I J C T A, 9(26) 2016, pp. 73-80

An Effective Surveillance Video Coding System

K. N. Abdul Kader Nihal[1] and A. R. Mohamed Shanavas[2]

ABSTRACT

Video Surveillance Systems are required to improve public safety, Mitigate risks of crime, Protect assets, Prevent fraud, Improve efficiency, Automate more processes, Provide better healthcare. Current Video Surveillance Systems using Digital Technology [1] are more sophisticated, cost effective, efficient, flexible and simple to operate. Advantages of video surveillance systems are to Keep track of information (video data) for future use and help in identifying people in crime scenes. Some disadvantages are difficulty in maintaining (storing) large amount of raw video data and requiring higher bandwidth for transmitting the visual data. The purpose of this paper is to create a Video Surveillance System called an Effective Surveillance Video coding System for Colleges (ESVCS) to overcome problems of storage and transmission of heavy volume video data.

Keywords: Video Compression, Frame Compression, Temporal Compression, Video Surveillance Systems, CCTV, Frame Compression Video Coding, Frames, Motion JPEG, MPEG-4 and H.264

1. INTRODUCTION

Guarding one's home & business premises has become imperative to-day. Whether it is screening and stopping a visitor at the gate or detecting, preventing and investigating low productivity / theft / fire or sounding burglar alarms or any other security aspect has become easy in today's modern world. Various cost-effective & efficient security systems are available based on the exact needs of the customer for any sized establishment. A wide range of sophisticated, yet simple, state-of-the-art devices, equipment & systems are offered. The proposed system ESVCS aims to achieve bit-rate (the number of bits per second that can be transmitted along a digital network.) optimisation and creating of surveillance videos for storing and transmission purposes.

Video Surveillance Systems [1] mainly involve video coding and video compression. These systems are mainly classified as sensor systems and Vision based camera systems [1]. This paper deals with capturing of images using a higher end CCTV or digital camera and a set of frames are fed as input into a detector which searches for any area of interest in the frames. Video coding [3] involves Segmentation, fore ground coding and background coding. Motion estimation [4] is performed which involves background subtraction and selection policy. This can furthermore involve feature or object extraction and object recognition.

Video Capture: It is the use of video cameras to transmit a signal to a specific place, on a limited set of monitors. It can be done using Closed-Circuit Television (CCTV) [3] also called video surveillance or by using Digital camera or digicam, which is a camera that encodes digital images and videos digitally and stores them for later reproduction [6].

Video Coding: Scalable video coding (SVC) [7] is an extension of the H.264 video compression standard for video encoding. The video codec allows video transmission to scale so that content is delivered without degradation between various endpoints.

1 Assistant Professor of Computer Science / Jamal Mohamed College, Trichy-20
2 Associate Professor of Computer Science / Jamal Mohamed College, Trichy-20

Video Compression: Technologies are about reducing and removing redundant video data so that a digital video file can be effectively sent over a network and stored. With efficient compression techniques, a significant reduction in file size can be achieved with little or no adverse effect on the visual quality. Three different video compression standards are Motion JPEG, MPEG-4 and H.264 which is the latest and most efficient standard. Compression techniques in video surveillance systems are categorized into frame – by – frame compression and temporal compression [7].

Video Storage: The high priority captured images that is., the images having notable change which are to be stored may be further compressed using image compression techniques. Image compression [08] is the technique of reducing the image storage size without degrading the quality of the image.

PROPOSED SYSTEM ARCHITECTURE

The proposed ESVCS system involves processes explained as modules each performing a specific task. Video capture module – Scalable Video Coding Module - Motion Estimation Module – Storage Module – Image Compression.

Methodology: The scalable video coding module of the proposed ESVCS system has three stages: Encoding, Extraction and Decoding which is achieved by three different modules namely Encoder, Extractor and a Decoder as shown in Figure 1.

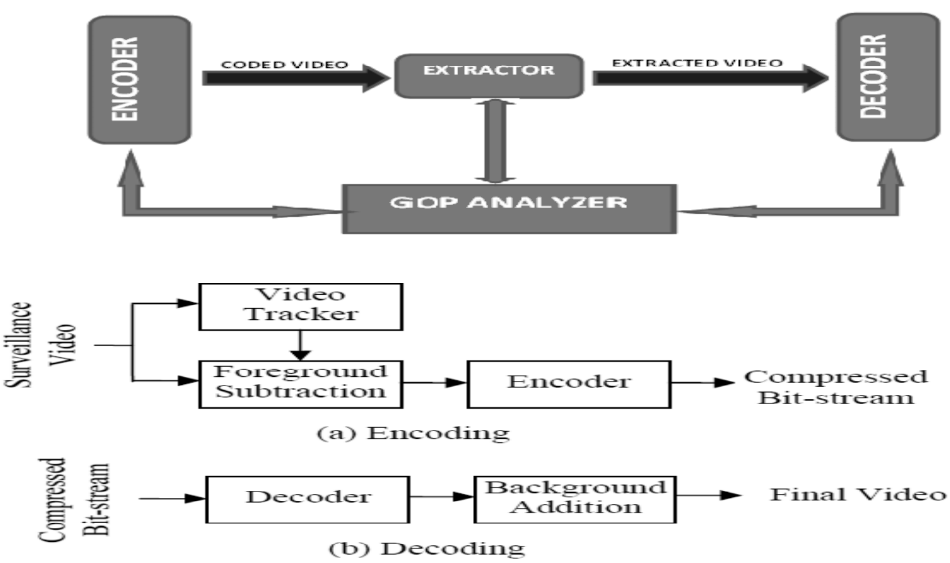

Figure 1: Block diagram of the video coding module. (a) Encoder. (b) Decoder

A GOP analyser to support event based features at GOP level for coding the surveillance videos is used to select a GOP from the input sequence and processes it to removes the GOP's dependency on the previous and next GOPs then the extractor extracts the GOPs with changed performance.

Workflow: The event of interest in the surveillance video is the occurrence of some change in motion or activity. The workflow of the proposed ESVCS System as shown in figure2 above starts with capture of frames using any device. The scalable video coding algorithm is implemented using GOP (group of pictures) structure, which specifies the order in which intra- and inter-frames are arranged. The GOP as explained in figure2 is a group of successive pictures within a coded video stream. Each coded video stream consists of successive GOPs. From the pictures contained in it, the visible frames are generated. GOP structure as in figure3 contains following 3 picture types:

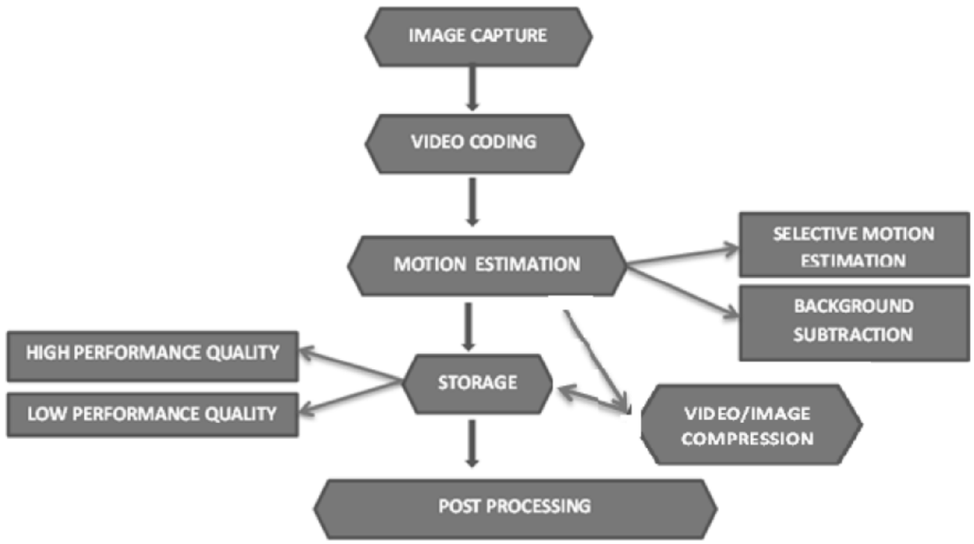

Figure 2: Work flow of the proposed ESVCS system

- I-frame (Intracoded Frame)
  - ➢ Coded in one frame such as DCT.
  - ➢ This type of frame do not need previous frame
- P-frame (Predictive Frame)
  - ➢ One directional motion prediction from a previous frame
    - ✓ The reference can be either I-frame or P-frame
  - ➢ Generally referred to as inter-frame
- B-frame (Bi-directional predictive frame)
  - ➢ Bi-directional motion prediction from a previous or future frame
    - ✓ The reference can be either I-frame or P-frame
  - ➢ Generally referred to as inter-frame

The work flow of the system depicted in figure 2 further indicates the motion estimation module where a selective motion vector search technique based on the motion detection module is implemented. This technique maintains the visual quality of the video as that of full search yet improves the processing efficiency for the coding of the surveillance videos.

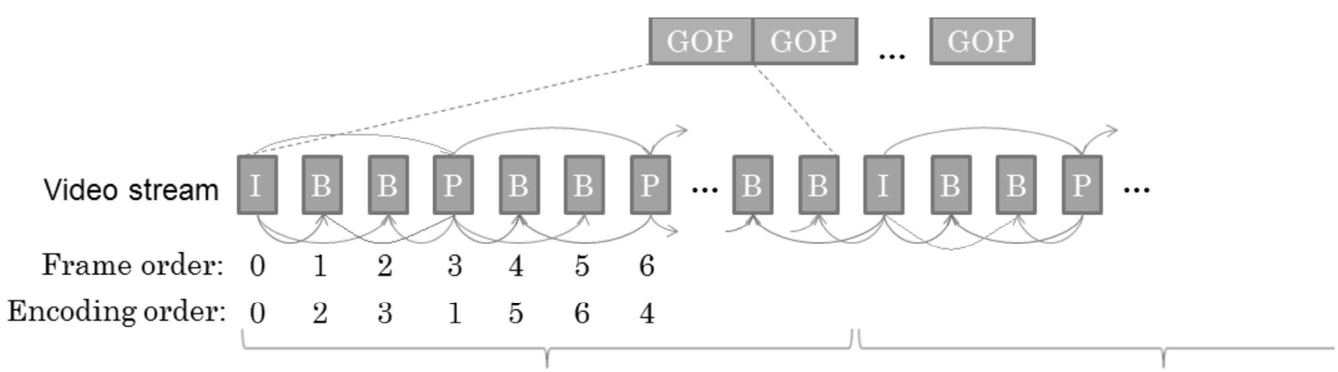

Figure 3: GOP Structure

Here object detection information generated is used to flag the frames which do not have any moving object. Based on this analysis, different selective motion estimation approaches were proposed depending on GOP level and Frame level. This module in turn consists of a real time back ground subtractor [09] that is used to detect motion change in the video. The subtraction module of the system tracks the foreground objects are by forming rectangular boxes. A Gaussian mixture model is used [09]. The Presence of bounding boxes as in figure 4 indicates motion activity in the frame. Then the presence of motion is marked and recorded. This information is used by the motion estimation module to perform selective motion estimation.

Figure 4: The foreground objects are tracked by rectangular boxes

The frame or part of the video with the event of interest may be categorized into high performance quality and the part with not much change may be of low performance quality which is used in image storage. The information which determines if an image needs to be stored depending on the motion estimation is called the quality levels of the captured images. The stored image quality is directly proportional to details of information [10]. In other words high information gives high image quality which in turn leads to high priority hence storing the image as shown in Figure5. Once motion estimation is done and after

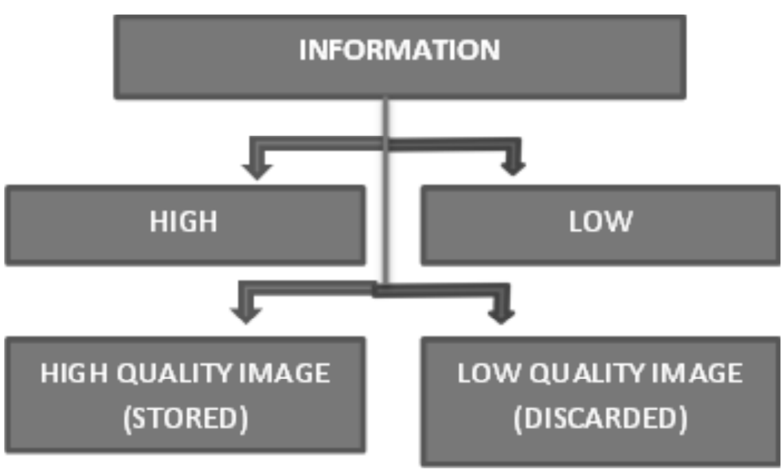

Figure 5: Information versus image quality

background subtraction is done then the change in the image is noted if the change is considerably high then the priority for storage is high and if no change or minimal change is noted then the image is not stored. Before the image can be stored it is compressed using any available image compression technique.Post processing is the last module of the proposed system which is performed to enable the smooth viewing of the decoded video after extracting it from the storage decompressing the images and bringing it to its original form.

2. IMPLEMENTATION & EXPERIMENTAL RESULTS

Some modules of the proposed ESVCS system are implemented using MATLAB. Quality measures namely Compression Ratio, Compression Factor, Space Saving Percentage, Compression Time and Decompression Time may be used to compare existing techniques. The proposed algorithm had been implemented using MATLAB. Basic two successive frame subtractions were carried out to estimate motion detection. The proposed algorithm assumed two threshold values TLow and THigh which stands for Low threshold and High threshold respectively. The Figure5 illustrates that if the difference between two successive frames (D) is less than TLow, then the image of the second frame will not be stored. Or if (D) is greater than THigh, then the quality of the image to be stored will be 100%. Otherwise the relation between the difference value (D) and the quality of the image is linear.

In the GOP Analyser, an adaptive background subtraction method based on mixture of Gaussians [11] is used. At each time instance Gaussians of the mixture that represent the background are identified according to the predefined threshold. The output of the background subtraction and tracking module is illustrated in Figure 7.

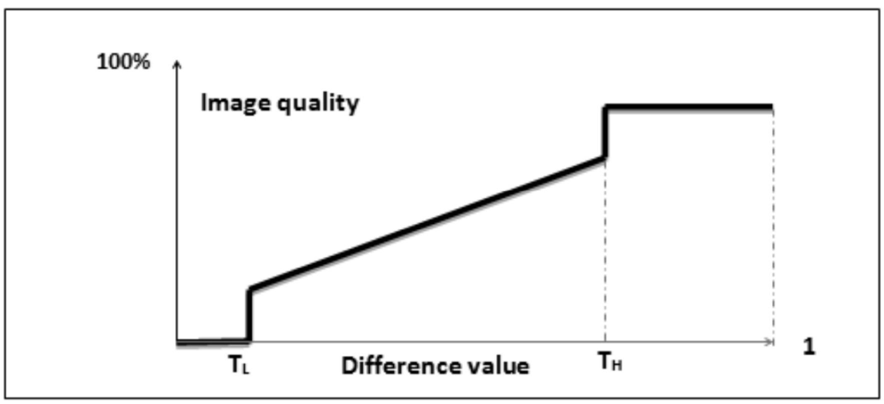

Figure 6: Relation between Frame Difference and Image Quality [13]

The performance evaluation of the proposed approach is carried out on three different surveillance sequences: Lane, Parking lot and a Corridor in a college as shown in Table 1. All of these sequences have CIF (352×288) spatial resolution and a frame rate of 30 Hz. Background in all the sequences is static throughout the length of the sequences.

Table 1
Surveillance Sequences

Sequence	*Total Frames*	*Description*	*Time(sec)*
Lane	750	Different objects moving in street	25
Parking	2100	Cars in the college parking lot	68
College corridor	300	Students and staff walking in the corridor	10

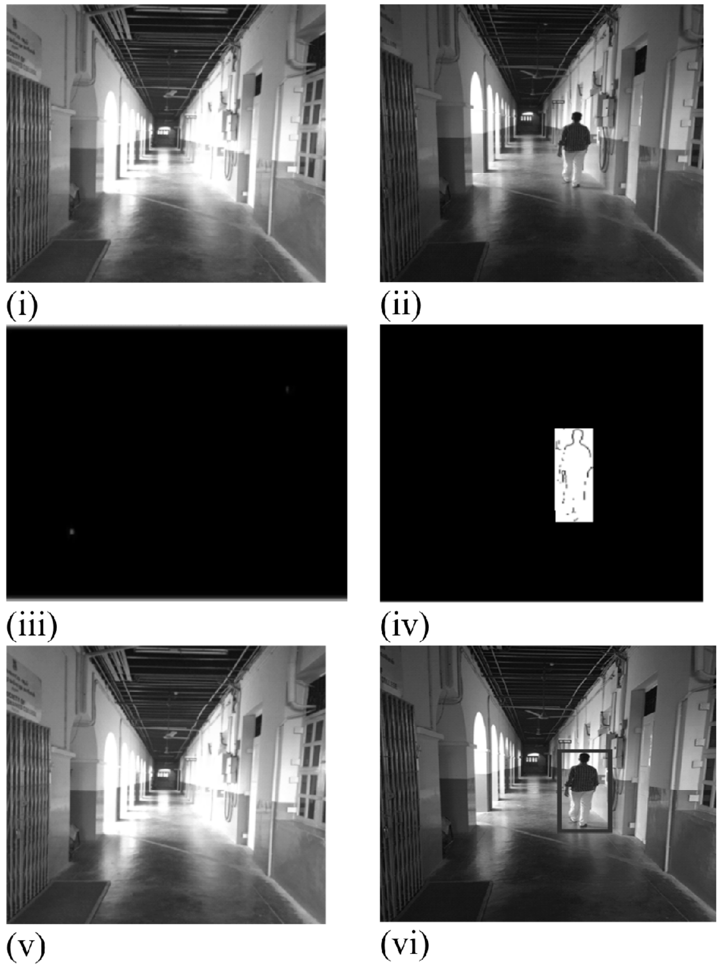

Figure 7: Background subtraction. (i) 7th frame and (ii) 153rd frame of the corridor sequence. Result of the background subtraction for the (iii) 7th frame and (iv) 153rd frame. Result of tracking for the (v) 7th frame and (vi) 153rd frame

The maximum number of reference frames is 5. All the videos are compressed for 196 kbps bit-rate. The true processing time is used to evaluate the performance of the proposed approach, while PSNR [14] is calculated to assess the image quality.

Peak Signal to Noise Ratio (PSNR) shown in Eqn.1 is measured on a logarithmic scale and depends on the mean squared error (MSE) of between an original and an impaired image or video frame, relative to $(2^n - 1)^2$ (the square of the highest-possible signal value in the image, where n is the number of bits per image sample).

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad \textbf{(1)}$$

PSNR can be calculated easily and quickly and is therefore a very popular quality measure, widely used to compare the ‘quality’ of compressed and decompressed video images.

The first row in Figure 8 shows original frames. The second row represents the binary mask of the original video, which is the output of the background subtraction module. The third row shows the reconstructed sequence whose essentially static segments were encoded.

PERFORMANCE RESULTS

Performance of the proposed system has been evaluated using 3 sequences listed in Table2 encoding was performed on the GOP by GOP sizes: 8, 16.

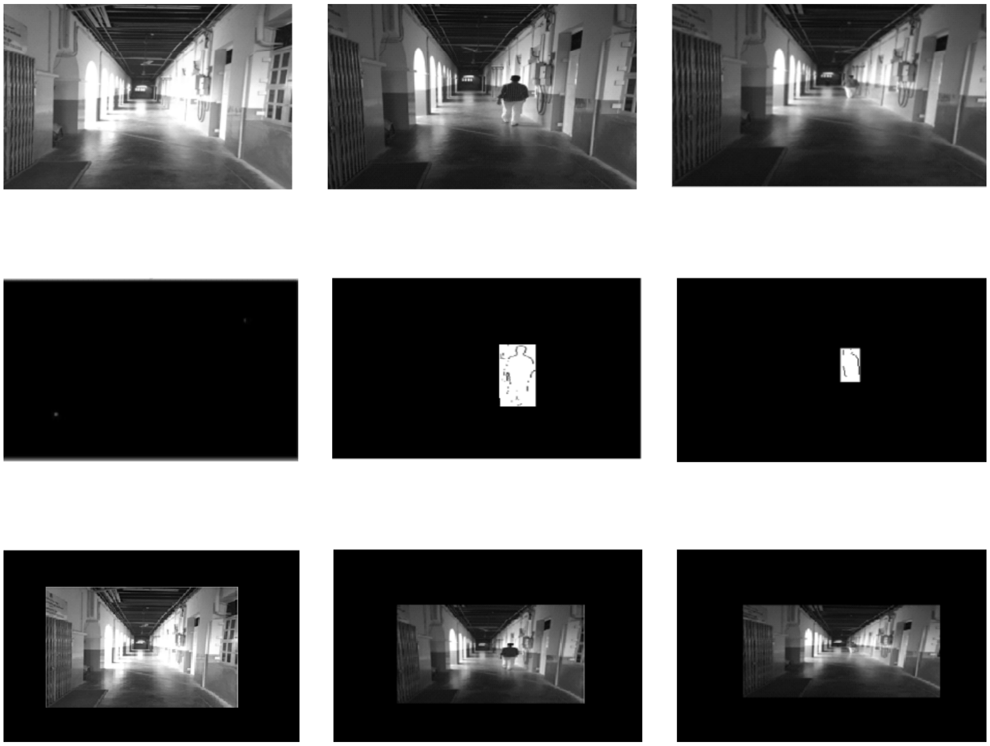

Figure 8: Frames of the reconstructed corridor sequence obtained. 1st row: original frames. 2nd row: binary mask. 3rd row: adapted Sequence

The relative byte saving [2] is calculated as:

$$ByteSaving = (1\text{-nbytes/nbytesConv})\ \text{x}100\% \qquad (2)$$

Table 2
Surveillance Sequences

Sequence	*Time(sec)*	*Frame/sec*	*GOP size =8*		*GOP size =16*	
			Time	*PSNR*	*Time*	*PSNR*
Lane	25	30.00	1734	27.86	2309	30.96
Parking	68	30.88	6080	37.15	8312	38.86
College corridor	10	30.00	787	33.90	1045	36.54

where nbytes represents the number of bytes of the compressed sequence using the proposed system from Figure 6 and nbytesConv represents number of bytes of the compressed sequence encoded at full resolution, frame-rate and quality for the whole length.

4. CONCLUSION

The proposed system is versatile and easy to implement as its main objective is organized using well-known techniques for estimating the amount of information in each captured image to be able to compress the required data from the GOPs and efficient storage which further leads to easy transmission. Future work may be performed on extracting important details especially at places where important and minute details are required for investigation such as noting the vehicle number, the faces of people inside vehicles before accidents or robberies. Transmission of the video data may be done as images which may then be decompressed and decoded into original video by using a table with supporting information such as the arrangement of images and the time of change in activity and so on.

REFERENCES

[1] Zgaljic, Toni, Naeem Ramzan, Muhammad Akram, Ebroul Izquierdo, Rodrigo Caballero, Alan Finn, Hongcheng Wang, and Ziyou Xiong. "Surveillance centric coding." In Visual Information Engineering, 2008. VIE 2008. 5th International Conference on, pp. 835-839. **IET, 2008**.

[2] Ding, Wei, and Gary Marchionini. "A study on video browsing strategies." (**1998**).

[3] Akram, M., N. Ramzan, and E. Izquierdo. "Event based video coding architecture." In Visual Information Engineering, 2008. VIE 2008. 5th International Conference on, pp. 807-812. **IET, 2008**.

[4] Akram, Muhammad, Naeem Ramzan, and Ebroul Izquierdo. "Selective Motion Estimation for Surveillance Videos." In International Conference on User Centric Media, pp. **199-206**. Springer Berlin Heidelberg, 2009.

[5] https://en.wikipedia.org/wiki/Closed-circuit_television.

[6] https://en.wikipedia.org/wiki/Digital_camera.

[7] https://en.wikipedia.org/wiki/Scalable_Video_Coding.

[8] Alam Jahaan ,Dr T.N. Ravi, "A Capsulization Of Image Compression Aspects", International Journal Of Advanced Research In Computer Science And Applications Issn 2321-872x Online Issn 2321-8932, Volume 3, Issue **7**, July 2015

[9] Stauffer, Chris, and W. Eric L. Grimson. "Learning patterns of activity using real-time tracking." IEEE Transactions on pattern analysis and machine intelligence 22, no. **8** (2000): 747-757.

[10] Zaghar, Dhafer R., and Tawfeeq Enad Abdulabbas. "A Compression Algorithm for Video Surveillance System." International Journal of Computer Applications 118, no. **24** (2015).