

Live Demonstration: On the distance estimation of moving targets with a Stereo-Vision AER system

M. Domínguez-Morales, A. Jiménez-Fernández, R. Paz-Vicente, G. Jiménez, A. Linares-Barranco

Robotic and Technology of Computers Lab. University of Seville.
Av. Reina Mercedes s/n, 41012-Sevilla, SPAIN.mdominguez@atc.us.es

Abstract—Distance calculation is always one of the most important goals in a digital stereoscopic vision system. In an AER system this goal is very important too, but it cannot be calculated as accurately as we would like. This demonstration shows a first approximation in this field, using a disparity algorithm between both retinas. The system can make a distance approach about a moving object, more specifically, a qualitative estimation. Taking into account the stereo vision system features, the previous retina positioning and the very important Hold&Fire building block, we are able to make a correlation between the spike rate of the disparity and the distance.

I. OVERVIEW

The demonstration consists of the system shown in the photo and a moving stimulus. The stimuli will be always in movement, getting closer to and further from the retinas. Our system will be able to indicate the estimation distance with a chain of led lights, which will indicate the distance from the point of view of our system. This system has been under test in our lab for the last few weeks and it is being improved by using other correlation mechanisms to obtain better results.

II. HARDWARE INVOLVED

The components shown in this demonstration are: two DVS128 retinas (with a resolution of 128 by 128 pixels each), two USB-AER (CAVIAR project), a Virtex-5 PFGA (VHDL algorithms) and the UBSminiAER2 (monitor – connection with the laptop). USB-AER boards have a hardware background activity filter to eliminate retinas' spurious; in the future, this filters will be integrated in the FPGA. Both retinas are calibrated with an angle of $86^{\circ}1387$ degrees to obtain a focal collision of one meter. The FPGA processes both streams of spikes, applies the Multi Hold&Fire algorithm, calculates the correlation between the spike rate before and after the Multi Hold&Fire to obtain a better distance approximation, and switches on the led after the distance estimation. The output stream of spikes of the FPGA board is monitored by the USBminiAER2, which collects the spikes into USB packages and send them to a laptop. After that, jaER software will process these packets and show the results on the screen.

The demonstration will use, as the moving object, a Scalextric circuit with a car moving at a constant velocity. The circuit arrangement will allow the car getting closer to and further from the retinas during its travel.

III. VISITORS EXPERIENCE

Visitors do not interact with the system because this fact can modify the previous retina positioning and make the system to work incorrectly. Led lights are put in such a way that they are positioned parallel to the moving object, and these led lights will be switched on to indicate where our system thinks the object is.

The visitors will be able to see the system working with a scalextric car running through the circuit and the led lights illuminating the course of the car to indicate where the system thinks the car is. Also, the visitors can observe the system disparity output shown in a laptop and using the jaER software tool.

In conclusion, using a moving target makes this an entertaining and showy demonstration, which seems to be simple. However, this work has very difficult and important aspects that the visitor would like to discuss.



Whole system with two DVS128 retinas (left), two USB-AER boards (middle-left), a Virtex-5 PFGA (middle-right) and the USBminiAER2 (right)

On the distance estimation of moving targets with a Stereo-Vision AER system

M. Domínguez-Morales, A. Jiménez-Fernández, R. Paz-Vicente, G. Jiménez, A. Linares-Barranco

Robotic and Technology of Computers Lab. University of Seville.
Av. Reina Mercedes s/n, 41012-Sevilla, SPAIN.mdominguez@atc.us.es

Abstract—Image processing in digital computer systems usually considers the visual information as a sequence of frames. Digital video processing has to process each frame in order to obtain a result or detect a feature. In stereo vision, existing algorithms used for distance estimation use frames from two digital cameras and process them pixel by pixel to obtain similarities and differences from both frames; after that, an estimation is calculated about the distance of the different objects of the scene. Spike-based processing implements the processing by manipulating spikes one by one at the time they are transmitted, like the human brain. The mammal nervous system is able to solve much more complex problems, such as visual recognition by manipulating neuron's spikes. The spike-based philosophy for visual information processing based on the neuro-inspired Address-Event-Representation (AER) is achieving nowadays very high performances. This study proposes a two-DVS-retina connected to a Virtex5 FPGA framework, which allows us to obtain a distance approach of the moving objects in a close environment. A Multi Hold&Fire algorithm in VHDL is also proposed, which obtains the differences between the two retina output streams of spikes; and a VHDL distance estimator.

I. INTRODUCTION

In recent years there have been numerous advances in the field of vision and image processing, because they can be applied for scientific and commercial purposes to numerous fields such as medicine, industry or entertainment. Trying to simulate the vision of human beings, researchers have experimented with two-camera-based systems inspired by human vision ([8][9]). Following this, a new research line has been developed, focused on stereoscopic vision [1]. In this branch, researchers try to obtain three-dimensional scenes using two digital cameras. Thus, we try to get some information that could not be obtained with a single camera, i.e. the distance at which the objects are.

By using digital cameras, researchers have made a breakthrough in this field, going up to create systems able to achieve the above. However, digital systems have some problems that, even today, have not been solved (like real time stereo vision matching, because of the features extraction phase [10]). A logical and important result in stereoscopic vision is the calculation of distances between the point of view and the object that we are focused in. This problem is still completely open to research. The problems related to this are the computational cost needed to obtain appropriate results

and the errors obtained after distance calculation. There are lots of high-level algorithms used in digital stereo vision that solve the distance calculation problem, but this implies a computer intervention into the process and it is computationally expensive. This makes it difficult to develop a real-time autonomous system.

Bio-Inspired and Neuro-Inspired systems or circuits are approaches to solve real problems by mimicking Biology in its efficient solutions [2]. Spiking systems is one of the neuroinspired alternatives of mimicking the neurons layers of the brain for processing purposes. These systems process the information into a continuous way, without discretization into frames. Hardware implementations of these systems are usually composed of several steps: sensors [6], filters [13], convolutions [3], actuators [14], etc... Engineers found a great problem at this point because they need to communicate thousands of neurons from one chip to the next chip, but they have a limitation in the number of pins. Address-Event-Representation solves this problem [4]. Usually these AER circuits are built using self-timed asynchronous logic. The success of these systems will strongly depend on the availability of robust and efficient development, debugging and interfacing AER-tools [15].

In this work, a new approach about distance estimation is presented, using an AER stereo-vision system. Also, some distance estimation results are presented, and they are compared with the real distance.

II. COMPONENTS.

In this section we describe the components used. We can see from figure 1 a block diagram of the whole system.

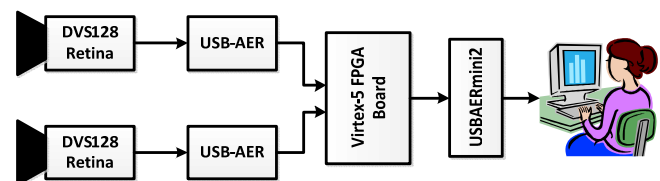


Figure 1. System components and interconnections between them.

All the elements that compose our system are these (from left to right): two DVS128 retinas [6], two USB-AER, a Virtex-5 FPGA board, an USB AER mini2 [5] and a computer

monitoring the resulting stream of spikes with jAER software [7]. Next, we will talk about the USB-AER and the Virtex-5 FPGA board.

The USB-AER board was developed in our lab during the CAVIAR project, and it is based on a Spartan II FPGA with two megabytes of external RAM and a Cygnal 8051 microcontroller. To communicate with the external world, it has two parallel AER ports (IDE connector). One of them is used as input, and the other is the output. In our system we have used two USB-AER boards, one for each retina. In these two boards we have synthesized in VHDL a filter called Background-Activity-Filter, which allows us to eliminate noise from the stream of spikes produced for each retina. This noise (or spurious) is due to the analog pixel nature of the retina. So, at the output of the USB-AER we have the information filtered and ready to be processed.

The other board used is a Xilinx Virtex-5 board, developed by AVNET [12]. This board is based on a Virtex-5 FPGA and mainly has a big port composed of more than eighty GPIOs (General Purpose Inputs/Outputs ports). Using this port, we have connected an expansion/testing board, which has standard pins, and we have used them to connect two AER inputs and one output. The Virtex-5 implements the whole processing program, which works with the spikes coming from each retina, processes them and obtains the differences between both retinas and the spikes rate of this difference. The whole program block diagram is shown in figure 3. The system behavior and its functionality are shown in the next sections.



Figure 2. Whole real system.

III. MULTI HOLD&FIRE ALGORITHM.

Traffic coming from both retinas must be subtracted on the fly in order to calculate the differences between them. To do that, we have used the idea of the Hold&Fire building block [11] to obtain the difference of two signals. This algorithm subtracts two pulses, received from two different ports. When it receives an event, it waits for a short fixed time to another one with the same address. If it does not receive a second event and the fixed time is over, it fires the pulse. Otherwise, if it receives another event, depending on its polarity and its source, the algorithm fires an event or not. Using this on retinas, if the second pulse received has the same address we have two options: if the new event comes from the other retina, the event is cancelled and no event is transmitted; but if this second event comes from the same retina, the first event is

dispatched and this second event take the role of the first event and the system waits again the short fixed time. This Hold&Fire operation for subtracting or cancelling two streams of spikes is described in depth in [11].

We have extrapolated the Hold&Fire block to a set of 128x128 signals system (one for each pixel of the retinas). We call this a Multi Hold&Fire system, which allows us to calculate the differences between two retina output spikes streams.

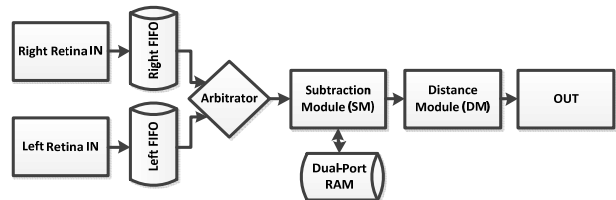


Figure 3. VHDL Blocks.

Our algorithm is based on multiple Hold&Fire blocks, as commented before (it has one H&F block for each two equivalent pixels of the two retinas). It treats each pixel separately and obtains the difference between this pixel in the left retina and the same pixel in the right retina. At the end, we have a stream of spikes that represents the difference of both retinas in our system output.

The VHDL project has these blocks:

- Two FIFOs: storage the great amount of spikes from both retinas.
- An Arbitrator: selects spikes from both FIFOs depending on their occupation.
- Subtraction Module: applies the Multi Hold&Fire algorithm to the stream of spikes received.
- Distance Module: estimates the distance. We will go deeper into it in the next section.
-

IV. DISTANCE ESTIMATION.

A. Description

As it was explained in the previous section, this system uses the distance module to obtain an approach of the distance at which the moving object is. The algorithm described is a logical approach for calculating distances. We will not obtain a quantitative result of the distance. This fact is not a failure, because human vision cannot calculate distances, it can only estimate distances, based on previous experiences. So, our algorithm is very close to the behavior of human vision.

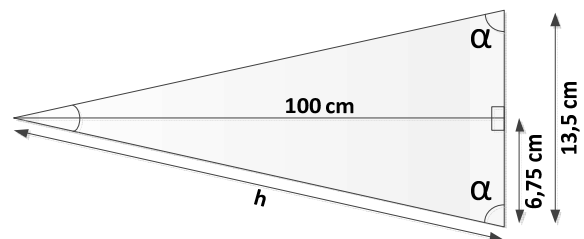


Figure4. Retinas positioning.

In our system, both retinas are positioned with a certain angle to obtain a focus distance of 1 m. To do that, we have put our retinas in a base, separated at 13'5 cm from each other. We have obtained the system shown below (Figure4).

Applying Pythagoras and trigonometric rules, we can obtain:

$$h^2 = 6'75^2 + 100^2; h = 100'227$$

$$\sin \alpha = \frac{100}{100'22755} = 0'99773$$

$$\arcsin 0'99773 = 86'1^{\circ}$$

Existing algorithms in digital systems extract features from both cameras, process them, and try to match objects from both cameras frame by frame [10]. This process has hard computational costs and does not work in real time. We want to do the same thing in real time using AER. As a first step to achieve this goal, we propose an algorithm based on the spikes rate of the Multi Hold&Fire output [17], but not only on it.

If two retinas are focusing on the same point, the AER bandwidth of both retina outputs should be very similar for a moving object on the focus point. Therefore, the difference of traffic of both retinas should be the lowest. If the object is moving on the direction of the retinas and it is closer than the focus point, the traffic in both retinas should increase because the object becomes bigger and there are more active pixels, but the traffic difference cannot be cancelled because the object is not focused, so it is represented in different parts of the two retinas. Therefore, the difference of traffic should be greatly increased when the object is closer. When the object is further, since less pixels of the retina are stimulated, and they are also different in both retinas, the traffic is also higher than on the focus point, but it should be lower than when the object is approaching.

The spike rate of the differences of both retinas (Multi Hold&Fire output) is not the same for every object. A bigger object will fire more spikes than a small one. In order to have a normalized distance estimation, we have to combine the spike rate from the subtraction with the spike rate before the Multi Hold&Fire modifications. Thus, we should obtain more accuracy, better results and a more real estimation.

This proposed technique is a first step in the distance calculation problem. Our intention in a near future is to use this mechanism combined with another set of modules that have already been made in this group. One of these modules is related to stereo matching process [10], thereby, we think that the results may be better and we will improve the accuracy of this method by applying this technique.

Our final aim is to obtain a multi-level neural network where we combine results of different modules to obtain better results and, as a final step, the combination of all of them will result in the final distance calculation, which will have better accuracy than the algorithm proposed in this work alone. We are now working on a calibration module, so we would combine calibration, matching and distance estimation in a unique way.

B. Results

After the previous calculations and calibrations, our retinas are positioned with an angle of 86'1387° to obtain a focal

distance of one meter. After that, we have measured the spikes rates at the output of our Multi Hold&Fire algorithm using a vibrating object. This object has been moved closer and further from the retinas many times, so we obtain data from different distances; and we have been monitoring the number of spikes fired during this process. We have used this data with the spikes rate before the MH&F algorithm, like we said before. The resulting spike streams have been recorded using jAER software [7]. After measurements, we have recorded all the results and have made a graph with all of them. This graph indicates the number of spikes versus distance (see figure 5).

At the central match point of the focal length of each retina, the MH&F works like a perfect subtractor and will fire very few spikes (mainly retinas' spurious), so the spikes rate at this point is the smallest of the whole measuring points. If we go closer to the retinas, the spikes rate will be increased because the object becomes bigger and each retina sees it from a different point of view, so the subtractor will not act so perfectly. On the other hand, if we put the vibrating object further from the retinas, the spikes rate will be slightly increased because of the subtraction result (different retinas points of view), but the object becomes smaller, so it compensates this failure: the farther the object, the smaller it is; and, therefore, the lower the spike rate is (less spikes are fired, but the subtraction acts worse and fires more spikes). That is why these aspects balance each other.

As it is indicated by the experimental results in Figure 5, the number of spikes is greatly increased near the retinas and is slightly increased as we move away the object from the focal collision point. We have stimulated our system by using two objects with different sizes that produce different output retina bandwidth. Measurements have been taken from the starting point of 10 centimeters to 150 centimeters. They were taken every 10 centimeters.

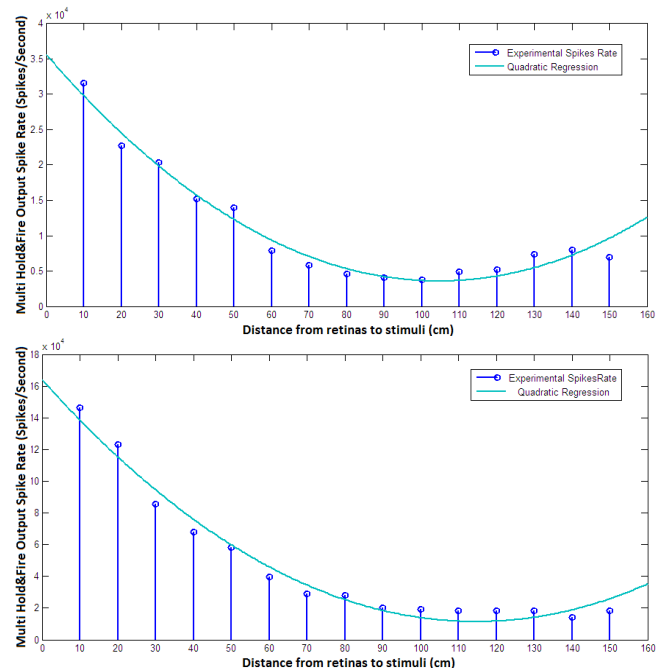


Figure5. Spikes Rate versus Distance with 2 different stimuli.

Figure 5 shows the experimental results obtained with two different stimuli: the first graph corresponds to a pendulum and the second one is related to an oscillatory ruler. It is interesting to remark that, approximately, at a distance of 100 centimeters (focal collision of both retinas) we obtained the lowest spikes rate. If we see measurements taken closer, it can be seen that the spikes rate increases, and far away from the focal collision point, the spikes rate is increased a little. The distance of the object in movement can be estimated with a quadratic regression (line in blue shown in figure 5).

Our system behaves quite similar to human perception: a priori, without knowing the size of the object, we cannot give exact distances, but only an approximation. This approximation depends on our experience, but the system proposed does not learn. However, we can measure the distance qualitatively, and interacting with an object in a nearby environment.

With regard to the comparison between this method and the classical computer vision methods, we have to say that this algorithm is only one of the steps implemented on a computer vision distance calculation. In order to be able to compare AER distance estimations with the classical computer vision one, we need to join it with a previous matching process and, in some cases, with a pre-calibration step.

Regarding the classical vision, the mechanism provided is somehow related to the method called "distance calculation using disparity map", which tries to determine a first-approach estimation about the distance of an object using the disparities between both cameras in a stereo vision system.

We are working on improvements for this system, using several cascaded filters that address strictions to the system and improve efficiency: AER stereo-vision matching [10], spatial corrections [16] and object tracking [10].

V. ACKNOWLEDGEMENTS

We want to thank the contribution of Tobias Delbruck and Raphael Berner, whose work has been used by our group for many years. This work has been supported by the Spanish government granted project VULCANO (TEC2009-10639-C04-02) of the MICINN.

VI. CONCLUSIONS

The difficulties to calculate distances in digital systems have been presented. That is why a biological approach to work with has been introduced (Address-Event-Representation), like a new paradigm in Neuromorphic Engineering. The advantages of this method have been explained and evaluated.

After that, a distance approach method has been proposed for working with moving objects using Address-Event-Representation in a near environment. To accomplish this goal, a stereoscopic vision system with two DVS retinas, working with VHDL over a Virtex-5 FPGA, was used.

The whole system used has been described and shown (and every element in it) to obtain the distance approach. With the hardware system described, the algorithms used have been explained in detail. First algorithm uses a method to obtain differences between both retinas in real time and without

sampling. With these differences, the second algorithm has been explained, which works with the spikes rate obtained in our system after the calculation of the differences. And, to obtain more accurate and realistic results, this algorithm has used the information of the original spikes rate.

With the results of these two algorithms, we have been able to model the spikes rate versus the distance of the object. The simulation results are very encouraging, because it can be seen in the graphs that there is a relation between distance and the spikes rate after our processing, and that this system functioning is quite similar to human perception.

REFERENCES

- [1] S. T. Barnard, M. A. Fischler: Computational Stereo. Journal ACM CSUR. Volume 14 Issue 4 (1982).
- [2] G. M. Shepherd: The Synaptic Organization of the Brain. Oxford University Press, 3rd Edition (1990).
- [3] A. Linares-Barranco et al: AER Convolution Processors for FPGA. ISCAS (2010).
- [4] M. Sivilotti: Wiring Considerations in analog VLSI Systems with Application to Field-Programmable Networks. Ph.D. Thesis, Caltech (1991).
- [5] Berner, R., T. Delbruck, A. Civit-Balcells and A. Linares-Barranco: A 5 Meps \$100 USB2.0 Address-Event Monitor-Sequencer Interface. ISCAS, New Orleans, 2451 – 2454 (2007).
- [6] P. Lichtsteiner, C. Posch, T. Delbruck: A 128x128 120dB 15 us Asynchronous Temporal Contrast Vision Sensor. IEEE Journal on Solid-State Circuits, vol. 43, No 2, pp. 566-576, (2008).
- [7] jAER software: <http://sourceforge.net/apps/trac/jaer/wiki>
- [8] Benosman, R. and Devars, J.: Panoramic stereo vision sensor. International Conference on Pattern Recognition, ICPR (1998).
- [9] Benosman, R. et al: Real time omni-directional stereovision and planes detection. Mediterranean Electrotechnical Conference. MELECON (1996).
- [10] M. Dominguez-Morales et al: Image Matching Algorithms using Address-Event-Representation. SIGMAP (2011).
- [11] A. Jimenez-Fernandez, et al: Building Blocks for Spike-based Signal Processing. IEEE International Joint Conference on Neural Networks, IJCNN (2010).
- [12] AVNET Virtex-5 FPGA board: <http://www.em.avnet.com/drc>
- [13] R. Serrano-Gotarredona, et al.:AER Building Blocks for Multi-Layer Multi-Chip Neuromorphic Vision Systems, NIPS (2005).
- [14] A. Linares-Barranco et al.: AER Neuro-Inspired interface to Anthropomorphic Robotic Hand, IJCNN (2006).
- [15] R. Serrano-Gotarredona, et al.: CAVIAR: A 45k-neuron, 5M-synapse AER Hardware Sensory-Processing-Learning-Actuating System for High-Speed Visual Object Recognition and Tracking. IEEE Trans. On Neural Networks, Volume 20, Issue 9, Sept. 2009.
- [16] A. Jimenez-Fernandez, et al.: Neuro-inspired system for real-time vision sensor tilt correction, ISCAS (2010).
- [17] M. Dominguez-Morales et al.: An Approach to Distance Estimation with Stereo Vision Using Address-Event-Representation. ICONIP (2011)
- [18] F. Gómez-Rodríguez et al.: Real Time Objects Tracking Using a Bio-Inspired Processing Cascade Architecture. ISCAS 2010