Hardware Implementation of Smart Embedded Vision Systems

Elisa Calvo-Gallego, Piedad Brox and Santiago Sánchez-Solano
Instituto de Microelectronica de Sevilla IMSE-CNM, CSIC - University of Seville, Seville, Spain

1 INTRODUCTION

The research presented in this contribution is focused on the efficient hardware implementation of image processing algorithms that are present at different levels of a smart vision system. The system is conceived as a reconfigurable embedded device which, in turn, will be a node of a collaborative sensor network.

The inclusion of fuzzy logic techniques is explored to improve the performance of conventional vision algorithms.

This project is integrated into the research line dedicated to 'Embedded Systems' of the 'Micro-electronics' PhD Program from the University of Seville ("Microelectronics doctorate program"). The author is financially supported by one of the most prestigious fellowship in Spain, the FPU Program from the Ministry of Education of the Spanish Government. The work is partially funded by project TEC2011-24319 from the Spanish Government with support from FEDER.

2 RESEARCH PROBLEM

Digital image/video processing is a key discipline due to the wide range of applications in which it could be used. Not only it is required by necessity in professional areas such as industry (in sectors as automotion, packing, robotics, etc.), medicine (real-time monitoring of cells or viruses, rehabilitation-physical therapy, etc.), environment (fire detection, animal population monitoring, etc.) or security (building/area surveillance), but also it is required in consumer products for applications more related to entertainment and enjoyment (Figure 1).

As a consequent of this relevance, many efforts are being invested in the development of new systems able to provide improved functionalities and to support emerging applications. Many of these new applications require the integration of vision systems in embedded devices, like PDAs, mobile phones or sensor networks (Figure 2).

The majority of the vision algorithms proposed in literature are not conceived to be implemented on platforms with limited resources being necessary an adaptation process. In addition, many applications require autonomous devices, which demand solutions with low-power consumption. However, these systems have to work in real-time demanding a high processing time, which is besides continuously increasing due to the use of high resolution cameras. These facts force designers to study different options as alternative to classical software implementations on CPUs or GPUs, whose computational/ programming models are far to satisfy real-time requirements of the majority of video standards. Moreover, they involve a high economic cost and they are not a low-power consumption solution.

In this sense, the use of hardware/software co-design methodologies and Field Programmable Gate Arrays (FPGAs) could be a good idea to accelerate

Figure 1: Examples of applications where image or video processing is essential (a) Aerial Topography. (b) Medicine. (c) Security. (d) Environmental Sciences. (e) Industry. (f) Entertainment. (g) Robotics.

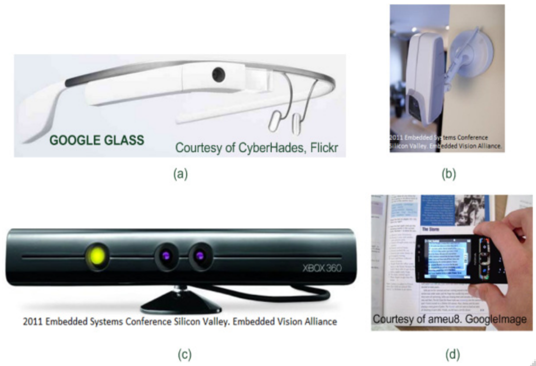

Figure 2: Examples of vision embedded systems (a) Google Glass (b) Security Camera (c) Kinetic. XBox Game Console (d) Text Reader on a mobile phone.

the development of smart video embedded systems required by this kind of applications.

The hardware/software co-design concept stated to emerge in 1990s. It consists of designing a mixed hw/sw system where the coexistence between the two kinds of components is taken into account along all the design process. This design strategy allows designers to accelerate software implementations by moving timing-critical tasks to hardware. In the last years, a set of new computer aided design tools (i.e. Vivado HLS) has appeared which allows the exploitation of hw/sw co-design techniques in FPGAs. These electronic components are essentially high density arrays of uncommitted logic. They are very flexible devices where developers can establish trade-offs between resources and performance by selecting the appropriate level of parallelism to implement an algorithm. In this way, FPGAs could be excellent platforms for final hardware realization or for systems prototyping to be implemented as application specific integrated circuits (ASIC).

In addition to the challenges of adapting vision algorithms to new platforms and its integration on reconfigurable devices, others research problems to tackle are related to common difficulties in image and video processing. For example, noise or illumination changes could be a problem in all the stages of a vision system.

These obstacles can be faced up modifying the existing algorithms with the knowledge coming from other areas such as statistic or soft-computing. This last term was introduced in the middles of 90s by Zadeh to encompass a set of techniques (fuzzy logic, neuro-computing, etc.) that allow designers to manage uncertainty and vagueness inherent to many natural problems, trying to emulate the human reasoning. These properties have been widely exploited in low- and middle-level tasks in image and video processing. Some examples are noise filtering and video de-interlacing, where the use of neuro-fuzzy techniques improves the performance of conventional algorithms.

Finally, more complex vision systems are distributed. This means that this kind of algorithms are implemented in a collaborative network, where several nodes co-operate to carry out the assigned work. This strategy provides a lot of improvements in performance and computing. However, it presents some problems as a consequence of the limited network capacity, the way in which the information of the different nodes is exchanged, shared, protected and processed to validate a decision or to rectify errors, and the manner in which the partition of tasks is done.

3 STATE-OF-ART

As this PhD project addresses the general flow of a vision system, it would be difficult to include here a complete review of the contributions in the state-of-art. A block diagram of a complete vision system is illustrated in Figure 3. Each one of the algorithms that are implemented in each stage is briefly introduced herein.

A large variety of papers has been published regarding low-level processing operations. Among them, lens distortion / correction, color space conversion, feature detection (edges or corners), filtering (noise reduction) or picture enhancement (contrast improvement in the image by means of the redistribution of the pixel values) could be found. Normally, low-level processing algorithms are relatively simple and they could be processed in pixel time without needing to consume a large amount of resources. Some works in this level are (Bailey, 2011)(Wnuk, 2008). The experience of our research group in this area has allowed the development of a library of hardware Intellectual-Property modules (Garcés-Socarrás et al., 2013).

Some examples of medium-level processing algorithms are background subtraction (moving objects in a scene are identified), labeling (connected components in an image are identified in a unique way) or segmentation (objects or regions with similar properties in an image are isolated). Hardware implementation of this kind of algorithms is more complex. They could need several frame-buffers to save intermediate results and its implementation in real-time is not always achieved.

Labeling algorithms are classified in the literature attending to multiples criterions such as the level of parallelization or the way in which the

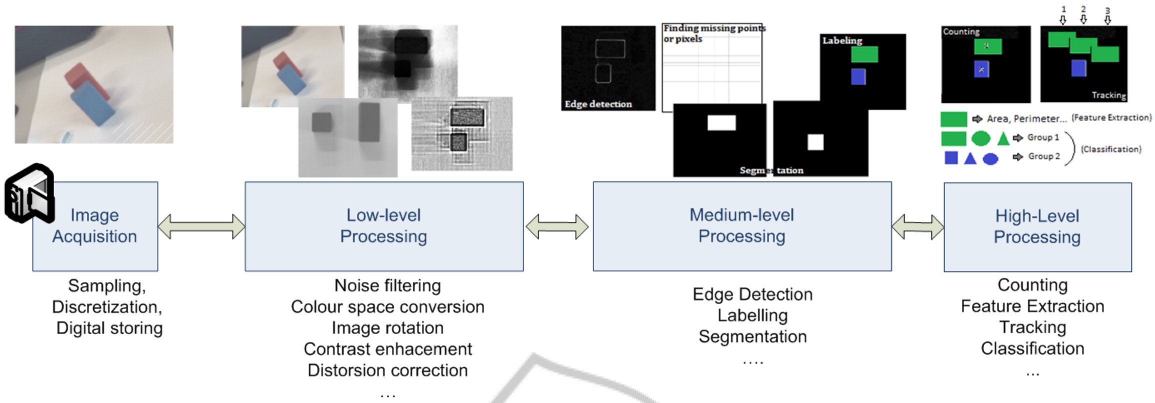

Figure 3: Complete vision system.

image is represented. However, a classification according to the number of image scans is found in (Calvo-Gallego, 2011). In a first group, one-scan algorithms, as region growing, contour and feature extraction algorithms, are described. The main drawback of many of them is an irregular and random mode to access to memory. Multi-scan algorithms, in the second group, have simple hardware implementation due to its regular accesses to memory, but their execution time depends on the position of the pixels in image, so it is impossible to determine its duration and, therefore, to achieve real-time operation. Third group is composed by two-scan algorithms. Proposals of two-scan algorithms differ from each other in the method and data structure used to save label equivalences, and in the way it is performed the final resolution. A good algorithm of labeling is provided on (Bailey and Johnston, 2007).

On other hand, background subtraction algorithms are usually classified into different categories: basics (frame differencing, mean filtering, median filtering, etc.); statistical (Gaussian model-based, support vector based, learning subspaces based); of estimation (application of Kalman filter, wiener filter, etc.); neural-network and fuzzy logic modeling; and clustering. A review of existence methods could be found in (“BS_Rev,” n.d.). In terms of complexity, basic methods could be implemented in hardware, although they consume a lot of memory since they are based in the analysis of several previous frames. Other options are the simplification and adaptation of complex methods (Appiah and Hunter, 2005).

Finally, complex feature extraction (color, texture, position, shape, motion, etc.), stereo vision and tracking techniques are included among high-level processing algorithms. Hardware implementation of feature extraction algorithms have been recently proposed in (Svab et al., 2009)(Schaeferling and Kiefer, 2010). Stereo vision systems obtain the position of the points in the scene from several images. The key problem is the selection of characteristics points in one of the images and its identification in the other ones. Hardware implementations are performed using sum-of-absolute-differences or sum-of-squared-differences. A complete revision can be found in (Lazaros et al., 2008). Tracking techniques are used to monitor the movement of an object. A review of classical algorithms can be found in (Yilmaz and Javed, 2006). Hardware implementations are provided on (Cho et al., 2006) and (Fan Yang and Paindavoine, 2003).

High-level algorithms are complex and, in occasions, it is necessary to use more than one reconfigurable device to provide a real-time solution.

Concerning collaborative sensor networks, some fundamental ideas have to be explored. Although there are some publications in which the network is composed by independent cameras (Stillman et al., 1999), it is more frequent to find works focused on the learning of a topology of a network (Zhao et al., 2008), the way of calibrating the system (Lobaton et al., 2010) or the control or parameters definition in pan, till and zoom networks (Everts et al., 2007). Regarding the applications developed over these distributed networks, object detection, tracking, recognition or pose estimation could be found in (Sankaranarayanan et al., 2008), (Chen Wu and Aghajan, 2008).

4 OBJETIVES

The main objective of this research is the design of efficient image/video processing algorithms tailored for hardware implementation on reconfigurable devices. Based on this idea, the research will cover these three lines:

- The improvement of existing algorithms with a double purpose: its integration in embedded devices and the increasing of its performance by means of the use of soft-computing techniques (Nachtegael et al., 2007).
- The efficient hardware implementation of algorithms into reconfigurable devices.
- The use of design methodologies that allow us to reduce implementation and verification times. Specifically, a model-based design methodology based on Matlab/Simulink and Xilinx Tools, which provides a common integrated framework to cover all the steps in the design flow (from software implement-tation to hardware co-simulation), will be used.

5 METHODOLOGY AND TOOLS

The followed methodology to develop each block of the system will be:

- Review of the State-of-Art: The initial step in each block is to review the fundamentals as well as previously published works. In this way, enough knowledge to face up to the problem will be acquired.
- Software Implementations: For a better understanding of the studied methods, and in order to compare the results with the obtained in other works, software implementations of analyzed algorithms will be developed.
- Studies about improved Algorithms: Once the limits of current methods have been evaluated, the incorporation of new proposals will be analyzed. Soft-computing techniques could be applied in some cases. In this point, algorithms suited for a hardware implementation will be especially considered.
- Design and Hardware Implementation of Final Algorithms: A microelectronic design of the algorithms for a reconfigurable device will be developed. Different options to optimize area and timing will be considered to achieve the goals. Moreover, advantages, constraints and cost of a possible hw/sw partition must be studied to find the optimal solution.
- Verification stage: To verify the desired behavior of the block and characterize it from the point of view of resources, operation speed, etc.
- Integration as IP Core: The adaptation of the designed blocks for its integration as IP Core of standards embedded microprocessor on FPGA.

Once completed the design of the considered blocks, a demonstrator of a whole system and a prototype of the network will be built. Among the applications, environmental, security and surveillance will be considered.

5.1 State of Research

This research work started after finishing the Master in Microelectronics (“Microelectronics Master,”). As final project of this master some hardware implementations of connected component labeling algorithms were developed (Calvo-Gallego et al., 2012b). Two simple demos to illustrate theirs applications in counting and tracking were also included. After that, a new implementation that takes advantage of the blanking periods in video standards and temporal parallelism was proposed. This last implementation was integrated on a Spartan 3A DSP 3400 development board and it was able to process VGA (640x480) video sequences from Micron MT9V022 camera (Calvo-Gallego et al., 2012a).

After developing a deep study about state-of-art in background subtraction, the student proposed in a recent publication an algorithm to improve background subtraction using fuzzy logic (Calvo-Gallego et al., 2013).

Currently, her work is centered on developing an efficient hardware implementation of this algorithm.

6 EXPECTED OUTCOME

Efficient hardware implementation of a smart embedded vision system is going to be carried out. This system will be a node of a distributed sensor network, able to tackle complex tasks. Environmental and surveillance applications for this network will be considered. It is expected to transfer this knowledge to industrial companies.

REFERENCES

Bailey, D.., Johnston, C.., 2007. Single Pass Connected Components Analysis.

Bailey, D.G., 2011. Design for embedded image processing on FPGAs. John Wiley & Sons (Asia), Singapore.

BS_Rev [WWW Document], n.d. URL https:// sites.google.com/site/thierry- bouwmans/background-subtraction

Calvo-Gallego, E., 2011. Implementación sobre FPGAs de algoritmos de procesamiento de imágenes para etiquetado de componentes conectados (Trabajo Fin de Máster Máster en Microelectrónica: Diseño y Aplicaciones de Sistemas Micro/Nanométricos). Sevilla.

Calvo-Gallego, E, Aldaya-Cabrera, A., Brox, P, Sánchez-Solano, S, 2012a. Real-time FPGA Connected Component Labeling System. Presented at the 19th *IEEE International Conference on Electronics, Circuits and Systems* (ICECS),.

Calvo-Gallego, E., Brox, P., Sanchez-Solano, S., 2012. Un algoritmo en tiempo real para etiquetado de componentes conectados en imágenes, in: *Proceedings of the XVIII International IBERCHIP Workshop.*

Calvo-Gallego, E, Brox, P, Sanchez-Solano, S, 2013b. A Fuzzy System for Background Modeling in Video Sequences. *Springer, Lecture Notes in Artificial Intelligence* (LNAI) 184–192.

Chen Wu, Aghajan, H., 2008. Real-Time Human Pose Estimation: A Case Study in Algorithm Design for Smart Camera Networks. Proc. IEEE 96, 1715–1732.

Cho, J.U., Jin, S.H., Dai Pham, X., Jeon, J.W., Byun, J.E., Kang, H., 2006. A real-time object tracking system using a particle filter, in: Intelligent Robots and Systems, 2006 *IEEE/RSJ International Conference* On. pp. 2822–2827.

Everts, I., Sebe, N., Jones, G.A., 2007. Cooperative Object Tracking with Multiple PTZ Cameras. Presented at the *14th International Conference on Image Analysis and Processing, 2007.* ICIAP 2007, pp. 323–330.

Fan Yang, Paindavoine, M., 2003. Implementation of an rbf neural network on embedded systems: real-time face tracking and identity verification. *IEEE Trans. Neural Networks* 14, 1162–1175.

Garcés-Socarrás, L.., Sánchez-Solano, S., Brox, P., Cabreara Sarmiento, A.., 2013. Library for model-based design of image processing algorithms on FPGAs. Rev. Fac. Ing. Univ. Antioquia, nº68 3–5.

Lazaros, N., Sirakoulis, G. C., Gasteratos, A., 2008. Review of Stereo Vision Algorithms: From Software to Hardware. Int. J. Optomechatronics 2, 435–462.

Lobaton, E., Vasudevan, R., Bajcsy, R., Sastry, S., 2010. A Distributed Topological Camera Network Representation for Tracking Applications. IEEE Trans. Image Process. 19, 2516 –2529.

Microelectronics doctorate program [WWW Document], URL http://www.phdmicroelectronica.us.es/eng/?pag= general_description

Microelectronics Master [WWW Document], URL http://www.mastermicroelectronica.us.es/)

Nachtegael, M., Van der Weken, D., Kerre, E.E., Philips, W., 2007. Soft Computing in Image Processing, Studies in Fuzziness and Soft Computing,. *Springer Verlag.*

Sankaranarayanan, A.C., Veeraraghavan, A., Chellappa, R., 2008. Object Detection, Tracking and Recognition for Multiple Smart Cameras. *Proc. IEEE 96*, 1606–1624.

Schaeferling, M., Kiefer, G., 2010. Flex-SURF: A flexible architecture for FPGA-based robust feature extraction for optical tracking systems, in: Reconfigurable Computing and FPGAs (ReConFig), 2010 International Conference On. pp. 458–463.

Stillman, S., Tanawongsuwan, R., Essa, I., 1999. Tracking multiple people with multiple cameras, in: International Conference on Audio-and Video-based Biometric Person Authentication.

Svab, J., Krajník, T., Faigl, J., Preucil, L., 2009. Fpga based speeded up robust features, in: Technologies for Practical Robot Applications, 2009. TePRA 2009. IEEE International Conference On. pp. 35–41.

Wnuk, M., 2008. Remarks on Hardware Implementation of Image Processing Algorithms. Int. J. Appl. Math. Comput. Sci. 18.

Yilmaz, A., Javed, O., 2006. Object Tracking: A Survey. ACM Comput. Surv. 38.

Zhao, J., Cheung, S.-C., Nguyen, T., 2008. Optimal Camera Network Configurations for Visual Tagging. IEEE J. Sel. Top. Signal Process. 2, 464 –479.