Address-Event Image Sensor Network

Eugenio Culurciello
E-Lab
Yale University
New Haven, CT 06520
eugenio.culurciello@yale.edu

Andreas Savvides
Embedded Networks and Applications Lab
Yale University
New Haven, CT 06520
andreas.savvides@yale.edu

***Abstract*— We describe a sensor network based on smart imager sensors able to extract events of interest from a scene. We present a framework towards the use of image sensors for realizing sensor networks that can interpret behaviors in physical space. An array of heterogeneous event-based CMOS imagers is employed at front-end, to convert light intensity, motion and other vision primitives into pulse-density modulated stream of address-events. The wireless nodes are Yale XYZ motes employing high data rate communication and large processing capabilities. Energy-aware communication is implemented at the sensor level by employing an event-based readout with pixel-level filtering capability. This paper provides an overview of the research challenges followed by a brief description of our current research efforts towards creating a functional experimental system for investigating these intelligent networks.**

I. Introduction

The last few years have experienced remarkable progress and increasing interest in the field of sensor networks. Several research groups have demonstrated that networks of compact wireless devices can advance the scientific efforts to understand how nature works by providing information from places that were not reachable before [1], [2], [3].

In this paper we analyze and present the research tradeoffs towards an image sensor network operating with a front-end composed of smart image sensors. An image sensor network is desirable for several reasons: first imagers very often provide a large quantity of information, that can be filtered on-demand or utilized to improve measurement quality and reliability, Second it has a natural interface to humans-in-the-loop, manual interpretation and is less prone to false-alarms, Finally imagers can be instrumented to provide analog computation on the image plane to sense intensity, motion, velocity and act as data-driven triggers and as feature extractors [4].

Our research vision is to combine smart image sensor technology with event-based computer vision algorithms to create an active sensor network that can detect complex behaviors in physical space. In this regard, we plan to pursue an integrative approach that exploits physical layer properties to build sensing functionalities in hardware that result in cheaper, simpler, smaller and power efficient sensors. These intelligent image sensors will be able to classify sensor data directly into a set of primitive events that can be directly used in a learning framework. This approach will essentially extract only the few useful bits of information at the sensor level, thus minimizing the computation and communication requirements of the network. This will provide a new approach for composing scalable systems capable of interpreting a wide range of behaviors in physical space. Example applications of such technology are plentiful and range from training scenarios for sports, search and rescue operations, indoor games, security and safety.

II. Smart image sensors for Sensor Networks

In order to be able to identify behavior in a scene sensors will be able to identify the *phonemes* of an action, while a sensor network node will process the phonemes in a time sequence to reconstruct *words*. The network will then be able to infer actions from the data forwarded by individual nodes and label them with text. This event-driven architecture requires a bottom-up approach where individual sensor are capable of pushing relevant information to the node and network, as opposed to being polled for information. Event-based image sensors and networks can detect features in a scene and trigger nodes to analyze and process the relevant data, while requiring network operation only after presence confirmation has been corroborated. In a sensor network communication is the most expensive operation. An event-based architecture will reduce communication costs by compressing and filtering information at individual nodes.

To make an event-based architecture plausible, image sensors have to be designed with a stringent set of requirements, here summarized:

- Power efficiency: ultra-low-power operation, consuming microwatts of power and allowing sensors to be always active.
- Large dynamic range and adaptability: sensors need to operate in a variety of lighting and environments while producing the same quality and quantity of information.
- Efficient use of bandwidth: filtering of data needs to occur at the focal plane of the sensor, thus requiring smart pixel architectures.
- Event-driven architecture: sensors will push data to a receiver using the address-event representation while filtering information.
- Common interface to heterogenous sensors and processing nodes: the network can be augmented with different image and non-imaging sensors that need to communicate with the same protocol.

0-7803-9390-2/06/$20.00 ©2006 IEEE

- Low manufacturing and operational costs: sensor network operate with large numbers of nodes so sensors need to be cheap and reliable.

In a sensor network requiring an image sensor as an event-based front-end detector, commercial off-the-shelf (COTS) image sensors cannot be used. Firstly because these are polled sensors and are not able to push data to the output. COTS *blindly* collect all the visual information whether the scene contains new and useful data or not. Image data is notoriously large even for small image sensor operating at video rates (33fps). As an example COTS VGA sensor generates 2.4Mbits/s because of the large imaging array. Large arrays are not needed in a sensor network because of the multiplicity of sensors and the need of extracting information. In sensor network application aimed at understanding behavior the information sought are often very few bits, usually 1 bit: is there a person in the scene? Is there motion in the scene? Identifying behavior requires only few additional bits: is she sitting? Sleeping? Walking? Exercising?

The Cyclops COTS module [5] for MICA motes developed by UCLA and Agilent employs an external processor a CPLD, SRAM and FLASH memory, offering CIF-resolution image processing at the expense of large power consumption (excess of 50mW). COTS camera module, although often labeled 'low-power' consume large amount of power, often more than 50mW. The power consumption is attributed to the relatively high resolution of the sensor, color processing, compression, multiple readout and interfacing standards. All of these features are of little significance to sensor network nodes, where low power consumption is instead a much more desirable feature. COTS modules can last only 1 day running on two AA batteries, while measured data from custom image sensors in research papers yields 6 days [6], 13 days [7], and 4.5 years of our most recent prototype [8].

We argue that a sensor network that recognizes behavior cannot be implemented without efficient sensing of events, rather than bits [9]. Furthermore, if data reduction is not performed at the sensor level, vision algorithms must be used to extract features from an image [10], [11]. In this case, the amount of computation to extract features from the visual data is practically always too high for the limited power budget of a sensor network node. In this context we propose the use of an ultra-low-power image sensor that can detect specific features of interest. Our image sensors use an event-based digital representation of the image data to encode and compress information. This encoding can be used for heterogenous sensors thus simplifying data processing. They use a thousand times less power than conventional image sensor because they only output the relevant information (intensity of brightest pixels, motion, contour) as opposed to entire frames. They provide very large dynamic range and require no external components. They are cheap to manufacture because of the small array sizes and the use of older CMOS processes. We will here describe our initial prototypes of efficient image sensor that encode light intensity into a frequency of events.

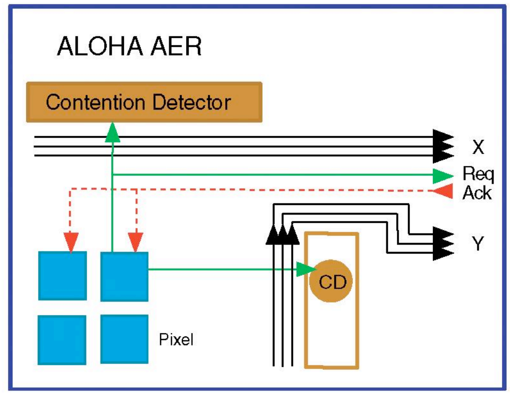

Fig. 1. ALOHA image sensors architecture. A pixel generates and event, the event generates a request (Req) and outputs the pixel X, Y addressed on the output bus. A collision detector circuit detects multiple rows or columns trying to access the output bus at the same time. The receiver circuits responds with an acknowledge (Ack) as a confirmation of detecting the event and reading the image sensor output bus. CD is the output bus collision detection circuitry.

III. Address-Event Image Sensors

Our image sensors utilize the asynchronous event-based digital representation called Address-Event Representation (AER) [12]. In the AER terminology, *events* are communication primitives sent from a sender to one or more receivers. For an image sensor, events are individual pixels reaching a threshold voltage and accessing the bus for communication with an outside receiver. An AER image sensor is composed of an array of pixels, and the digital output of the image sensor is the address of the pixel that is to communicate an event (refer to Figure 1). The integration time of each pixel varies in relation to the incident light intensity. Since the activity of the array is generated by the light intensity of the scene, and not an external scanning circuitry, the rate of collection of frames can be modulated by varying the *request-acknowledge* cycle time between the imager and the receiver circuitry. *Thus information can be extracted on demand from individual nodes in a wireless sensor network.* The address-event representation of an image can be thought as a realization of a delta-sigma pixel-parallel analog to digital converter [6]. This corresponds to a dynamic allocation of the output bandwidth determined by *need* as opposed to scanned sensors, which allocate uniformly and inefficiently the bandwidth across pixels. Thus, representing intensity in the time domain allows each pixel to have large dynamic range [13], [14].

A. The ALOHA Image Sensor

Our second-generation prototype camera for sensor networks [15] is a CMOS image sensors with digital output fabricated in a 0.6m CMOS process. Figure 1 shows the architecture of the address-event sensor. The imager, named ALOHA, performs a light-to-frequency conversion at the pixel level. The four quadrant, 32 x 32 pixels image sensors present very high dynamic range and ultra-low power operation targeted to energy aware sensor network applications. Two samples collected, respectively of the 'analog devices' sign

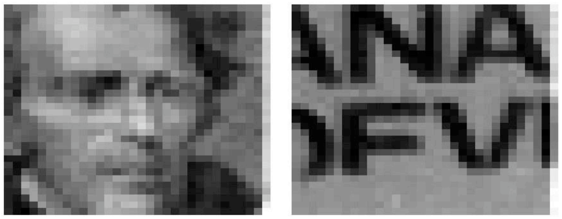

Fig. 2. Example images collected with image sensor ALOHA. President Jackson and 'Analog Devices' text.Images were collected with the sensor operating at 7.6Kevents/s.

and President Jackson of a 20 dollars bill are reported in Figure 2. The image sensors produced approximately 10,000 events in a period of 1375ms, while the current consumption of each quadrant for this output event rate was $2.5\mu A$ at 3.3V. This corresponds to a power consumption of $5.75\mu W$ for each quadrant, a value that would scale to $60\mu W$ for a VGA size. With this power consumption the sensor can run on two 1000mAh AA batteries for 4.5 years [15].

IV. Camera Enabled Sensor Nodes and Testbed

The sensor network nodes used are the custom designed XYZ sensor nodes [16] designed for mobility and learning applications. The XYZ sensor nodes run SOS, a lightweight operating system, that supports dynamic module loading and unloading at runtime [17]. This provides a powerful development tool that allows us to update the applications running on our testbed from a remote locations. Furthermore, it provides a flexible framework to start developing the proposed learning applications using real data and real-time operation. We have instrumented the XYZ sensor nodes with three different imager setups.

- The first one is a custom interface to the ALOHA address-event camera prototypes. Our prototype is shown in Figure 4 together with a sample image collected by the network.
- The second is a small board based around a COTS camera module from Omnivision (OV). The XYZ-OV node is shown in figure 3. This is used for acquiring images over the network and for fast prototyping of standard vision and image processing techniques. The COTS camera was also designed to provide a comparison with the performance of the ALOHA sensor node.
- The third is also based on the COTS camera module by OmniVision, but with the addition of an Analog Devices BlackFin DSP. The XYZ-OV-DSP is used to model the operation of the address-event image sensors in software. Since the design and optimization of custom image sensors requires approximately 1 month for design and 2-3 months for fabrication and testing, it is desirable to be able to test ideas, advanced focal-plane processing, and pixel-level computation using a model. This model provides the same address-event output, thus allowing to focus on the development of event-based algorithms to process image data and recognize behavior in the scene. Lastly, the XYZ-OV-DSP module is used to prototype computation demanding vision and image processing algorithms requiring a larger computational power than the one available on the XYZ node.

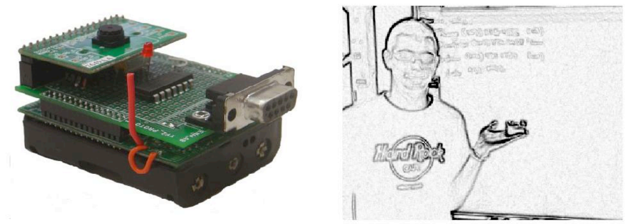

Fig. 3. XYZ sensor node with a COTS camera module (left) and an example of a captured image after *Sobel* edge detection (right).

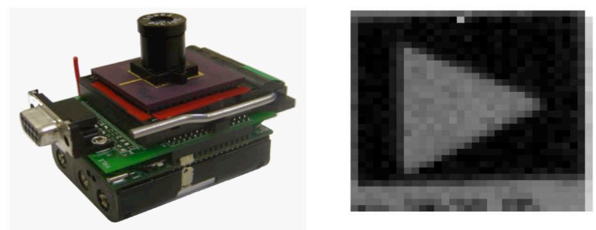

Fig. 4. XYZ sensor node with the ALOHA image sensor (left) and example of a captured image (right).

To experiment with learning behaviors from heterogeneous sensor networks we have constructed a multimodal sensor network testbed that is deployed inside the computer science building and the Morse Teaching Center at Yale University. Half of the nodes are deployed on a three-dimensional testbed installed in a single room. The rest of the nodes are deployed across a busy corridor in the Morse Teaching Center that connects the computer science building with the engineering building at Yale. This testbed features a heterogeneous set of sensors including imagers, passive infrared, light, temperature and acoustic sensors.

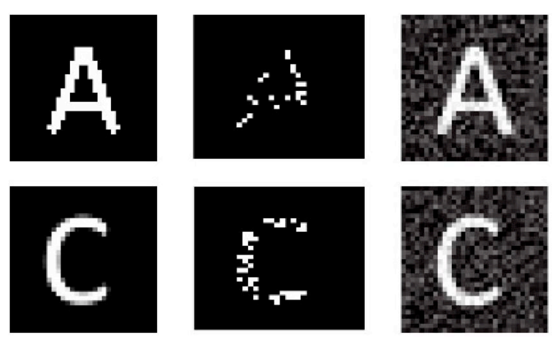

Fig. 5. Experiments in character recognition using the ALOHA sensor.

We are currently testing a few recognition algorithms using the intensity-based ALOHA image sensor. One experiment consists in detecting letters of the alphabet and sign-language. Figure 5 shows the characters used as input stimuli (left), the data received in a 20ms interval (middle), and the reconstructed image after 500 events (right). Due to the nature of address-event data, it is possible to provide recognition of an item in a small set by monitoring and computing an histogram of the incoming events. Since only a fraction of the pixel is needed to recognize an item, the algorithm will be much faster and power-efficient than using a COTS and a DSP to collect and process the image. In order to improve the speed of recognition, we use rank encoding of plausible targets based on recent work on primate vision studies [4].

V. Advanced Address-Event Image Sensor Networks

In order to improve the performance of address-event sensors for sensor networks, we have compiled a detailed list of additional specifications and improvements that will be featured in the next generations of sensors:

- Enhanced ALOHA access technique: we will add an event queue as an interface between the pixel array and the address-event output bus. The queue will reduce collision on the ALOHA bus and provide a less stringent asynchronous timing interface for the readout of events.
- Pipelined readout architecture: adding buffers and queues to individual rows and columns of pixels enhances the throughput of the sensor.
- Serial output using standard protocols: the output event data will be serialized with an high-speed SPI channel to reduce the wiring to the node and take advantage of its SPI serial inputs and DMA channels.
- Global adaptation of output event rate: the event rate of the sensors can be monitored and fed back to analog bias networks to maintain a constant data rate. This is desired for adaptation and so that the receiver is not flooded with a sudden burst of data.

All the improvements on sensor and interfacing are targeted to reduce the power consumption to the theoretical minimum in CMOS process and to streamline the delivery of data to the sensor network nodes. Fast readout will contribute to the reduction of latency for timely recognition of events in the scene. Finally, we are designing a ultra-low power motion sensor to act as triggering device.

VI. Conclusion

In this paper we motivated the use of address-event image sensors for application in a sensor networks for understanding behaviors. To realize these networks we identified the main research directions for the development of custom multifunction imager. We have also described some of the hardware infrastructure towards achieving this goal. In the near future, we will provide details on our preliminary event-based algorithms to detect features towards a learning based framework. Up-to-date information on this effort can be obtained from our websites at http://www.eng.yale.edu/enalab and http://www.eng.yale.edu/elab.

Acknowledgment

This work was funded in part by the National Science Foundation CISE/CNS award #0448082. The authors would also like to thank the members of ENALAB, Dimitrios Lymberopoulos, Andrew Barton-Sweeny and Evan Park for their help with this article.

References

[1] M. Duarte and H. Yu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, pp. 826–839, July 2004.

[2] R. J. Nemzek and J. Dreicer, "Distributed sensor networks for detection of mobile radioactive sources," *IEEE Transactions on Nuclear Science*, vol. 51, pp. 1693–1701, August 2004.

[3] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Destrin, "Habitat monitoring with sensor networks," *Communications of the ACM*, vol. 47, pp. 34–41, June 2004.

[4] S. Thorpe, A. Delorme, R. Van Rullen, and W. Paquier, "Reverse engineering of the visual system using networks of spiking neurons," *IEEE International Symposium of Circuits and Systems, ISCAS*, vol. 4, pp. 405–408, 2000.

[5] M. Rahimi, D. Estrin, R. Baer, H. Uyeno, and J. Warrior, "Cyclops: image sensing and interpretation in wireless networks," in *Second ACM Conference on Embedded Networked Sensor Systems, SenSys*, Baltimore,MD, November 2004.

[6] L. McIlrath, "A low-power low-noise ultrawide-dynamic-range CMOS imager with pixel-parallel A/D conversion," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 846 – 853, May 2001, issue 5.

[7] K. Cho, A. Krymski, and E. Fossum, "A 1.5-V 550-μW 176 x 144 Autonomous CMOS Active Pixel Image Sensor," *IEEE Transactions on Electron Devices*, vol. 50, no. 1, pp. 96–105, 2003.

[8] E. Culurciello and A. G. Andreou, "ALOHA CMOS imager," in *IEEE International Symposium on Circuits and Systems, ISCAS*, Vancouver, Canada, May 2004, pp. IV– 956–9.

[9] R. Van Rullen and S. Thorpe, "Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex," *Neural Computation*, vol. 13, pp. 1255–1283, 2001, massachusetts Institute of Technology.

[10] E. M. C.F. Chiasserini, "Energy-efficient coding and error control for wireless video-surveillance networks," in *Telecommunication Systems*, vol. 26, June-August 2004, pp. 369–387.

[11] M. Tehrani, P. Bangchang, T. Fujii, and M. Tanimoto, "The optimization of distributed processing for arbitrary view generation in camera sensor networks," in *IEICE Transactions On Fundamentals Of Electronics Communications And Computer Sciences*, vol. 8, August 2004, pp. 1863–1870.

[12] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 47, no. 5, pp. 416–434, 2000.

[13] E. Culurciello, R. Etienne-Cummings, and K. A. Boahen, "A biomorphic digital image sensor," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 2, pp. 281 –294, February 2003.

[14] O. Yadid-Pecht and A. Belenky, "In-pixel autoexposure CMOS APS," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 1425–1428, August 2003.

[15] T. Teixeira, E. Culurciello, and A. Andreou, "Event-based imaging with active illumination in sensor networks," in *IEEE International Symposium on Circuits and Systems, ISCAS*, Kobe, Japan, May 2005.

[16] D. Lymberopoulos and A. Savvides, "Xyz: A motion-enabled, power aware sensor node platform for distributed sensor network applications," in *Proceedings of Information Processing in Sensor Networks (IPSN), SPOTS track, April 2005*, Los Angeles, CA, April 2005.

[17] C.-C. Han, R. Kumar, R. Shea, E. Kohler, and M. Srivastava, "A dynamic operating system for sensor nodes," University of California Los Angeles, Networked Embedded Systems Lab, Tech. Rep. NESL-TM-2004-11-01, November 2004. [Online]. Available: http://nesl.ee.ucla.edu/projects/sos