

Smart Cameras: Enabling Technology for Proactive Intelligent CCTV

Abbas Bigdeli†, Brian C. Lovell‡ and Ting Shan†‡
†SAFE Smart Sensors Group, Safeguarding Australia Program

National ICT Australia (NICTA) Ltd.

and

‡Intelligent Real-Time Imaging and Sensing Group

School of ITEE, The University of Queensland

{Abbas.Bigdel,Brian.Lovell, Shan.Ting}@nicta.com.au

ABSTRACT: A proactive Intelligent CCTV system must ideally be capable of processing the video stream on the fly recognising events or objects of interest in real-time, hence making it possible for a timely response to imminent threats such as terrorism acts. The main bottleneck in realising a system with such capabilities is two folds. For accurate and efficient recognition, a high resolution camera is needed which in turn requires a large bandwidth for transferring the image to the CPU. The resulting data handling and memory management requirements in effect paralyses the CPU leaving no processing power to actually process the image and perform the recognition. We aim to overcome this by discarding the unused information at the sensor side before the entire image is transferred to the CPU therefore freeing up CPU time for actual processing. In this paper we present an architecture for high performance embedded systems used in computer vision applications in general and the smart camera concept in particular.

Introduction

Video surveillance systems have attracted worldwide attention since they were used to such great effect to track the movements of the four suicide bombers in the days before their attack on the London Underground in July 2005. Despite their usefulness, most current surveillance systems only provide reactive security by enabling the analysis of activities after the terrorist attack has already occurred — what is needed is proactive security to help prevent future attacks. Intelligent Closed-Circuit TV (ICCTV) systems use powerful computers to analyse the video feeds to assist human operators to detect events of interest as they occur — an example might be recognising the face of a suspected terrorist in a crowded railway station.

A proactive Intelligent CCTV must ideally be capable of processing the video stream on the fly recognising events or objects of interest in real-time, hence making it possible for a timely response to imminent threats such as terrorism acts. The main bottleneck in realising a system with such capabilities is two folds. For accurate and efficient recognition, a high resolution camera is needed which in turn requires a large bandwidth for transferring the image to the CPU. The resulting data handling and memory management requirements in effect paralyses the CPU leaving no processing power to actually process the image and perform the recognition.

The motivation for our project is to overcome the high bandwidth requirement of high resolution cameras by discarding the unused information at the sensor side before the entire image is

transferred to the CPU therefore freeing up CPU time for actual processing. In the first instance, we set to develop a general purpose computer system (embedded) for acquiring dual resolution video. Such system utilizing hardware/software solutions on a Field Programmable Gate Arrays (FPGA) platform provides a mechanism that enables both a full frame view for finding objects of interest and full resolution details of those objects.

Furthermore, the reconfigurable nature of FPGA's makes it possible for some part or even the entire processing algorithm to be done in hardware hence achieving real-time performance in many applications. At the later stages of this project, it is envisaged to develop an integrated HW/SW solution whereby the FPGA can be reconfigured on the fly as the requirement of the software changes. An example of this would be a system where the pre-processing stages such as edge detection, histogram equalization, and various transforms are performed in HW. When this system is required to recognize a face, the HW is configured to perform certain pre-processing; while for a different application the HW is configured to perform other processes. Such system constitutes a reconfigurable Smart Camera.

Smart Camera

The concept of Smart Camera has been floating for quite some (Wolf et al. 2002). According to Wikipedia, "A smart camera is an integrated machine vision system which, in addition to image capture circuitry, includes a processor, which can extract information from images without need for an external processing unit, and interface devices used to make results available to other devices". This definition is very general and there are not many Smart Cameras with full processing capability. In reality, as Andrew Wilson, the editor of Vision Systems Design argues (Andrews 2003), any imaging sensor that is housed alongside integrated circuits with some computational power, where part or all of the processing is transferred from software to the IC can be considered a smart camera system.

The types of computation performed on Smart Cameras vary between applications. In most cases, it is limited to compressing the image capture and transmitting it to the host computer. In more recent years however, a wide range of applications have emerged. Popular algorithms such as Face Recognition are good example of Smart Camera application. Fatemi et al. (2003) introduced a real-time face recognition System on a Smart Camera while Kleihost et al. (2004) took this concept further by developing a dedicated Smart Camera for Face Recognition. Ozer & Wolf (2001) have also implemented Human Activity recognition on a Smart Camera.

With growing interest in security and surveillance, Smart Cameras are finding their way in intelligent surveillance systems too. Matsushita et al. (2003) used Smart Camera for scene capturing and identity recognition. A more complex application of Smart Cameras includes Traffic Surveillance (Bramberger et al. 2004). One of the key features of most Smart Camera Systems is their ability to process areas of the image that contain information of interest. This is usually referred to as Region of Interest (ROI) processing.

Region of Interest Processing

ROI processing has been an area of research for several groups world-wide (Bradley 2003). Much of this research is for server-side ROI extraction prior to image coding for enhanced

quality at low bit rates. Unfortunately the improvements are frequently marginal at best due to the unnatural artefacts created by spatial variations in resolution (Bradley 2003). A much more sensible approach would be client-side specification of the true region of interest required for the application at hand. In the case of a single human observer, say, this could be performed by instantaneously monitoring the angle of gaze and enhancing image resolution in the centre of the field of vision. If done convincingly, the viewer would have the impression that the entire field of view was high resolution — yet bandwidth could be dramatically reduced (factor of 100 estimated based on full visual field).

Recently, researchers have tried to do this in real-time, but have used large and power-hungry processing boards to perform the ROI processing. Ever-increasing transistor densities and integration of camera sensors on silicon now make it possible to embed a whole image processing system upon a single chip. In the work reported by Wolf et al. at Princeton University an embedded system was designed that did ROI type processing for a single application (that of gesture recognition) and implemented a prototype using Trimedia VLIW processors and a PC. Nicholescu & Medioni (200) merged the images of five radially-mounted CMOS sensors into a single wide-angle “Region of Awareness” (portion of the environment being monitored). The high-resolution ROI within this region could then be panned, tilted, and zoomed to examine objects of interest such as moving people. The 5 camera panoramic merging and calibration was performed in software resulting in a serious processing bottleneck, yet the concept showed great promise in terms of cost and monitoring performance compared to competing technologies such as mechanical pan-tilt-zoom and wide angle lens cameras. To circumvent this processing bottleneck, practical imaging systems need to perform smart data reduction as close to the imaging sensor as possible, essentially putting the whole system on a single chip.

Dual Resolution Camera

Conventional cameras provide the full frame at full resolution, and the detailed data for the object of interest is extracted within the computer application. This means that much unnecessary data is transferred which wastes bandwidth and limits the frame rate of the video streams.

The dual-resolution camera would reduce this problem by providing two video streams: a low-resolution, full frame view for finding the object of interest, and a full-resolution sub-frame or ROI to provide a detailed image of the object. The location of the ROI within the camera’s field of view changes while the camera is streaming data, allowing for example for moving objects to be tracked. Another example is using the field of view to locate a face in a scene and then assigning a ROI to the face hence transferring the face window at full resolution. Figures 1 clearly illustrates how a dual resolution camera may be used for number plate recognition or face recognition application.

A Smart Camera once fitted with the Dual-Resolution capability, may further process the ROI so even smaller more meaningful information is sent to the host computer. In Figure 2, the ROI where number plate is of interest, rather than transferring the actual colour image containing the number plate, a Smart Camera with appropriate edge detection filter can further process the image and only transmit either the recognised characters or the edge outline of the number plate.

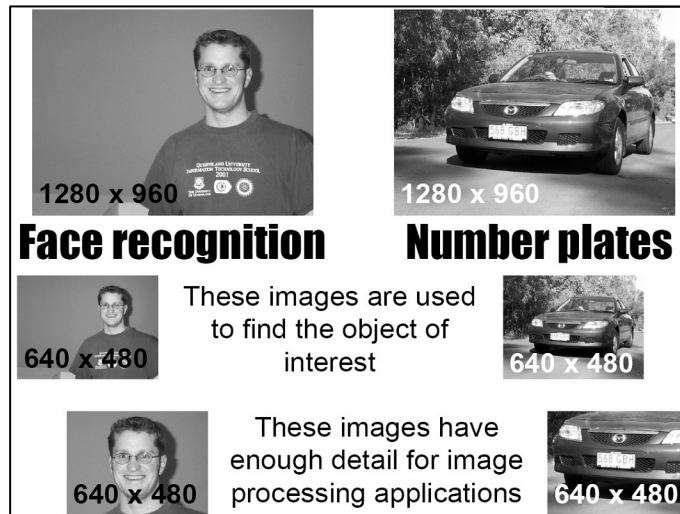


Figure 1. Dual-Resolution Camera Examples



Figure 2. Dual-Resolution Camera with embedded edge detection

Ideally a dual-resolution camera should be implemented on a reconfigurable system-on-chip. Using a camera sensor directly interfaced to an FPGA will enable variable and multiple ROI extraction, along with the necessary ROI application processing and subsequent communication to the client. The Image transfer and control communication between the Camera and the host PC is shown in Figure 3.

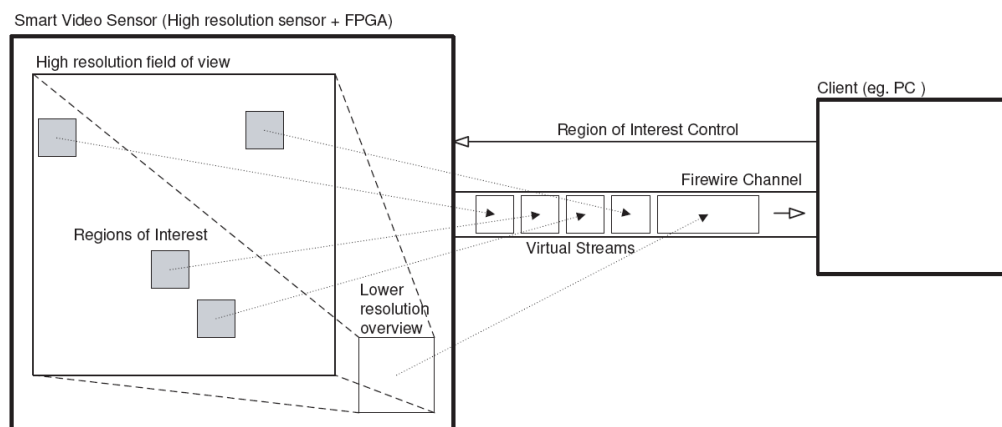


Figure 3. Image Transfer and Control of a Dual-Resolution Camera

Conclusion

Smart Cameras are slowly being introduced in the emerging surveillance systems. They usually perform a set of low-level image processing operations on the input frames at the sensor end. This is done usually to improve the video compression. The majority of the video processing and analysis in the existing surveillance systems is executed at a central host using standard workstation racks. Up-and-coming embedded computing technology is being used to implement Smart Cameras with greater processing capabilities. Such capabilities include face and number plate recognition, human activity recognition etc.

The adoption of Smart cameras is resulting in a paradigm shift from a central to a distributed control surveillance system. The main motivation for this shift is increasing the surveillance system's functionality, availability, and autonomy.

Acknowledgement

This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet. National ICT Australia is funded by the Australian Government's Backing Australia's Ability initiative and the Queensland Government, in part through the Australian Research Council.

AUTHOR'S BIOGRAPHY:

Abbas Bigdeli: Researcher of the SAFE Sensors group in National ICT Australia. His main research interests are is high performance real-time embedded systems for computer vision applications.

Brian Lovell: Research Leader of the SAFE Sensors group in National ICT Australia and Professor in the School of ITEE, The University of Queensland. His main interest is the analysis of video streams to recognize human activities.

Ting Shan: PhD student of University of Queensland, his research interest is in face detection, multi-view face recognition and real-time face recognition system.

REFERENCES

- Bradley, A. P. 2003. Can Region of Interest Coding Improve Overall Perceived Image Quality? *In proceedings of WDIC2003*, Feb 2003: 41-44.
- Bramberger, M., Brunner, J., Rinner, B., Schwabach, H. 2004. Real-Time Video Analysis on an Embedded Smart Camera for Traffic Surveillance. *In proceedings of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'04)*, May 2004: 174- 181.
- Fatemi, H., Kleihorst, R., Corporall, H., & Jonker, P. 2003. Real-time Face Recognition on Smart Camera. *Proceedings of Acivs 2003 (Advance Concepts for Intelligent Vision Systems)*, Belgium, September 2003: 222-227.
- Kleihorst, R., Reuvers, M., Krose, B. & Broers, H. 2004. A Smart Camera For Face Recognition. *Proceedings of International Conference on Image Processing (ICIP 2004)*: 2849-2852.
- Matsushita, N., Hihara, D., Ushiro, T., Yoshimura, S., Remkimito, J. & Yamamoto, Y. 2003. ID CAM: A Smart Camera for Scene Capturing and ID Recognition. *In proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)*. October 2003: 227- 236.
- Nicolescu, N. & Medioni, G. G. 2000. Electronic Pan- Tilt-Zoom: A Solution for Intelligent Room Systems, *In proceedings of IEEE International Conference of Multimedia and Expo (ICME 00)*, IEEE CS Press, Los Alamitos Calif., 2000: 1581-1584.
- Ozer, B. & Wolf, W. 2001. A smart camera for real-time human activity recognition, *In Proceedings, SIPS-01*, IEEE: 217-225.
- Wolf, W., Ozer, B. & Lv, T. 2002. Smart Cameras as Embedded Systems. *IEEE Computer*, 35(9):48-53.
- Wilson, A. 2003. Smart Cameras embed processor power, *Vision Systems Design*, September 2003: 95-99.