



Demo: Results of 'iCaveats', a Project on the Integration of Architectures and Components for Embedded Vision

R. Carmona-Galán,
J. Fernández-Berni,
Á. Rodríguez-Vázquez

Instituto de Microelectrónica de Sevilla
IMSE-CNM
CSIC-Universidad de Sevilla (Spain)
rcarmona@imse-cnm.csic.es

P. López-Martínez,
V. M. Brea-Sánchez,
D. Cabello-Ferrer

Centro Singular en Tecnologías de la Información (CiTIUS)
Univ. Santiago de Compostela (Spain)
p.lopez@usc.es

G. Domenech-Asensi,
R. Ruiz-Merino,
J. Zapata-Pérez

Dpto. de Electrónica, Tecnología de Computadoras y Proyectos
Univ. Politécnica Cartagena (Spain)
gines.domenech@upct.es

ABSTRACT

iCaveats is a Project on the integration of components and architectures for embedded vision in transport and security applications. A compact and efficient implementation of autonomous vision systems is difficult to be accomplished by using the conventional image processing chain. In this project we have targeted alternative approaches, that exploit the inherent parallelism in the visual stimulus, and hierarchical multilevel optimization. A set of demos showcase the advances at sensor level, in adapted architectures for signal processing and in power management and energy harvesting.

CCS Concepts

• **Hardware** → **Communication hardware, interfaces and storage** → **Sensor devices and platforms.**

Keywords

Vision chips, smart image sensors, embedded vision systems, parallel concurrent processing.

1. INTRODUCTION

The inclusion of embedded vision systems in mobile platforms for the intelligent transportation of people and goods would represent an important technological leap forward. It is so also in surveillance applications for security and defense. However, the implementation of a compact autonomous vision system, with low power consumption is not an easy task to accomplish [1]. The visual stimulus contains a great deal of data. In order to process them, a considerable computational effort needs to be put in place. In practical terms, this means to be able to realize several million operations per second. Achieving this performance with a restricted power budget is very difficult. In a conventional image processing chain—where images are captured, then digitized and stored in a memory and then processed—we will most probably arrive to impossible specifications for the sensor, the analog-to-digital converter, processor and memory.

A viable alternative is to take advantage of the inherent

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

ICDSC '18, September 3-4, 2018, Eindhoven, Netherlands

© 2018 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6511-6/18/09

<https://doi.org/10.1145/3243394.3243707>

parallelism of early vision tasks. In order to do that, part of the low level processing can be conveyed to the focal plane [2]. The distributed implementation of the processing resources implies a reduction on data transfers to and from the memory. Besides, these processing elements can be built with analog and mixed-signal circuit blocks, which can be very efficient. The main problem of this approach is that an ad hoc implementation lacks flexibility to be migrated to other application fields. In addition, computer vision experts and application developers, who work at higher abstraction levels, do not easily handle hardware programming at low level. In order to bridge this gap, an important industrial consortium has been created very recently to generate standards for the hardware acceleration of computer vision and, in general, massive sensory signal processing. OpenVX [3], for instance, defines, at a layer that is just above the hardware, a set of functions that can be employed by computer vision application developers who also search for a power-optimized implementation.

The objective of project 'iCaveats' is the capitalization of the acquired know-how by developing a library of hardware components and architectures that follow these principles. Along the project, we have worked in adapted processing blocks for hardware acceleration of low-level and medium-level image processing tasks, in new sensor abilities like photon counting and time-of-flight estimation in concurrency with conventional imaging, and aspects more related to the system level, like energy management and interfacing with other signal processing chips.

We present a set of demos that showcase the advances towards an integrated vision system on-a-chip for intelligent transportation and security applications.

2. VISITOR EXPERIENCE

A combination of posters, videos and experimental setups has been arranged to demonstrate advances towards an integrated vision system. These contributions are organized into four different sections: the smart image sensor, adapted image and video processing architectures, power management and energy harvesting, and embedded vision systems based on feature learning.

Smart image sensors

In embedded vision systems, a large fraction of power is dedicated to the storage, processing and shuffling of data that later on will become irrelevant. In this project we have explored different architectural alternatives oriented to the reduction of redundancies, like close-to-sensor or in-sensor realization of early

vision tasks, feature extraction and key points detection and the fusion of 2D and 3D information.

Two different sensors are showcased: a CMOS-SPAD image sensor able to generate 2D and 3D representations of the scene, and a linear range HDR image sensor. The first one employs photon counting and direct ToF estimation to capture perfectly aligned intensity and depth images. The second is based on the asynchronous tagging of pixel saturation events to extend the dynamic range. In both cases, signal processing at pixel-level provides additional functionality.

Adapted image processing architectures

In this work package, the complete SIFT processing chain has been implemented in dedicated hardware in a FPGA. To achieve real time requirements, pipeline structures have been widely exploited both in the keypoint extraction and in the descriptor generation stages. Simplifications to the original algorithm have been applied. The proposed architecture has been synthesized on a Xilinx Virtex 5 FPGA. It generates 3072 descriptor vectors for VGA images at 99 frames per second at a clock rate of 100 MHz.

Energy harvesting in image sensors

This live demo shows a micro-energy harvesting system which includes a 1 mm² solar cell as the unique power source and a Power Management Unit (PMU) on the same substrate in standard 0.18 μ m CMOS technology. The PMU has cold start-up from nW and it also performs a continuous and two-dimensional maximum power point tracking using analog strategies to meet very low power consumption, managing a high input power range.

Vision systems based on feature learning

At this level, we present a system for multiple object detection and tracking on the Jetson TX2, a Nvidia platform for embedded AI. The system is based on a combination of a hardware-oriented pixel-based adaptive segmenter (HO-PBAS) and the GoTURN a tracker based on CNNs, with extensions for multi-object tracking.

Besides, six state-of-the-art CNN models for 1000-category classification running on the most popular Deep Learning (DL) frameworks have been evaluated. Three key performance metrics are benchmarked, namely power consumption, throughput and precision. Further assessment is provided through a Figure of Merit (FoM) based on high-level specifications. Accordingly, we report the reachable performance of DL on embedded vision systems, and also enable the comparison of DL components across different application requirements.

3. CONCLUSIONS

This project has implemented an integrative approach to incorporate efficiency at sensor, processing and system levels. We have considered multi-level hierarchical optimization in order to benefit from emerging capabilities but, at the same time, keeping a recognizable and operable system architecture. In summary, we have advanced towards an integrated power-efficient vision system for the coming vision-enabled IoT.

ACKNOWLEDGMENTS

The author wants to acknowledge the work of the rest of the participants in this project, namely: B. Pérez Verdú, J. A. Leñero-Bardallo, I. Vornicu, A. Darie, F. Bandi, M. Trevisi and D. Velasco-Montero from IMSE-CNM, E. Ferro-Santiago, J. Illade-Quinteiro, D. García-Lesta from CiTIUS, J. A. López-Alcantud, J. Á. Díaz-Madrid, J. J. Martínez-Álvarez and P. Rubio Ibáñez from UPCT; and also the assistance of the external collaborators: M. Pelcat from Université Rennes, F. Berry from Université Clermont-Auvergne, T. J. Kazmierski, University of Southampton, M. Balsi, Università di Roma, La Sapienza, C. G. Bran, Universidad Don Bosco, El Salvador and J. M. Carrillo-Calleja, Universidad de Extremadura.

This work was mainly supported by the Spanish MINECO and the European Region Development Fund (ERDF/FEDER) through project ‘iCaveats’ (Ref. TEC2015-66878-C3-1-R, TEC2015-66878-C3-2-R and TEC2015-66878-C3-3-R), and partially supported by Junta de Andalucía through project ‘SmartCIS3D’ (Ref. TIC 2338-2013), by the Consellería de Cultura, Educ. e Ord. Universitaria and ERDF (accreditation 2016-2019, ED431G/08 and reference competitive group 2017-2020, ED431C 2017/69) and by EU-REA through project ‘Achieve’ (EU H2020 MSCA-ITN 2017, Grant No. 765866).

REFERENCES

- [1] J. Fernández-Berni at al. *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*, Springer (2012)
- [2] A. Zarandy (Ed.), *Focal-plane sensor-processor chips*. Springer (2011).
- [3] Khronos group, “OpenVX 1.0.1, hardware acceleration API for computer vision applications and libraries,” <http://www.khronos.org/openvx> (2015).

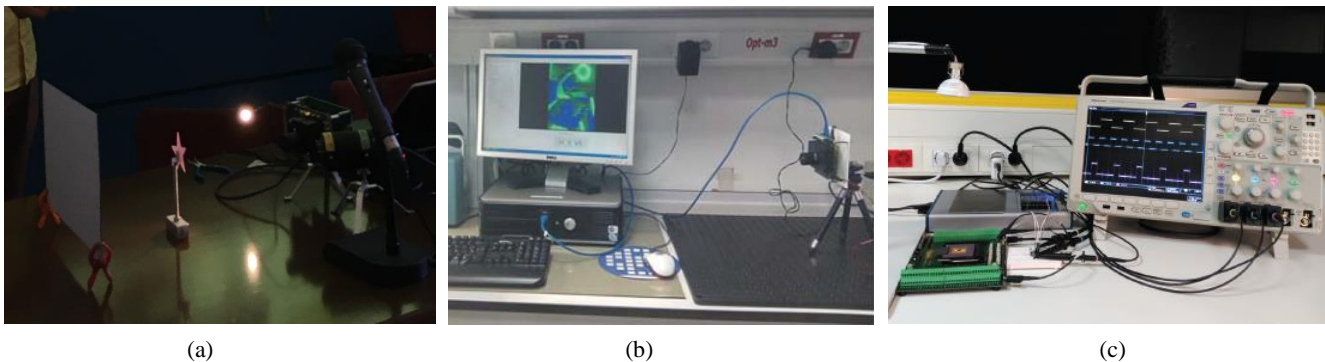


Fig. 1. Experimental setups for different chips designed, fabricated and tested in project ‘iCaveats’: (a) CMOS-SPAD sensor, (b) Event-driven HDR sensor, (c) Energy harvester chip.