

1-1-2005

Novel architecture for surveillance cameras with complementary metal oxide semiconductor image sensors

Igor Kharitonenko
University of Wollongong, igor@uow.edu.au

Wanqing Li
University of Wollongong, wanqing@uow.edu.au

Chaminda Weerasinghe
Toshiba, Australia

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Kharitonenko, Igor; Li, Wanqing; and Weerasinghe, Chaminda: Novel architecture for surveillance cameras with complementary metal oxide semiconductor image sensors 2005.
<https://ro.uow.edu.au/infopapers/178>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Novel architecture for surveillance cameras with complementary metal oxide semiconductor image sensors

Abstract

This work presents a novel architecture of an intelligent video surveillance camera. It is embedded with automated scene analysis and object behavior detection, so that operators can monitor more venues relying on the system that provides immediate response to suspicious events. The developed camera turns passive video data recording systems into active collaborators with security operators leaving to them only high-level decision making, while automatically carrying out all monotonous work on continuous video monitoring. When there is no alarming activity inside a restricted area the camera automatically turns back to the whole view mode.

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as: Kharitonenko, I, Li, W & Weerasinghe, C, Novel architecture for surveillance cameras with complementary metal oxide semiconductor image sensors, International Conference on Consumer Electronics, 8-12 January 2005, 251-252. Copyright IEEE 2005.

Novel Architecture for Surveillance Cameras with Complementary Metal Oxide Semiconductor Image Sensors

Igor Kharitonenko, Wanqing Li, and Chaminda Weerasinghe
University of Wollongong, Australia Toshiba, Australia

Abstract—This paper presents a novel architecture of an Intelligent Video Surveillance Camera. It is embedded with automated scene analysis and object behavior detection, so that an operators can monitor more venues relying on the system that provides immediate response to suspicious events. The developed camera turns passive video data recording systems into active collaborators with security operators leaving to them only high-level decision making, while automatically carrying out all monotonous work on continuous video monitoring. When there is no alarming activity inside a restricted area the camera automatically turns back to the whole view mode.

I. INTRODUCTION

Video surveillance and monitoring systems have become important components in the modern security infrastructure. More and more cameras are installed to provide efficient surveillance, but this also requires employing a sufficient number of skilled personnel for monitoring. According to [1], the most vulnerable part of video surveillance systems is video monitoring personnel. Most of the time, there are no alarming events and the staff gradually loses concentration on duty. When alarming events are captured by the system, the human response is very often delayed and is not optimal.

Another problem that reduces efficiency of using video surveillance systems is a contradiction between the required sector of view and the sharpness of the captured objects. The wider area captured by a camera, the less resolution it can provide to represent detailed objects. As a result, image quality is usually not sufficient to recognize facial or other important features.

One may conclude that the fundamental problem restricting broader utilization of video surveillance systems is caused by the concept when the system is considered only as an observation and recording device, entirely relied on human attention and decision-making. There is a distinct trend in video surveillance market towards using intelligent systems, which can provide efficient assistance to the operators. In this regard, development of technologies for intelligent video surveillance systems that can automatically optimize system parameters and provide assistance to the personnel becomes an important practical issue. This research area is related to automatic scene analysis and image understanding. Although tremendous efforts have been made to solve these problems, there are no efficient generic solutions and theoretical concepts applicable to intelligent vision systems. However, understanding of video surveillance environment, industry needs and application specific conditions can lead to development of reliable technologies. This paper presents a video camera architecture that can turn passive surveillance

systems into active collaborators with security operators to provide immediate and efficient response to suspicious events.

II. VIDEO CAMERA ARCHITECTURE

The Intelligent Video Camera incorporates an automatic pan, tilt and zoom (PTZ) functionality with the parameters decided upon captured activity within a specified region of interest (ROI). This provides more detailed visual information in zoomed modes while monitoring the full view at the same time. The automatic PTZ adjustments can be instantly made based upon motion activity, color and illumination changes of targeted objects. The block diagram of the camera is shown in Figure 1.

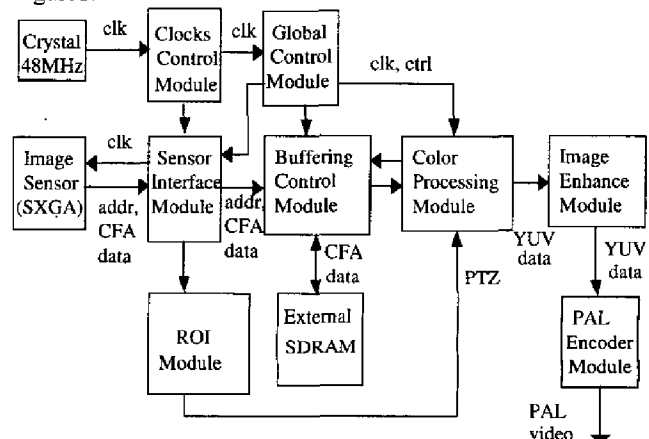


Figure 1. Intelligent Video Camera architecture

The camera is built upon a 1.3-megapixel CMOS image sensor. Such high resolution of the sensor allows displaying zoom 1 (normal zoom, whole image), zoom 2 and zoom 4 without compromising resolution at the video output. This is achieved electronically by cropping a required area of SXGA image sensor just within 40ms. The output video frame from the Color Processing module is 640×256 per PAL field (640×512 interleaved). However, the ROI Module always uses ROI data for monitoring of suspicious events regardless of the currently displayed area. Moreover, PTZ values are directly used to optimize Color Processing Module parameters in such a way that no specific scaling step is performed before or after the interpolation or color reconstruction of the video frame.

Adaptable image enhancement and noise reduction features are implemented on the interpolated YUV color space within the Image Enhance module. These functions applied only to the luminance channel (Y), result in significant improvement in image quality, especially while the camera

operates in poor lighting conditions.

III. MOTION DETECTION AND OBJECT TRACKING

Motion analysis and object tracking have been studied for several decades. The algorithms of motion detection cover temporal frame difference methods, including mean-absolute or sum-of-squares difference, and cross-correlation technics [2], background modelling [3] and optical flow methods. The optical flow method is the most accurate, but also it is the most computationally demanding and is very difficult to implement in a camera-based processing unit.

In order to provide robust motion detection that is not affected by insignificant motion activity, the high-resolution input frame (1280 x 1024 pixels) is divided into a number of blocks 64x64 pixels each. Therefore, there are 320 (20 x 16) blocks $B(m,n)$ described by (1) in the image.

$$B(m,n) = \frac{1}{k^2} \sum_{x=0}^{k-1} \sum_{y=0}^{k-1} f(km+x, kn+y) \quad (1)$$

where $f(x,y)$ is the initial high-resolution frame acquired from the image sensor and k is a block size in pixels. It is 64 in our case. This size can be further adjusted by specifying ROI of the input image, the area, where the monitored motion activity is expected. The ROI coordinates are set as most left, top, right and bottom block coordinate.

The reference frame calculation is based on the weighted block difference algorithm, which takes into account the block values from several consecutive frames and thus provides some temporal filtering in addition to spatial, performed by previous operation. The value of the block difference $D_t(m,n)$ at time t is calculated as follows:

$$D_t(m,n) = \frac{3D_{t-1}(m,n) + (B_t(m,n) - B_{t-1}(m,n))}{4} \quad (2)$$

If at a particular moment in time the absolute value of $D_t(m,n)$ for a particular block is equal or greater than the specified threshold, then this block is marked as a motion active block. The motion active blocks are used for calculating PTZ parameters.

IV. PTZ BASED COLOR INTERPOLATION

Color interpolation is based on a weighted average bilinear algorithm. The weights are calculated taking into account the zoom mode and the distance from the intended interpolated pixel location to the original R, G and B components in the CFA. The processed CFA data window is always equal to 4x4 regardless of the zoom value.

In the Zoom 1 mode, CFA frame of size 1280 x 1024 is interpolated to produce a full color frame of size 640 x 512. In the Zoom 2 mode, CFA frame of size 640 x 512 is used. In the

Zoom 4 mode, CFA frame of size 320 x 256 is interpolated to produce a full color frame of size 640 x 512.

Color correction is performed using an algorithm described in detail in [4]. Color space conversion from RGB to YUV is performed using the coefficients that are adjusted to be hardware friendly, without compromising visual quality.

$$Y = 0.5 \times R_c + 1.25 \times G_c + 0.125 \times B_c \quad \text{range } [0 \dots 511] \quad (3)$$

$$U = B_c - 0.5 \times Y \quad \text{range } [-256 \dots 255] \quad (4)$$

$$V = R_c - 0.5 \times Y \quad \text{range } [-256 \dots 255] \quad (5)$$

V. IMAGE ENHANCEMENT

Sharpening and noise reduction is implemented on the YUV color space and affects only Y channel. To avoid noise amplification, a simple non-linear sharpening filter is employed. The filter has the following properties:

Filter taps: $S_{-2}, S_{-1}, S_0, S_{+1}, S_{+2}$

$$Y' = |S_1 - S_{-1}|$$

Filter coefficients: If $(Y' > 20)$ then $\{-0.25, -0.25, 2.00, -0.25, -0.25\}$ else $\{0.00, 0.00, 1.00, 0.00, 0.00\}$.

The implemented noise suppression algorithm is based on anisotropic diffusion [13]. Anisotropic diffusion is performed on Y samples along each video line. This minimizes the buffer size needed for storing intermediate data between iterations. Only three iterations are performed.

VI. CONCLUSION

The described architecture was implemented on a Xilinx Virtex II FPGA-based platform. It was tested in different surveillance scenarios. These scenarios included both indoor and outdoor scenes, such as monitoring of the vehicles at the highway, surveillance at the exhibition pavilion, hallway and lab monitoring. The tests showed that camera had a quick response to the suspicious activity in the ROI, automatically zooming into the area, which encloses all moving objects and tracking these objects when they move.

REFERENCES

- [1] C. Regazzoni, G. Fabri, G. Vernazza. *Advanced Video-Based Surveillance Systems*. Kluwer Academic Publisher 1999.
- [2] D.J. Connor and J.O. Limb. *Properties of frame-difference signals generated by moving images*. IEEE Trans. COM-22(10): 1564-1575, 1974.
- [3] Y.W. Huang, B.Y. Hsieh, S.Y. Chien, and L.G. Chen, *Simple and effective algorithm for automatic tracking of a single object using a pan-tilt-zoom camera*. Proc. of ICME 2002, Switzerland.
- [4] I. Kharitonenko, S. Twelves and C. Weerasinghe, *Suppression of noise amplification during color correction*, IEEE Trans. Consumer Electronics, Vol. 48 (2), pp. 229 - 233, 2002