A Solid-State, Simultaneous Wide Angle – Detailed View Video Surveillance Camera

Ryutaro Oi†, Marcus Magnor‡, Kiyoharu Aizawa†

†University of Tokyo
Dept. of Frontier Informatics
7-3-1 Hogo, Bunkyo-ku, Tokyo, Japan

‡MPI Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken, Germany

Email: ooi@hal.t.u-tokyo.ac.jp

Abstract

We have developed a simultaneously wide-angle and detailed-view surveillance camera. For the purpose of surveillance, detailed views for suspicious objects are needed. Conventional motorized zoom cameras, however, are fragile and provide only a small region-of-interest at a time. We propose a system which is capable of obtaining wide-range views as well as detailed views of multiple regions of interest simultaneously. We present a prototype solid-state system based on a CMOS random access image sensor array. The system outputs a wide-angle, low-resolution view for monitoring the entire scene. If moving objects enter the scene, the system outputs detailed views of all objects in addition to the wide-angle view.

1 Introduction

Digital image acquisition is becoming ubiquitous. Mass-produced, inexpensive CMOS imaging chips, wireless cellular phone networks, and Internet technology enable mounting large quantities of electronic eyes almost anywhere for remote observation.

In 2001, more than 2.5 million closed circuit television cameras (CCTV) were installed in the United Kingdom alone, recording every UK citizen on average 300 times a day [1]. While the installations' prime objective is surveillance and crime prevention, the camera infrastructure is also used for 'civil' applications, e.g. to automatically collect the fee for driving in Downtown London by recognizing the license plate number and billing the car owner's bank account. Notwithstanding the social aspects of this development, the demand for automatic visual data collection and processing is increasing as fast as its technology is progressing. Remote visual surveillance calls for two different operational modes: during scanning, a large area is continuously surveyed for any suspicious activity, which if present must then be scrutinized more closely at high resolution. For this dual-mode application, currently a high-resolution video camera equipped with a motorized zoom lens on a motor-driven two-axis mount is necessary. Due to the mechanical elements involved, such systems are expensive and require frequent maintenance. Furthermore, they can only work in one mode at a time, losing the ability to monitor the big picture when zooming into some region of interest. Finally, camera zooming takes some time which may correspond to lost opportunity for getting good images of criminals etc. There are some approaches that use two cameras, one with wide-angle and the other with motorized [2][3]. This setup addresses the 2nd problem, but the problem caused by motors remains. In addition, it is very difficult to follow multi object, because each motorized camera can track only one object at a time.

This paper presents an alternative approach for monitoring a wide-angle scene while at the same time observing multiple regions of interest in more detail. Based on CMOS imaging technology, a solid-state multi-sensor array is employed whose pixels can be read out at random. Views of different scene regions can be synthesized at different resolution by sampling the sensor's pixels appropriately. When different sampling

VMV 2003 Munich, Germany, November 19-21, 2003

patterns are multiplexed, the entire visible scene and multiple regions of interest can be observed simultaneously. Our system doesn't have any mechanical parts, so that it can track arbitrary many different objects very quickly. A camera array of 4 x 4 CMOS imagers is used to continuously monitor the entire visible scene and at the same time track and show in maximum resolution any moving object within the scene.

2 Real-Time Multi-Camera Video Imaging

To observe a wide area, we employ a multi-camera array. This method doesn't demand any additional optics such as mirrors or lenses. At the same time, the multi-camera system provides high resolution to scrutinize individual objects in the scene at high detail. However, the multi-camera system has the disadvantage that it produces a large amount of image data. For example, 16 channels of NTSC color video correspond to more than 3.98 gigabytes per second of data.

In order to overcome this problem, the employed solid-state multi-sensor array can be accessed randomly, and sub-sampled versions of the full view as well as full-resolution images of small regions-of-interest can be provided by the hardware for fast transmission and analysis.

2.1 CMOS Imager Array Hardware

Figure 1 shows the block diagram of our CMOS image sensor array hardware. The sensor array consists of 16 custom-made random-access image sensors and one FPGA (Field Programmable Gate Array). The FPGA controls all image sensors. Each image sensor has a resolution of 128 x 128 pixel and is optimized for image processing by providing

- A frame shutter
- Pixel-based random access

This ensures a uniform integration time of all image sensors throughout the readout and processing operations. Each sensor can output a randomly accessed pixel value at 330 frames per second, 5.5 times faster than the normal TV rate [4]. The sensor array can be directly controlled by PC via the parallel port. The necessary control information is the output resolution and the position of the upper left and lower right pixel position of the region-of-interest. The amount of control data is small compared to the image pixel data, so that real-time random access operation is easily achieved. The sensor array outputs analog, gray-level pixel-values. By appropriately timing the output signal, the image can be directly displayed on an NTSC TV monitor. All sensors are synchronized to a reference clock unit (Tektronix DG2020) and have the same exposure time. The sensor array can be operated at well above 60 frames/sec, including the controller overhead.

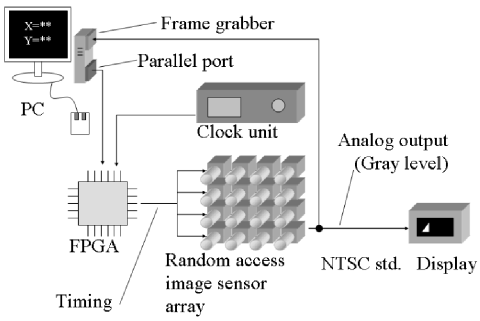


Figure 1: Block diagram of the imager array

Figure 2 shows the layout of the CMOS image sensor array. Each sensor is packaged and the distance between sensors is 4.57cm. Connectors J1 and J2 in the figure provide the interface to the FPGA. The entire planar sensor array consists of a single 2-layered printed circuit board.

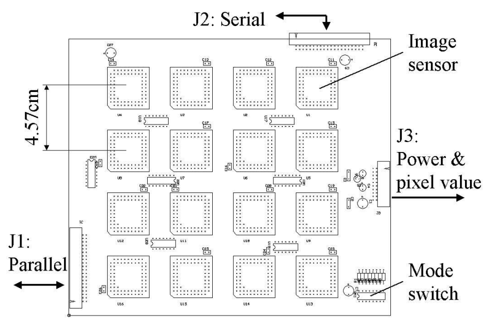


Figure 2: Layout of the sensor array

2.2 Digitization

In order to perform image processing for sensor calibration, photometric correction and motion detection, the computational power and versatility of modern a CPU is very useful. We employ a frame grabber card to digitize the CMOS sensor array output. Fig.3 shows the entire configuration of our system. The frame grabber is a Bt878 based PCI card (IOdata GVBCTV5/PCI). It is controlled via the Video4Linux API [5] on a Linux system (Redhat Linux 8.0). In this configuration, the PC is responsible for capturing, processing and displaying the image data as well as controlling the CMOS imager array. Thus, a feedback control loop between CPU and the array is established.

By introducing the frame grabber card, the output timing of the CMOS imager array must conform to the NTSC standard, providing 640x480-pixel images at 30 frames/sec, interlaced from 2 image fields.

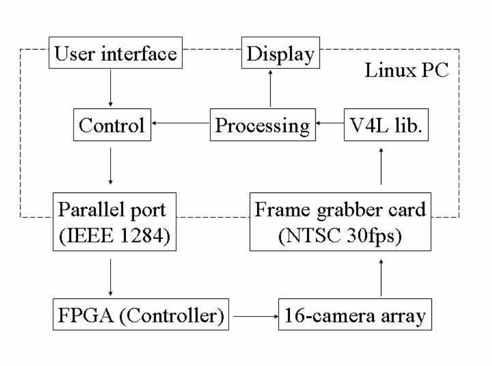


Figure 3: Video capture and control feedback loop

2.3 Geometric Camera Array Calibration

Although the sensor chips are equally spaced on the planar circuit board and the optical axes are aligned approximately in parallel, the output images still need to be calibrated to obtain one big image plane. To this goal, a planar calibration object, consisting of an arrangement of different ring patterns and positioned at the distance of the later scene, is recorded from all 16 sensors, Fig.4. Each individual ring pattern can be automatically recognized and its center position determined. In addition, the special arrangement of different ring patterns allows for non-ambiguously and automatically determining the absolute position within the entire calibration pattern if 4 ring patterns are in view. From at least 4 calibration points in view, the homography matrix is calculated to reproject each sensor's image plane onto one common plane, Fig.5. The correspondences between sensor pixel and reprojected coordinates in the common image plane are pre-calculated and stored in a look-up table for fast and efficient display during online processing.

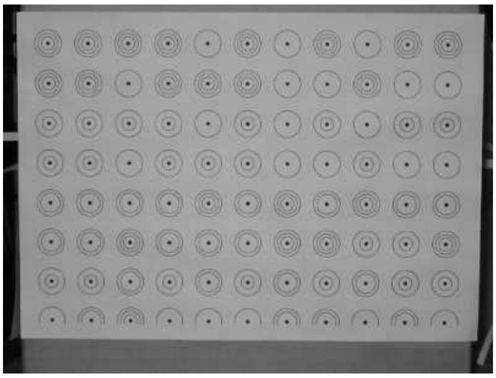


Figure 4: The calibration object

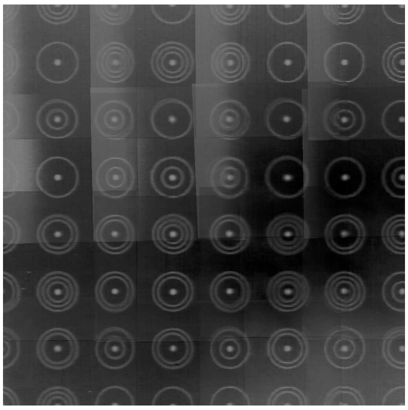


Figure 5: Calibrated full-view output image

2.4 Photometric Calibration

A real image sensor array does not possess uniform sensitivity over the entire pixel array. Also, sensitivity can differ significantly from one sensor to another. In addition, there are faulty pixels which do not respond to the incident light at all. Finally, the analog output signal of the CMOS imager array must be matched to the dynamic range of the frame grabber card. In order to correct for these effects, photometric calibration is necessary.
From the physical processes of CMOS imaging, a linear response curve to light intensity can be assumed for each pixel. To determine the individual photometric calibration parameters, two images, one of a uniformly black and one of a uniformly white target, are recorded, while the illumination is kept approximately constant over the entire area. The response to the black and white target enables determining the offset and slope of the photometric calibration curve for each pixel. Faulty pixels are detected by discarding those pixels whose difference in response to the black and white target is smaller than a preset threshold. The photometric calibration parameters are stored for each pixel individually for later online display.

3 Simultaneous Multi-Object surveillance

Our proposed surveillance system has two operational modes. In the monitoring mode, the entire scene is displayed as wide-angle view at reduced resolution. In the alert mode, additional high-resolution images of multiple regions-of-interest are provided.
In order to reduce the bandwidth from the sensor array to the monitoring site, a low-resolution video signal showing the entire scene is output by the CMOS imager array. The reduced resolution saves bandwidth, reduces storage capacity at the monitoring site, and/or enables high frame rates. The resolutions can be changed at the sensor by hardware-subsampling each sensor's pixel array. Hardware sub-sampling consists of outputting every n-th pixel of the pixel array and assembling one "big picture" from all sensors.
On the other hand, detailed views of regions-of-interest must be provided when needed. However, while a surveillance camera outputs a detailed view, important things in the "big picture" might be missed. Therefore, the wide-angle view should be provided at any time.

Our system performs the following steps:

STEP 1: Wide-angle monitoring. The system outputs a low-resolution video of the entire scene. (monitoring mode)

STEP 2: Automatic motion detection in the wide-angle view. The system multiplexes between the full-resolution image of the moving region and the low-resolution image of the entire scene. (alert mode)

STEP 3: Tracking. The full-resolution image follows the moving object. (alert mode)

STEP 4: Detecting other moving objects. The system multiplexes between full-resolution images of all moving regions and the low-resolution wide-angle view. (alert mode)

STEP 5: Tracking multiple moving objects. The full-resolution images follow the moving objects. (alert mode)

The PC automatically detects motion in the wide-angle view and by controlling the sensor array switches between the "big picture" and the full-resolution images of the regions-of-interest. Since the FPGA and the random access image sensor hardware do not exhibit any mechanical delay, the views can be switched in less time than it takes to read out one image field (16.5ms).
Fig. 6 shows the operational states of the CMOS imager. The FPGA returns to the wait state everytime after outputting image data. In this state, the pixels of all image sensors are reset, and communication to the PC and the sample/hold for the next field is performed (1.2ms). Note that all this is done in less time than the vertical blanking period of the NTSC standard TV signal.

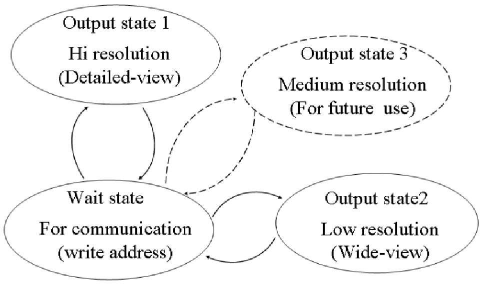


Figure 6: Operational states of CMOS imager

3.1 Motion detection

In our prototype of an automatic surveillance system, the alert state is entered based on detecting motion in the wide-angle view. A frame differencing approach is employed to detect changes in the scene, interpreted as movement.
Even if lighting conditions are constant, the output signal from an actual CMOS imager will exhibit noise. Thus, we adopt a statistical per-pixel background model based on each pixel's mean value and standard deviation when observing the static background. The time-averaged mean value and variance of each pixel are stored and used for motion detection in the following way:

1. Find candidates for movement: If the current pixel value differs from the mean of the background value by more than 3 times its standard deviation, the pixel is marked as foreground.
2. Find all connected foreground pixel areas.
3. If a foreground area is larger than a preset threshold, the connected foreground pixels are treated as a moving object to be surveyed.
4. Measure the area of all foreground objects and sort them by descending size.

   In our prototype surveillance system, the two largest suspicious objects are displayed.

3.2 Tracking

Tracking of moving objects is implemented by following the center of gravity of each connected foreground object from frame to frame. The system alternates between wide-angle monitoring and displaying detailed views of the moving objects. This way, the frame rate decreases proportionally to the number of tracked objects. Since the frame grabber card is designed for a fixed frame rate of 30 fps, there is a trade-off betweeen frame-rate and the number of tracked objects.
When the system switches to the alert state, the detailed views pop up next to the wide-area video image. This way, the observer can watch the entire scene and the detailed views at the same time.

4 Results

4.1 System setup

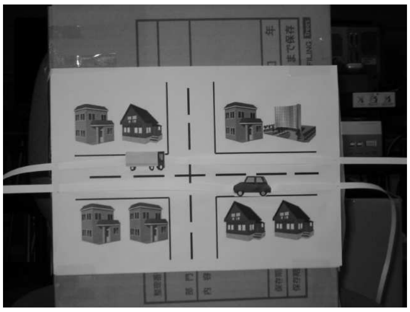
Fig.7(a) depicts the CMOS image sensor array, and Fig.7(b) shows the target object plane of our system: Pictures of two cars can be drawn along the street. The task of the system is to find and track these cars in this laboratory-level prototype system. The physical dimensions are described in Tab.1. The distance between the sensor array and the scene is chosen such that the sensor's images of the scene overlap partially.
In this prototype system, we used inexpensive fixed-focus, fixed-iris lenses for CMOS imagers. (The focal length is 6mm for 1/3-inch optical format imager.)

Table 1: Dimensions of the prototype system

Number of sensors	16
Camera's view angle	32.6 degree
Horizontal size of array	200mm
Vertical size of array	233mm
Target plane distance	105mm

We can scale up this system by choosing the appropriate lens for each sensors and widening the camera separations. By putting these cameras to the ceiling, this system can be used for monitoring a large room such as museum or bank.

(a) Sensor array (FPGA on the backside)



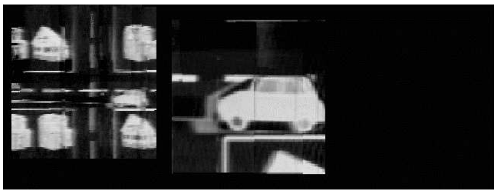
(b) Observed scene (street with two cars)

Figure 7: Prototype system

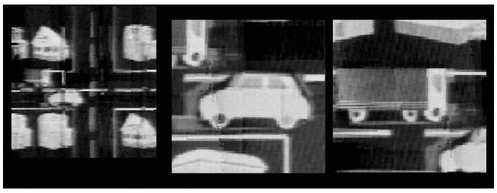
4.2 Experiment

Fig.8 depicts surveillance images of the prototype system. Fig.8(a) shows the normal state. There are no moving objects in the monitored scene. Fig.8(b) shows the wide-angle view as well as the full-resolution image of one moving target: A car has entered the monitored scene, and the system has switched to the alert state. Fig.8(c) shows the presented views when two moving objects are present: The car and a truck move in opposite direction along the street. The prototype system tracks both objects smoothly and does not confuse them when they meet [6].

(a) Normal state: The wide-angle, low-resolution monitoring view can be seen in the leftmost window. (Frame #0)



(b) Alert state: A car is entering the scene, and the system outputs a detailed view of the car (middle window) in addition to the wide-angle view. (Frame #62)



(c) Alert state: In addition to the car, a truck is entering the scene from the opposite direction. The system automatically outputs high-resolution images of the car (middle window) and the truck (rightmost window) as well as the wide-angle view. (Frame #93)

Figure 8: Frames from the surveillance video sequence [6]

The frame rate of the prototype is approximately 5fps with 2 tracked objects and 1 monitoring wide-angle image. The limiting factor at this point is the software processing on PC (Intel Cerelon 1.8GHz) which encompasses geometric and photometric rectification, motion detection and tracking.

The hardware-subsampled wide-angle images is always displayed, Fig.8. This way, the system can robustly monitor the scene while simultaneously following in detail multiple objects.

The sub-sampling ratio for the wide-angle image is 4:1 in horizontal and vertical direction. Thus, the transferred pixel data is only 1/16th of the total image data collected at by the sensor array.

Because hardware sub-sampling is currently done by simply passing over pixels and leaving out image information, aliasing artifacts are visible in the wide-angle view. To overcome this problem, a space-variant sampling control sensor can be employed [7], or a CMOS imager providing locally averaged pixel values may be developed.

5 Conclusions

In this paper we propose a system to monitor a wide-angle scene while at the same time observing multiple regions of interest in detail. Based on CMOS imaging technology, a solid-state multi-sensor array is employed whose pixels can be read out at random. The proposed system doesn't demand any mechanical elements, so that the time-lag when switching between monitoring and detailed views is negligibly short (less than the vertical blanking period). The hardware-subsampled wide-angle view is always visible to continuously detect survey the entire scene.

In the future, a multi-resolution random-access CMOS imager array will be developed to obtain anti-aliased wide-angle views.

References

[1] BBC, "Watch your every Move", *http://news.bbc.co.uk/1/hi/sci/tech/1789157.stm*

[2] Steven G. Goodridge, "The environment understanding interface: detecting and tracking human activity through multimedia sensors", *Proceedings of the 1995 conference of the Centre for Advanced Studies on Collaborative research*, pp.24-31, 1995.

[3] Yong Rui, Liwei He, Anoop Gupta, Qiong Liu, "Building an intelligent camera management system", *Proceedings of the ninth ACM international conference on Multimedia*, pp.2-11, 2001.

[4] R.Ooi, T.Hamamoto and K.Aizawa, "Image acquisition by pixel based random access image sensor for a real time IBR system", *IEICE Transactions.* Vol.E85-C, No.3, pp.505-510, 2002.

[5] V4L, "Video4Linux API specifications", *http://www.exploits.org/v4l/*

[6] M.Magnor, R.Oi, K.Aizawa, "Web page", *http://www.hal.t.u-tokyo.ac.jp/~magnor/*

[7] Y.Ohtuka, T.Hamamoto, K.Aizawa, "A new image sensor with space variant sampling control on a focal plane", *IEICE Transactions*. Vol.E83-D, No.7, pp.1331-1337, 2000.