ARCHITECTURE AND APPLICATIONS OF WIRELESS SMART CAMERAS (NETWORKS)

Richard Kleihorst, Ben Schueler, Alexander Danilin

NXP Semiconductors Research, High Tech Campus, Eindhoven, The Netherlands

ABSTRACT

A network of (wireless smart) cameras can analyse the scene from different views. Wireless smart cameras challenge the hardware for low-power consumption and high imaging performance. In this paper we introduce a wireless smart camera based on an SIMD video-analysis processor and an 8051 microcontroller as a local host. Wireless communication is through the IEEE802.15.4 standard. The camera constructed in this paper is to enable application research into distributed smart camera systems.

Index Terms— Real-time vision, smart camera, parallel processing

1. INTRODUCTION

Imaging plays an important role in sensing devices for ambient intelligence [1]. Computer vision can for instance be used for recognising persons and objects and recognising behaviour such as illness and rioting. Having wireless cameras in a network opens the way for distributed scene analysis. "More eyes see more than one" and a camera system that can observe a scene from multiple directions is able to overcome occlusion problems and describes objects in their true 3D appearance. These approaches are a recent field of research [2, 3, 4].



Fig. 1. The new wireless smart camera mote contains 2 VGA colour sensors and a high-performance vision system.

There are several key aspects relevant for wireless vision. Among those are programmability, power consumption, response time and the demands on the image processing performance. Because of power consumption reasons, it makes sense to do at least the event detection (image processing) on the ambient sensing device itself and then start sending data only instead of raw video streams. The systems that do this are cameras with embedded processing, known as "smart cameras". A wireless smart camera can operate in intelligent networks sharing the view of the environment with other cameras and sensor nodes.

Lately, there have been several smart cameras made that are suitable to be connected to a wireless network for event broadcasting. Although there is an increasing amount of commercial smart IP cameras for industrial or surveillance applications with dedicated processing on board, we will focus here on lower-cost cameras that are completely programmable and for development and research applications. Among the most visible ones are the CMUCams developed at Carnegie Mellon University [5] and the Cyclops camera developed by Agilent and UCLA [6, 7]. While the latter one has a wireless connection, the CMUcams can easily be made wireless. As a research prototype, Stanford university is developing a wireless smart camera with 3 image sensors for low-power imaging [8]. The core programmable blocks for these cameras are CPLDs, micro-controllers (CMUcam and Cyclops) and an ARM7 processor for the Stanford camera. We want to argue in this paper that the above used processor choices are able to do life video analysis in an efficient manner.

We propose a smart camera that has a fully programmable Single-Instruction Multiple-Data (SIMD) processor dedicated for video-analysis (see Figure 1). The processor has been proven to give ultra-high pixel processing performance for milli-Watts through its massively parallel SIMD architecture. Previous smart cameras of the authors were more in the industrial domain were the SIMD processor was linearly coupled to a TriMedia host processor [9]. In the new camera an 8051 microcontroller is used as host and the two processors are coupled using a shared (dual port) memory in order to have unsynchronised tasks on the processors. A feedback mechanism allows multiple passes of the SIMD processor on image data, allowing lens-distortion correction and geometrical transformations.

The remainder of the paper is organized as follows: Section 2 demonstrates that local processing reduces energy consumption. Section 3 introduces hardware kernels for image processing. Sections 4 describes the wireless smart camera. While Section 5 discusses present projects of the camera.

2. WHY "SMART" WIRELESS CAMERAS?

Broadcasting the raw video data to a PC takes in the order of 400mWatts from the camera budget for a digital 15fps gray level VGA wireless link. Actually, for short distance transmission most of the broadcast energy is dissipated in the DA convertor in the transmitter. Current technology of DA converters is able to reach very close to the lowest power consumption limit, dictated by thermal noise. For DSP on silicon however, power consumption continues to reduce because of techniques like technology- and voltage scaling, lazy computation and the use of low-energy architectures. This reduction of energy consumption can go magnitudes further before it reaches the intrinsic minimum of silicon [1]. The field of research into giving more system performance per Watt is finding solutions through the use of strong data-level parallelism and advances in integration techniques. Real-time video processing on low-cost and low-power programmable platforms is now becoming possible [10, 11]. This draws us to the conclusion that it is a better proposition to invest more into computing at the camera node itself and sending only event detections to the central host and/or to the other cameras in the connected environment.

3. HARDWARE KERNELS FOR EFFICIENT IMAGE PROCESSING

The algorithms in the application areas of smart cameras can be grouped into 3 levels: *low-level, intermediate-level* and *high-level* tasks. The *low- or early- image processing level* is associated with typical kernel operations like convolutions and data dependent operations using a limited neighbourhood of the current pixels. The processors exploiting the inherent parallelism in this level have an SIMD architecture, where the same instruction is issued on all data items in parallel [12]. The parallel architecture reduces the number of memory accesses, clock speed, and instruction decoding, thereby enabling higher arithmetic performance at lower power consumption [11].

Next to object and pixel operations, also image based operations such as lens distortion correction, geometrical transforms and image pyramids are important kernels for vision systems. Traditionally these tasks were run on general purpose processors, often using look-up tables. Recently it was shown how to do these operations efficiently using SIMD processors connected to scratch-pad memories [13].

In the *high*- and *intermediate-level part* of image processing, decisions are made and forwarded to the user. General purpose processors are ideal for these tasks because they offer the flexibility to implement complex software tasks and are often capable of running an operating system and doing networking applications.

4. CAMERA HARDWARE PLATFORM

With earlier stated considerations in mind we developed a wireless smart camera system that can operate stand-alone or in a network of cameras. This camera is depicted in Figure 1, it consists of basically four components, one or two VGA color image sensors, an SIMD processor for low-level image processing, a general purpose processor for intermediate and high-level processing and control and a communication module. Both processors are coupled using a dual port RAM that enables them to work in a shared workspace on their own processing pace (see Figures 2 and 4).



Fig. 2. Complete architecture of the wireless camera showing all processing and hardware blocks

The IC3D, a member of the NXP semiconductors' Xetal family of SIMD processors, shows 5 specific internal blocks, (see Figure 3). Two of the blocks function as video input and output processors respectively. They are capable of streaming in and out 3 digital video signals to the internal memory. The heart of the chip is formed by the Linear Processor Array (LPA) with 320 RISC processors with data paths of 10-bits wide. Each of these processors has simultaneous read and write access within one clock-cycle to memory positions in the parallel memory. Both the memory address and the instruction of the processors are shared in SIMD sense. All processors can also read the memory data of their left and right neighbors directly for filter kernel operations. The RISC processors have instructions ranging from arithmetic and singlecycle multiply-accumulate to compound instructions. In addition to these, there are conditional guarding instructions, enabling data-dependent operations.

The line-memory block stores 64 image lines. By automatic interleaving mechanisms lines of images with different sizes and all their colour components can each be processed as



Fig. 3. Architecture of the "IC3D" which is a member of the "Xetal" family of SIMD chips

single vectors. The GCP (Global Control Processor) is a processor dedicated to control the IC3D and to do some global DSP operations on the data. It takes care of video synchronization, program flow and also communicates with the LPA and the outside world. The peak pixel performance of IC3D is around 50GOPS. It can do real-time video processing simultaneously on all colour components of the 2 VGA sensors and an additional feedback channel (see Figure 4).

Despite its high pixel-performance, the IC3D is an inherently low-power processor as not only instruction decoding is shared between all 320 processors, but also memory access addresses ultra-wide memory words containing complete video lines. For typical applications, such as feature finding or face detection, the power consumption is well below 100mWatt.

The Dual Port (DP) RAM functions as an asynchronous connection between both processor cores. While the IC3D processor works on streaming data, processing the frames at sensor speed, the 8051 host-processor (discussed later) will not be working on streaming speed. Moreover, the high-level processing task on the 8051 processor is a non-constant time program that takes shorter or longer depending on the number of objects of interest in the scene.

With that in mind, the IC3D writes information from the video, such as feature points or coordinates of objects, or even (parts of) images in the DPRAM. The 8051 can than leisurely read and analyze that information and make decisions of the position, scale or movement direction of objects from the scene.

The memory uses semaphore techniques to avoid corruption of data if both processors try to write to the same address at the same time. Also, the memory has banks that can be allocated to a specific process.

The size of the memory is now 128K words of 8bits, divided into two banks of 64K words each. If the system wants to store in image format, 2 images of up to 256x256 pixels are directly storable. Applications where image format data is stored are for instance dynamic background subtraction and motion estimation. For applications like lens distortion correction, camera calibration, image pyramids or geometrical transforms, the IC3D computes the address transformation using the efficiency and performance of the SIMD engine. The architecture enables this by connecting the DPRAM addresses to the data outputs of the IC3D as shown in Figure 4. Although the IC3D can work with images from 320 to 1280 pixels wide, its output pins that are now used for memory addressing are only 8 bits wide hence the limited size of image that can be stored. Earlier work used a direct connection between the SIMD and the general purpose processor, but this caused inefficient programs on the latter processor to keep up with the SIMD and did not allow dynamic image transformations and feedback loops involving the SIMD [9].



Fig. 4. The connections between the two processors and the DPRAM

To save components and to keep the power consumption low a top-of-the-range 8051 is used. This device has all necessary components inside to make a small, yet complete system, it has a large number of usable I/O pins to control the camera and its surroundings. The 8051 has a 16bit-wide external address bus for memory which fits nicely with the dual port memory connected to the IC3D. To indicate special data transfers between the IC3D and 8051, an interrupt line on the 8051 is used that can be triggered by IC3D. The internal EEP-ROM is used to store parameters and instruction code for the IC3D. Communication to the outside world is done via the UART.

The Aquis Grain ZigBee module is the transceiver part of the wireless camera. It was made by Philips Research around ChipCon's CC2420 SOC, see the top PCB in Figure 1. The radio system implements a MAC layer in IEEE 802.15.4 spec. The software radio system is programmed on an additional 8051 processor and can be modified for special purpose applications. The 802.15.4 standard offers wireless communication upto about 5 meters distance. In the communication network the device that starts up first acts as coordinator. The peer-to-peer structure offers direct camera to camera communication [15]. It is also quite robust as cameras can be switched on or off (even the coordinator) and the network will remain stable and take action automatically for the changes. The communication module is attached to the camera as a wireless UART port of limited capacity. The maximum data-rate of around 10kB/second will only leave room to communicate about details or events in the scene. On a non-real-time rate images or parts of images can be send. The network is quite capable however to send for instance face images of people who are present in the scene to each-other or a host processor. Although the low bit-rate seems to pose problems for present-day approaches, it also solves a number of problems and creates new challenges. The low bit-rate for instance enables a low-power solution as discussed before and also from a legal and privacy point of view: the cameras will technically not be able to stream live video data, which could make the acceptance rate of cameras in home environments higher.

The complete camera can be programmed wirelessly or remotely due to "in-system" programmability capabilities of the 8051. An external I²C EEPROM can store 16 application programs which can be used for content switching approaches.

The software for the wireless camera consists of 3 parts that are almost independently developed. Programs for the IC3D are written in a C^{++} language with image lines as implicit parallel data-types. The programs on the 8051 host-processor are dedicated to keep track of data and events over time. It performs the host function (running the operating system) and can decide to transmit events to a host system.

5. CURRENT PROJECTS

Within the Philips and NXP semiconductors Labs there are several projects that implement applications on the presented wireless smart camera which was named "WiCa". Among the projects that were proven on this platform are the development of layered communication protocols that permit the processors on different cameras to contact each-other directly [15]. These techniques rely on parts of the ZigBee stack and will be used in future distributed imaging applications. On the imaging side, we mapped a face detection method that is based on detection of vertical and horizontal patterns in smooth gradient regions [16]. By combining the detection results from the two image sensors and incorporating information about the distance to the cameras, robust face and eye detections are obtained that are good enough for commercial applications. The power consumption of IC3D for this stereo face detection algorithm is 40mWatts. Also, first attempts for gesture recognition based on hand detection were demonstrated. While the technique is able to recognize finger positions (and report this through the network), we are suffering with hand detection if the background is too busy. The IC3D power consumption for hand detection and finger position recognition is now around 15mWatts. We hope that the technique becomes more stable when more cameras around the scene are involved in the detection.

All projects run in real-time (VGA video at 30 fps) on the described wireless smart camera system. Future algorithmic projects will focus on collaboration between cameras for distributed scene analysis. Future hardware projects will involve migrating to a combination of Xetal-II and TriMedia with a larger DPRAM, enabling more low-level and high-level performance [10].

6. REFERENCES

- Richard Kleihorst, Ben Schueler, Anteneh Abbo, and Vishal Choudhary, "Design challenges for power consumption in mobile smart cameras," in *COGIS2006*, Paris, France, Mar. 2006.
- [2] S. Velipasalar and W. Wolf, "Multiple object tracking and occlusion handling by information exchange between uncalibrated cameras," in *Int. Conf. Image Proc. (ICIP'05)*, Genova, Italy, Sep. 11–14, 2005.
- [3] J. Mallet and V. M. Bove, "Eye society," in *ICME2003*, Baltimore, MD, USA, July 2003.
- [4] Arezou Keshvarz, Ali Maleki Tabar, and Hamid Aghajan, "Distributed vision-based reasoning for smart home care," in *DSC'06*, Boulder, USA, 2006.
- [5] "The CMUcam vision sensors, website," http://www.cs.cmu.edu/ cmucam/.
- [6] "Welcome to the Cyclops Camera!, website," http://www.cyclopscamera.org/.
- [7] Mohammad Rahimi, Rick Baer, Obimdinachi Iroez, Juan Garcia, Jay Warrior, Deborah Estrin, and Mani Srivastava, "Cyclops: In situ image sensing and interpretation," *Third ACM Conference on Embedded Networked Sensor Systems*, November 2-4 2005.
- [8] Stephan Hengstler and Hamid Aghajan, "A smart camera mote architecture for distributed intelligent surveillance," in DSC'06, Boulder, USA, 2006.
- [9] R. Kleihorst et al., "An SIMD smart camera architecture for real-time face recognition," in *Abstracts of the SAFE & ProR-ISC/IEEE Workshops on Semiconductors, Circuits and Systems* and Signal Processing, Veldhoven, The Netherlands, Nov 26– 27, 2003.
- [10] A. Abbo, R. Kleihorst, V. Choudhary, L. Sevat, P. Wielage, S. Mouy, and M. Heijligers, "Xetal-II: A 107GOPS, 600mw massively-parallel processor for video scene analysis," in *ISSCC2007 Digest of technical papers*, San Fransisco, Ca, USA, 2007.
- [11] A. Abbo, R. Kleihorst, V. Choudhary, and L. Sevat, "Power consumption of performance-scaled simd processors," in *PAT-MOS2004*, Santorini, Greece, Sept 2004.
- [12] P. Jonker, "Why linear arrays are better image processors," in *Proc. 12th IAPR Conf. on Pattern Recognition*, Jerusalem, Israel, 1994, pp. 334–338.
- [13] R.T.A. van Rootseler, "Inter-frame operations on the WiCa platform," Tech. Rep., Twente University of Technology, Enschede, The Netherlands, Dec. 2006.
- [14] Erik Ljung, Erik Simmons, Alexander Danilin, Richard Kleihorst, and Ben Schueler, "802.15.4 powered wireless smart cameras network," in *DSC'06*, Boulder, USA, 2006.
- [15] Vincent Jeanne, Francois-Xavier Jegaden, Richard Kleihorst, Alexander Danilin, and Ben Schueler, "Real-time face detection on a "dual-sensor" smart camera using smooth-edges technique," in DSC'06, Boulder, USA, 2006.