter, D. Matolin, G. Vanstraelen, P. Schön, N. Donath and M. Litzenberger, “A Dual-Line Optical Transient Sensor with On-chip Precision Timestamp Generation,” IEEE International Conference ISSCC, Digest of Technical Papers , pp., Feb. 11-15, 2007