

SMART CAMERA MOTE WITH HIGH PERFORMANCE VISION SYSTEM

Richard Kleihorst, Ben Schueler, Alexander Danilin, Marc Heijligers

Philips Research Laboratories, WAY4.1,
Prof. Holstlaan 4, NL5656AA Eindhoven, The Netherlands

ABSTRACT

As a non-invasive technology, imaging plays an important role in mobile sensing devices. Multiple communicating cameras viewing the same scene from different viewpoints can together construct a high performance surveillance system. 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 micro-controller as a local host. Wireless communication is through the IEEE802.15.4 standard. Multiple cameras can establish a peer-to-peer connection and analyse the scene in a distributed fashion from several views at once. The camera constructed in this paper is to enable application research into distributed smart camera systems.

1. INTRODUCTION

Ambient intelligence is defined as electronic environments that are aware of and responsive to the presence of people. Ambient intelligence is a lively field of research, pushing technology and relevant applications. So far, most applications are focused towards monitoring scenes and persons.

Imaging could play and plays an important role in sensing devices for ambient intelligence [1, 2]. Computer vision can for instance be used for recognising persons and objects and recognising behaviour such as illness and rioting. Having a wireless camera as a camera "mote" [3] 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 would be able to overcome occlusion problems and could describe objects in their true 3D appearance. In real-time, these approaches are a recently opened field of research [4, 5, 6].

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, or to even do the event description if the computation takes less power than broadcasting the raw data takes. Then, only the event descriptions are forwarded to a host system that takes appropriate action based on the output of more (communicating) sensing devices. Smart cameras are cameras with embedded processing that can analyze the scene themselves and contact a host only when some events happen.

Because it makes sense for the overall power consumption of wireless cameras to go to highly smart setups, a number of chal-

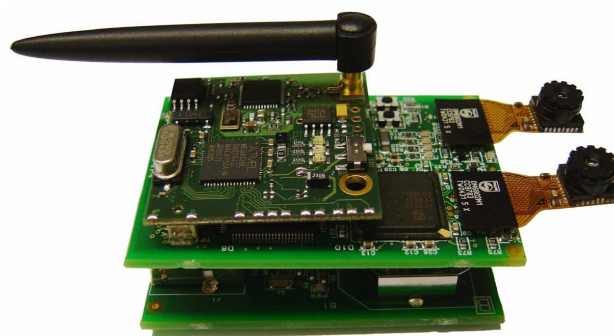


Fig. 1. The new wireless smart camera mote contains 2 VGA colour sensors and a high-performance vision system.

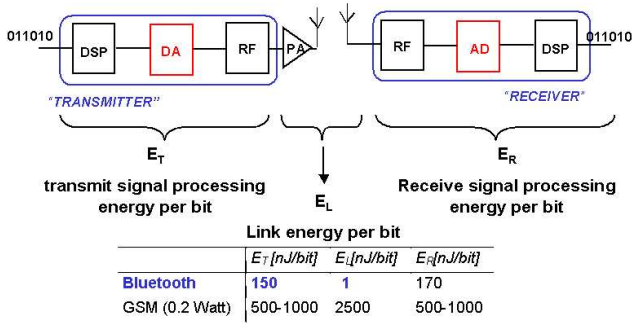
lenges have to be approached. Among them are the response time, the demand for the image processing performance and its associated power consumption.

In this paper we will describe the considerations taken for the hard- and software approach in the design of our wireless smart camera (see Figure 1). The remainder of the paper is organized as follows: In Section 2, it is demonstrated that it makes sense for camera power consumption reduction to do image processing on the camera itself. In Section 3, the essential kernels for embedded image processing are discussed. Section 4 describes in details the hardware side of our wireless smart camera and Section 5 describes the software side. Sections 6 and 7 describe respectively our current projects on the wireless smart camera and the conclusions.

2. SYSTEM ASPECTS

As low power consumption and high-performance are difficult to achieve for the functionality that the device is intended for, battery powered video processing, the functionality could be shifted towards a mains powered PC by broadcasting the raw video data. However, this takes in the order of 400mWatts for a digital 15fps graylevel VGA wireless link. It appears that continuously broadcasting live video from cameras to a computing engine like a PC takes more energy from the camera power source than the computation operations for scene analysis would take. Actually, for short distance transmission most of the broadcast energy is dissipated in the DA convertor in the transmitter [7]. This can be seen from Figure 2, where for Bluetooth transmitter electronics 150nJ per bit is needed and only 1 nJ per bit is needed for the actual transmission.

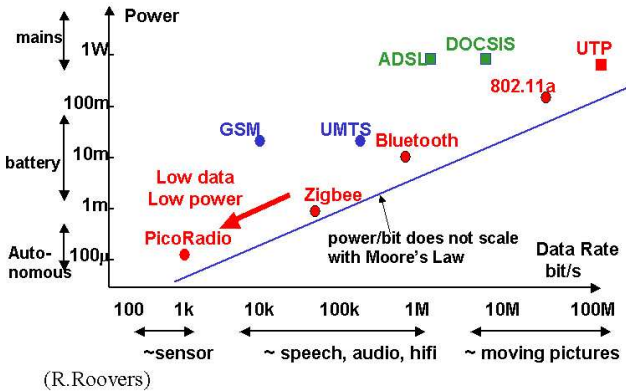
energy/bit in radio path



Raf Roovers

Fig. 2. Energy consumption of short and medium range transmission systems. It can be seen from the numbers in the table that most of the energy for the short range standard (Bluetooth) dissipates into the electronics and not into the air [7].

The technology of these DA converters is very close to their practical lowest power consumption limit, dictated by thermal noise. This becomes clear from the distribution of several modern low-range transmission systems such as Zigbee, PicoRadio: all are scattered slightly above the straight energy-per-bit line in Figure 2.



(R. Roovers)

Fig. 3. The fixed energy-per-bit line of short-range transmission systems. Note that all modern short range communication standards are scattered just above the linear line in the curve [7].

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 [8]. A clear graph to show this is in Figure 4, where silicon performances are rated according to how many million operations per Watt they perform. With the technology nodes on the X-axis, this graph continues growing. A clear distinction can also be seen for a general purpose sequential processor (Pentium) and a dedicated parallel processor (Xetal) [9], a derivative of which is used in the presented smart camera.

While wireless transmission is close to its energy efficiency limit, VLSI computation will continue to be more economical. The field of research into giving more system performance per Watt is starting to give the possibility, finding solutions through the use of strong data-level parallelism and advances in IC technology [10]. 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.

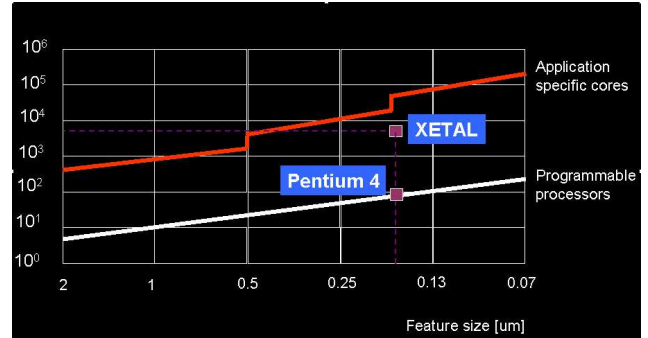


Fig. 4. The increasing silicon efficiency in Mega Operations Per Second per Watt for different technology nodes. The light curve shows the position of standard (sequential) processors, the darker curve shows the intrinsic performance of dedicated silicon solutions [8]. "Xetal" is a vector SIMD processor and one of the processors used in the camera platform.

3. HARDWARE KERNELS FOR EFFICIENT IMAGE PROCESSING

Real-time video processing on (low-cost and low-power) programmable platforms is now becoming possible thanks to advances in integration techniques [1, 9, 11, 12]. It is important that these platforms are programmable since new vision methods and applications emerge every month. The two types of programmable processors that we propose to be included in smart camera architectures are the SIMD (Single Instruction Multiple Data) massively parallel processor, and (one or more) general purpose DSPs [13, 14].

The algorithms in the application areas of smart cameras can be grouped into 3 levels: *low-level*, *intermediate-level* and *high-level* tasks. Figures 5 and 6 show the task classification and the corresponding data entities respectively.

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. In this part, often a classification or the initial steps towards pixel classification are performed. Because every pixel could be classified in the end as "interesting", the algorithms per pixel are essentially the same. So, if more performance is needed in this level of image processing, with upto a billion pixels per second, it is very fruitful to use this inherent data parallelism by operating on more pixels per clock cycle. The processors exploiting this have a SIMD architecture, where the same instruction is issued on all data items in parallel [14, 15]. From a power consumption point of view, SIMD proces-

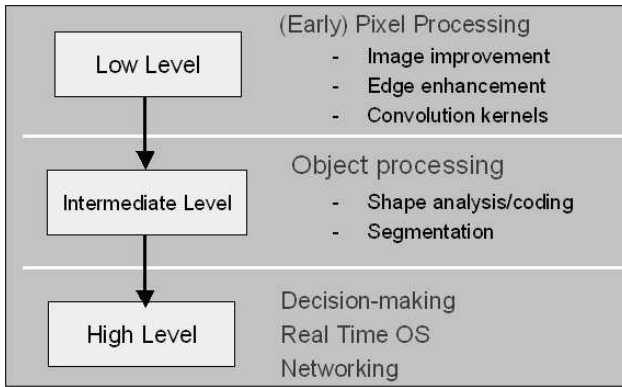


Fig. 5. Algorithm classification with respect to the type of operations

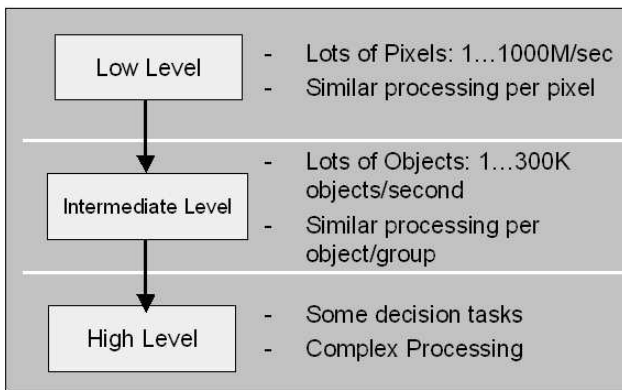


Fig. 6. Data entities with processing characteristics and possible ways to increase performance by exploiting parallelism

sors prove to be economical [16]. The parallel architecture reduces the number of memory accesses, clock speed, and instruction decoding, thereby enabling higher arithmetic performance at lower power consumption [1, 9].

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. The camera 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 Figure 7).

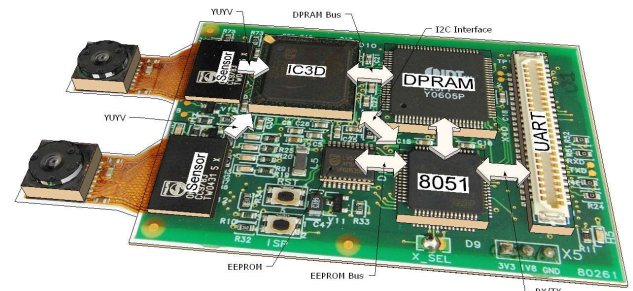


Fig. 7. Complete architecture of the wireless camera showing all processing and hardware blocks

4.1. IC3D SIMD Processor

The IC3D, a member of the Philips' Xetal family of SIMD processors, shows 5 specific internal blocks, see Figure 8. 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. 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. At the extremes of the linear array, the inputs of the processors are optionally coupled or mirrored. The processors have downloadable instructions ranging from arithmetic and single-cycle multiply-accumulate to compound instructions. In addition to these, there are conditional guarding instructions, enabling data-dependent operations. Data paths are 10-bits wide. Each processor has two word registers and a flag register. The line-memory

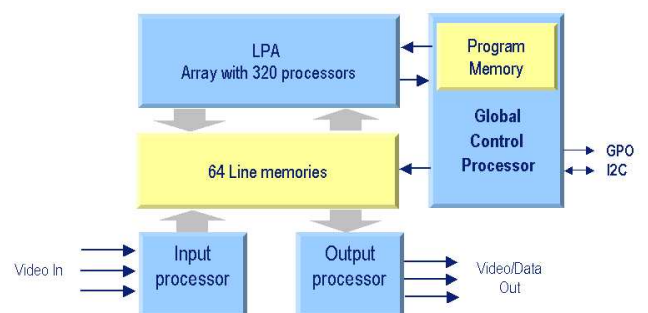


Fig. 8. Architecture of the "IC3D" which is a member of the "Xetal" family of SIMD chips

block stores 64 lines of 3200 bits. Pixels of the image lines are placed in an interlaced way on these memory. So CIF (320x240) images result in 1 pixel per processor, VGA (640x480) in 2 pixels per processor, etc. 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. 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 is on ultra-wide memory words that contain complete image lines instead of energy consuming access to multiple pixel-wide memory locations. For typical applications, such as feature finding or face detection, the power consumption is well below 100mWatt in active processing modes.

4.2. Dual Port RAM

The Dual Port (DP) RAM functions as an asynchronous connection between both processor cores. While the IC3D processor works on streaming (pixel) 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 then leisurely read and analyze that information and make decisions of the position, scale or movement direction of objects from the scene.

Writing information from the 8051 back to the IC3D is also possible via this DPRAM. 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.

4.3. The host controller 8051

To save components and to keep the power consumption low a top-of-the-range 8051 of ATMEL 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 easily to 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 used 8051 has 1792 bytes of internal RAM and 64kbyte of Flash to store its program or additional data. The internal 2KB EEPROM is used to store parameters and instruction code for the IC3D. Communication to the outside world is done via the UART. The UART has its own baud rate generator so all three timers of the 8051 are available for user applications. There are two 8bit timers and one 16bit timer. They are now partly used for task-switching in a (tiny) operating system.



Fig. 9. The ZigBee transceiver module used on the camera.

4.4. Aquis Grain ZigBee module

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 Figure 9 [17]. 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 [18]. 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.

5. SOFTWARE SYSTEM

The camera can be programmed wirelessly or remotely due to "in-system" programmability capabilities of the 8051. New IC3D programs can be uploaded from the 8051 via I²C in run time. An external I²C EEPROM can store 16 application programs which can be used for content switching approaches. The 8051 can load a program into the Xetal for a specific task that has to be carried out for a scene.

The software for the wireless camera consists of 3 parts that are almost independently developed. Programs for the IC3D processor are written in a C++ language with implicit parallel data-types. All programs are written in a line-based manner where complete image lines are processed in single-clock cycle instructions. By

guarding constructions, data-adaptive software structures can be implemented. Typical functions running on this processor are image improvement, motion analysis, object detection and tracking algorithms. The programs on the 8051 host-processor are dedicated to keep track of the object-data over time. It performs the host function (running the operating system) and can decide to transmit events to a host system.

6. CURRENT PROJECTS

Within the Philips Labs there are several projects that implement applications on the presented wireless smart camera. 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 [18]. On the imaging side, distributed face detection was mapped, where detection results from different image sensors are fused in order to decrease the false detection rate [19]. Also, first projects for gesture recognition based on hand detection were demonstrated. Our work on camera calibration techniques [20] will enable accurate distributed processing for unsystematically installed cameras.

All projects run in real-time (video at 30 fps) on the described wireless smart camera system. Future projects will focus more on collaboration between cameras for scene analysis and on lazy computation techniques to prolong the lifetime of the battery.

7. CONCLUSIONS

In this publication we presented the wireless smart camera platform that is being used in our research on distributed scene analysis. Smart cameras in a true sense, ie. cameras with build in processing are a key sensor for use in ambient intelligence. The local processing on-board results in a system that only sends keywords of information, by air, to a host system. Contrary to earlier believe, the advances in microprocessor architectures and silicon technology have made this more efficient for power consumption than broadcasting live video to an analysing host PC. Best performing architectures for smart cameras are built around an SIMD processor that benefits from the large amount of data-parallelism available in the pixel crunching part of the algorithms. Next to this processor a general purpose DSP does the control part of the algorithms. This processor benefits from the task level parallelism available in the intermediate and high-level part of the algorithms. The recently popular feature based approaches for object detection from natural scenes map very well to the aforementioned architectures and wireless cameras.

8. REFERENCES

- [1] A. Abbo and R. Kleihorst, "A programmable smart-camera architecture," in *ACIVS2002*, (Gent, Belgium), Sept 2002.
- [2] R. Kleihorst, B. Schueler, A. Abbo, and V. Choudhary, "Design challenges for power consumption in mobile smart cameras," in *COGIS2006*, (Paris, France), Mar. 2006.
- [3] I. Downes, L. B. Rad, and H. Aghajan, "Development of a mote for wireless image sensor networks," in *COGIS2006*, (Paris, France), Mar. 2006.
- [4] 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.
- [5] J. Mallet and V. M. Bove, "Eye society," in *ICME2003*, (Baltimore, MD, USA), July 2003.
- [6] H. Lee and H. Aghajan, "Vision-enabled node localization in wireless sensor networks," in *COGIS2006*, (Paris, France), Mar. 2006.
- [7] Raf Roovers, Philips Research Laboratories, "Personal Communication," 2005.
- [8] Engel Roza, Philips Research Laboratories, "Personal Communication," 2003.
- [9] R. Kleihorst, A. Abbo, A. van der Avoird, M. O. de Beeck, L. Sevat, P. Wielage, R. van Veen, and H. van Herten, "Xetal: A low-power high-performance smart camera processor," in *ISCAS 2001*, (Sydney, Australia), may 2001.
- [10] A. Abbo, R. Kleihorst, V. Choudhary, and L. Sevat, "Power consumption of performance-scaled simd processors," in *PATMOS2004*, (Santorini, Greece), Sept 2004.
- [11] J. Gealow and C. Sodini, "A pixel-parallel image processor using logic pitch-matched to dynamic memory," *IEEE Journal of Solid-State Circuits*, vol. 34, June 1999.
- [12] H. Yamashita and C. Sodini, "A 128 × 128 CMOS imager with 4 × 128 bit-serial column-parallel PE array," in *ISSCC2001 Digest of technical papers*, 2001.
- [13] P. Jonker, *Morphological Image Processing: Architecture and VLSI design*. Kluwer, 1992.
- [14] P. Jonker, "Why linear arrays are better image processors," in *Proc. 12th IAPR Conf. on Pattern Recognition*, (Jerusalem, Israel), pp. 334–338, 1994.
- [15] D. W. Hammerstrom and D. P. Lulich, "Image processing using one-dimensional processor arrays," *IEEE Proceedings*, vol. 84, pp. 1005–1018, jul 1996.
- [16] 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.
- [17] J. Espina, T. Falck, and O. Mühlens, "Network topologies, communication protocols and standards," in *Body Sensor Networks* (G. Z. Yang, ed.), pp. 145–182, Springer, 2006.
- [18] E. Ljung, E. Simmons, A. Danilin, R. Kleihorst, and B. Schueler, "802.15.4 powered wireless smart cameras network," in *DSC'06*, (Boulder, USA), 2006.
- [19] V. Jeanne, F.-X. Jegaden, R. Kleihorst, A. Danilin, and B. Schueler, "Real-time face detection on a "dual-sensor" smart camera using smooth-edges technique," in *DSC'06*, (Boulder, USA), 2006.
- [20] E. Simmons, E. Ljung, and R. Kleihorst, "Distributed vision with multiple uncalibrated smart cameras," in *DSC'06*, (Boulder, USA), 2006.