

Smart Cameras Enabling Automated Face Recognition in the Crowd for Intelligent Surveillance System

Y. M. Mustafah

National ICT Australia Ltd. (NICTA), School of ITEE, The University of Queensland

A. Bigdeli

National ICT Australia Ltd. (NICTA)

A. W. Azman

National ICT Australia Ltd. (NICTA), School of ITEE, The University of Queensland

B. C. Lovell

National ICT Australia Ltd. (NICTA), School of ITEE, The University of Queensland

ABSTRACT: Smart Cameras are rapidly finding their way into Intelligent Surveillance Systems. Recognizing faces in the crowd in real-time is one of the key features that will significantly enhance Intelligent Surveillance Systems. The main challenge is the fact that the enormous volumes of data generated by high-resolution sensors can make it computationally impossible to process on mainstream processors. In this paper we report on the prototyping development of a smart camera for automated face recognition using very high resolution sensors. In the proposed technique, the smart camera extracts all the faces from the full-resolution frame and only sends the image information from these face areas to the main processing unit — vastly reducing data rates. Face recognition software that runs on the main processing unit will then perform the required pattern recognition.

BIOGRAPHY: Y. M. Mustafah is a PhD student of The University of Queensland. His research interest is in face detection and recognition and embedded systems for computer vision. A. Bigdeli is a researcher of SAFE Sensors group in National ICT Australia. His main research is in high performance real-time embedded systems for computer vision applications. A. W. Azman is also a PhD student of The University of Queensland. Her research interest is in the software-hardware partitioning on embedded system and embedded systems for computer visions. B. C. Lovell is a Professor in School of ITEE, The University of Queensland. He is also the research leader of the SAFE Sensors group in National ICT Australia. His main interest is the analysis of video streams for human activities recognition.

Introduction

Video surveillance is becoming more and more essential nowadays as society relies on video surveillance to improve security and safety. For security, such systems are usually installed in areas where crime can occur such as banks and car parks. For safety, the systems are installed in areas where there is the possibility of accidents such as on roads or motorways and at construction sites.

Currently, surveillance video data is used predominantly as a forensic tool, thus losing its primary benefit as a proactive real-time alerting system. For example, the surveillance systems in London managed to track the movements of the four suicide bombers in the days prior to their attack on the London Underground in July 2005, but the footage was only reviewed after the attack had occurred. What is needed is continuous monitoring of all surveillance video to alert security personnel or to sound alarms while there is still time to prevent or mitigate the injuries or damage to property. The fundamental problem is that while mounting more video cameras is relatively cheap, finding and funding human resources to observe the video feeds is very expensive. Moreover, human operators for surveillance monitoring rapidly become tired and inattentive due to the dull and boring nature of the activity. There is a strong case for automated surveillance systems where powerful computers monitor the video feeds — even if they only help to keep human operators vigilant by sending relevant alarms.

Smart cameras can improve video surveillance systems by making autonomous video surveillance possible. Instead of using surveillance cameras to solve a crime after the event, a smart camera could recognize suspicious activity or individual faces and give out an alert so that an unwanted event could be prevented or the damage lessened. From another perspective, smart cameras reduce the need for human operators to continually monitor all the video feeds just to detect the activities of interest, thus reducing operating costs and increasing effectiveness.

Smart Cameras

Smart cameras are becoming increasingly popular with advances in both machine vision and semiconductor technology. In the past, a typical camera was only able to capture images. Now, with the smart camera concept, a camera will have the ability to generate specific information from the images that it has captured. So far there does not seem to be a well-established definition of what exactly a smart camera is. In this paper, we define a smart camera as a vision system which can extract information from images and generate specific information for other devices such as a PC or a surveillance system without the need for an external processing unit.

Figure 1 shows a basic structure of a smart camera. Just like a typical digital camera, a smart camera captures an image using an image sensor, stores the captured image in the memory, and transfers it to another device or user using a communication interface. However, unlike the simple processor in a typical digital camera, the processor in a smart camera will not only control the camera functionalities, but it is also able to analyse the captured images to obtain extra information.

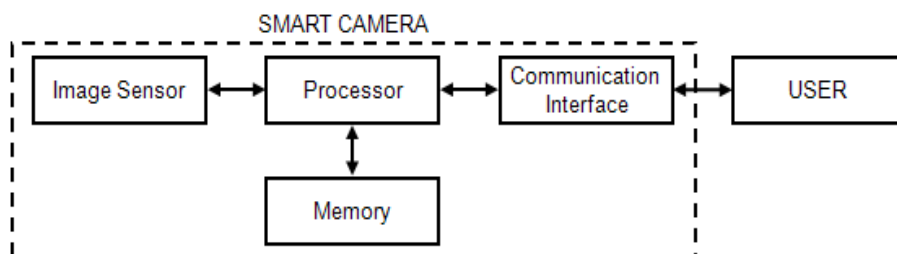


Figure 1: Basic Smart Camera Architecture.

There are many smart camera products available in the market today from a variety of manufacturers such as Tattile, Cognex, Matrix Vision, Sony, Philips, EyeSpector, PPT Vision, and Vision Components. However, this is still a very active area of research because of the wide range of capabilities of smart cameras that could be improved. One of the most popular works on the smart camera was by Wolf et al. (2006) where they introduced a system that can build a complete model of the torso and recognize various gestures made by a person. The work started with research on a human activity recognition algorithm and soon evolved to the implementation of the software algorithm onto hardware, including Hi8 cameras and Trimedia video capture boards. Another well-known research project was by Bramberger et al. (2006). They built a prototype camera called SmartCam which is a fully embedded smart camera system targeted for various surveillance applications such as traffic control.

Improving Smart Camera Design

We propose a smart camera system that can be used as an aid for face recognition in crowd surveillance. The camera utilizes a high resolution CMOS image capture device and an FPGA based processor for Region of Interest (ROI) extraction. The proposed system architecture is shown in Figure 2. The system has an internal processor to perform face detection to extract faces from the captured images in real-time. The main motivation to extract faces inside the camera is to conserve as much bandwidth as possible and to save processing time and memory on the client processor which performs the face recognition task.

Note that even in the dense crowd of Figure 3, the faces suitable for recognition only represent a very small proportion of the image area. In many scenes, faces would represent less than 1% of the image. Thus the smart camera would not overload the client processor by transmitting huge amounts of high-resolution image data which is destined to be discarded immediately after face detection. Such massive data reduction at source by up to two orders of magnitude is an immediate and significant benefit of this approach.

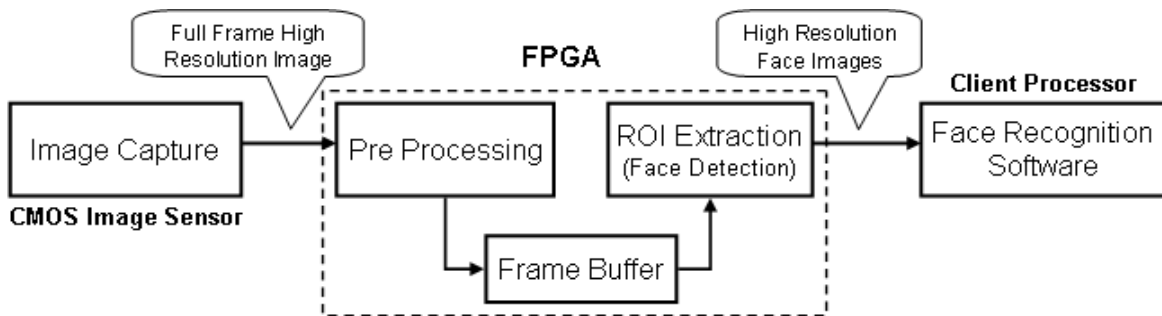


Figure 2: Proposed System Architecture.

High Resolution Image Sensor

While there are many smart camera products already available in the market today, most of them use a VGA resolution image sensor (640 x 480 pixels) and some use lower resolutions. Compared to the existing image sensor technology, VGA can be considered as low resolution and is quite unsuitable for many video surveillance applications especially in crowd surveillance. Crowd surveillance usually surveils a wide area with many of objects of interest in view, thus requiring a high resolution camera. High resolution images provide much more detailed information regarding objects in view. For example, in applications such as face recognition,

higher resolution images will help improve the recognition rate — indeed a very high resolution camera could even read nametags and other insignia. Figure 3 shows an example of a region of interest (ROI) extracted from a scene image of a crowd of people. In this simple experiment, the image window containing the face is extracted manually from the scene images (a) with 5 different resolutions. The face (b) extracted from a 7 MP (MegaPixel) high resolution image is much more recognizable than (f) extracted from the lower resolution (VGA) scene. The extracted image windows were tested for suitability for automatic face detection using a Viola-Jones face detection module (P. Viola and M. Jones. 2001) implemented in OpenCV library. The images (b), (c), (d) and (e) extracted from 7, 5, 3, and 1 MP images were suitable for face detection. However, the face cannot be correctly detected in the image (f) because the image does not have enough detail for the face detection module to work correctly.

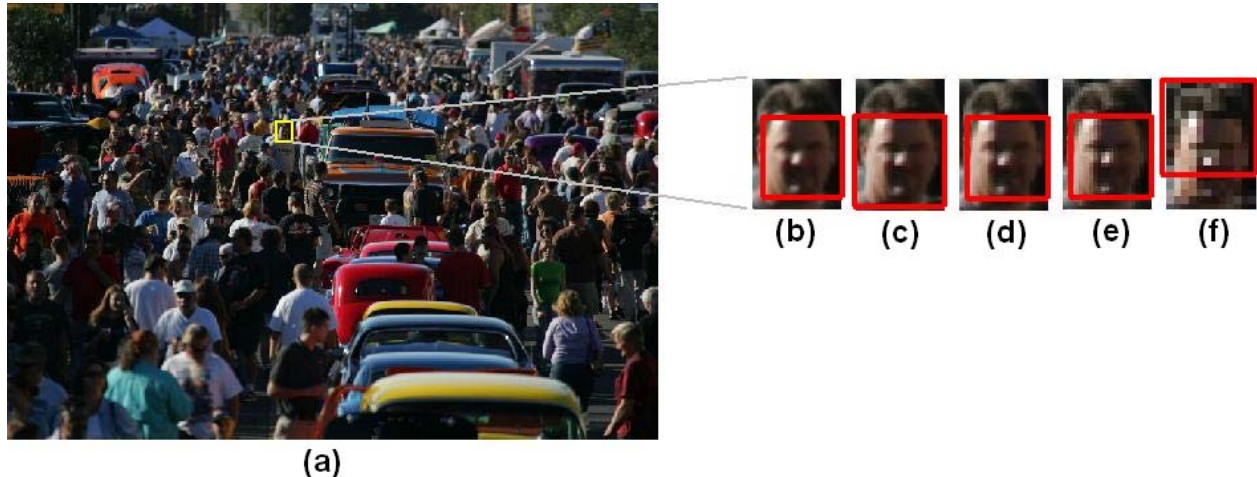


Figure 3: Overall Scene (a), ROI extracted from scene with resolution of 7MP (b), 5 MP(c), 3MP (d), 1MP (e) and VGA (f).

CMOS image sensor offers high resolution and low noise output. Due to the low power and high speed of CMOS, it is expected in the future that CMOS based image sensors will outperform CCD based image sensors (D. Litwiller. 2001). There are many CMOS image sensors in the market. Table 1 shows the highest resolution sensor product from three leading CMOS image sensor manufacturers; OmniVision, Micron and Kodak. It is noticeable that the frame rate is inversely proportional to the resolution of the camera. For wide angle surveillance, since no rapid movements of object of interest are expected, 5 high resolution frames per second can be considered as acceptable baseline performance for the prototype.

Table 1: High Resolution CMOS Image Sensors

Manufacturer	CMOS Sensor	Resolution (pixel)	Frame Rate at Full Resolution (fps)
OmniVision	OV5620	2608 x 1952	7.5
Micron	MT9E001	3264 x 2448	10
Micron	MT9P001	2592 x 1944	14
Kodak	KAC-5000	2592 x 1944	6

High Bandwidth Communication Interface

High resolution image sensors require a high data transfer rate. Although a smart camera could preprocess the captured images before they are sent to the external devices, it is sometimes necessary for the camera to output the RAW captured images. For a high resolution camera, a high bandwidth communication interface would be required. This is because, for example, a 5MP sensor working at the rate of 10 fps would require 50MP of data to be transferred every second if no compression is performed on the image. One pixel might represent several bits of data depending on the color depth of the image. Therefore without compression, a camera with a 5MP high-color (24-bit color depth) frame image working at 10 fps, would need a communication link with sustained transfer rate of 1200 Mega bits per second — no garden variety PC could keep up with this.

However, currently, there are five commonly used high bandwidth video interface standards available: FireWire 400 or IEEE 1394a, FireWire 800 or IEEE 1394b, USB2, Gigabit Ethernet or GigE, and Camera Link. Table 2 shows the general specification on the five interfaces. USB 2 and FireWire 400 can be considered as unsuitable in terms of data transfer speed if we compare them with the current resolution of CMOS image sensor technology. While Camera Link is suitable for very fast data transfer, it only supports one-to-one device connection. This means a network of cameras could not be supported by this interface. GigE and FireWire 800 interfaces can be considered as the most suitable interfaces for the purposed high resolution surveillance as they both have a considerable data transfer speed and allow for the networking of the cameras. For our first prototype, we decided to use FireWire 800 interface, because currently it is a more established interface. We will try to incorporate GigE interface in our future prototype once this interface become more mature. With the introduction of a new GigE Vision camera interface, it is expected GigE will become the dominant machine vision interface in the near future.

Table 2: Video Interface Standards

Interface	Data Transfer Rate (Mbps)	Max Cable Length (meter)	Max number of Devices
FireWire 400 (1394a)	400	4.5	63
FireWire 800 (1394b)	800	100	63
USB 2	480	5	127
GigE	1000	100	no limit
Camera Link	3600	10	1

Reconfigurable Platform for Hardware and Software Processors

Acquiring the appropriate target hardware for a smart camera processor is an important issue. While Application Specific Integrated Circuit (ASIC) provides a high performance and power efficient platform, it suffers from lack of flexibility and can be very expensive due to the high non-recurring engineering (NRE) cost. Digital Signal Processor (DSP) on the other hand has only a single flow of control which could pose problems in meeting real-time constraints. The general-purpose processor (GPP) also faces problems in meeting real-time constraint due to poor execution time predictability. Garcia et al. (2006) suggested that reconfigurable hardware as the

best option and is quite cost effective for an embedded system. Presently, Field Programmable Gate Arrays (FPGA) are one of the most widely used and competitive reconfigurable hardware platforms in the market.

One of the key aspects of FPGA is it has large number of arrays of parallel logic and registers which enable designers to produce effective parallel architectures. Parallel processing is an important feature especially for embedded systems that require high-level computation in real-time — for example, face detection on a smart camera processor. Parallel processing allows information to be transferred effectively and obtains end results faster since processing tasks are segregated to be carried out concurrently. At the same time, parallel processing reduces power consumption considerably, especially in processes which involve back-to-back memory access. Additionally, FPGA allows incorporation of a microprocessor on the same chip. For our smart camera prototype, the Spartan-3 FPGA was chosen as the main processing. We believe that the Spartan-3 platform imposes an interesting challenge where optimum hardware resources will be utilized in every aspect of the design.

Robust Face Recognition System

The uncontrolled environment of crowd surveillance makes a robust face recognition system a necessity. Ideally, a robust face recognition system would be able to recognize faces regardless of the face's expression, angle, features and lighting conditions. A face recognition system consists of face detection and a face classification part. In order for the system to recognize a particular face, the face must first be detected, and then extracted from the captured scene image. The face is then normalized and forwarded to the face classification processor where it could be recognized by comparing it to the faces stored in the database.

The face recognition to be implemented on the system is as proposed by Shan et al. (2006). Their system is comprised of three major components: 1) a Viola-Jones face detection module (Viola and Jones. 2001) based on cascaded simple binary features to rapidly detect and locate multiple faces, 2) a normalization module based on the eye locations, and finally 3) Adaptive Principal Component Analysis to recognize the faces. As stated earlier, the face detection part (1) will be implemented on the FPGA platform of the camera while the rest of the module will be implemented on the client PC.

The face detection and face recognition processes usually require considerable computing power and could require significant time when running on a standard PC. In a hardware implementation however, processes can be decomposed and run in parallel so that less time will be taken to execute the processes. FPGA's provide a flexible and suitable reconfigurable platform for applying suitable architecture of the processor. If sufficient parallelism is applied, it is possible for the overall process to run in real-time.

NICTA Smart Camera Prototype

Our smart camera platform was designed based on the principles outlined in the previous section. Table 3 summarizes the basic specifications of our prototype while Figure 4 shows our smart camera prototype.

Table 3: Specification of Smart Camera

Parameter	Value
Sensor Type	CMOS
Resolution	2592 x 1944
Processing Platform	Spartan-3 FPGA
Comm. Interface	Firewire800
Dimension	90 x 90 x 150 mm ³

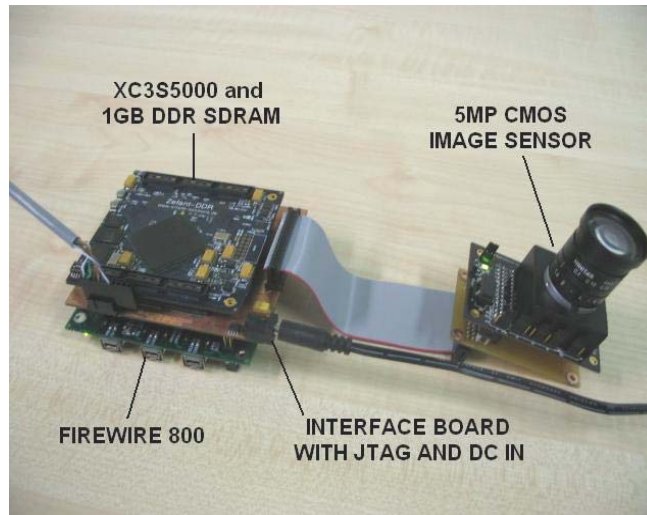


Figure 4: Smart Camera Prototype.

Conclusion and Future Work

Smart Cameras are slowly being introduced in emerging surveillance systems. They usually perform a set of low-level image processing operations on the input frames at the sensor end. This paper reported on our prototype development of a smart camera for automated face recognition using high resolution (5MP) sensors. In the proposed technique, the smart camera extracts all the faces from the full-resolution frame and only sends the pixel information from these face areas to the main processing unit. Face recognition software that runs on the main processing unit will then perform the required pattern recognition algorithm. The main challenge in this project is to build a stand-alone and low power smart camera system that integrates real-time face detection for crowd surveillance. Our future work would involve implementing a robust face detection algorithm on the camera.

Acknowledgements

NICTA is funded by the Australian Government's department of Communications, Information Technology, and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Research Centre of Excellence programs, and the Queensland State Government.

References

- D. Litwiller. 2001. CCD vs. CMOS: Facts and Fiction. *Photonics Spectra*.
- M. Bramberger, A. Doblander, A. Maier, B. Rinner, and H. Schwabach. 2006. Distributed embedded smart cameras for surveillance applications. *Computer*. 39: 68-75.
- P. Garcia, K. Compton, M. Schulte, E. Blem, and W. Fu. 2006. An Overview of Reconfigurable Hardware in Embedded Systems. *EURASIP Journal on Embedded Systems*. 1-19.
- P. Viola and M. Jones. 2001. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*. 511-518.
- T. Shan, B. C. Lovell, S. Chen, and A. Bigdeli. 2006. Reliable Face Recognition for Intelligent CCTV. *2006 RNSA Security Technology Conference*. Canberra. 356-364.
- W. Wolf, B. Ozer, and T. Lv. 2002. Smart cameras as embedded systems. *Computer*. 35: 48-53.