International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 7, No. 6 (2014), pp. 137-148
http://dx.doi.org/10.14257/ijsip.2014.7.6.12

Object Based Fast Motion Estimation and Compensation Algorithm for Surveillance Video Compression

Gopal Thapa[*1], Kalpana Sharma[2] and M. K. Ghose[2]

[1]*Department of Computer Application*
[2]*Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology*
[*]*gopal_ss@rediffmail.com, kalpanaiitkgp@gmail.com, deanacad.smit@gmail.com*

Abstract

In surveillance systems, the storage requirements for video archival are a major concern because of recording of videos continuously for long periods of time, resulting in large amounts of data. Therefore, it is essential to apply efficient compression techniques for compressing surveillance video. The techniques used for the general video compression may not be the efficient technique for the compression of surveillance video because of the use of static camera as compared to moving camera in general purpose videos. Generally surveillance video consist of multiple objects, smaller in size as compared to the background and they have frequents occlusion with each other. In this paper a new object based motion estimation and compensation technique for surveillance video compression is proposed. Background differencing and summing technique (BDST) is used for the segmentation of the moving objects. This technique not only identifies moving object but also the maximum distance moved by the object in given group of frames. A bonding box is created based on the movement of the object in order to segment the moving objects. For exploiting the temporal redundancy, the motion estimation and compensation is carried out for the bonding box region only. The multiresolution property of discrete wavelet transform is used for the motion estimation and compensation. Experimental results show that the approach achieves high compression ratios compared to MPEG-2 compression.

Keywords: *Surveillance video, video compression, object based motion estimation, background image, Discrete wavelet transform*

1. Introduction

Video has largely turned into a necessary component of today's multimedia applications. Surveillance video is one of the important multimedia applications of today's world. Surveillance videos are widely used in domains such as access control in airports, traffic monitoring, indoor surveillance and human identification [1]. In most of this application camera capture information over long period of time, resulting in a large amount of data. Such a large amount of data require huge amount of memory for storage and large bandwidth for their transmission. Therefore an efficient compression technique is required for the efficient storage and transmission of such huge amount of data.

Most of the present video compression techniques are intended for general purpose video compression i.e. when no assumption about camera motion is made. However, in surveillance application, videos are usually collected using stationary cameras, resulting in a large amount of temporal redundancy due to high inter-frame correlation. Therefore, techniques which are

ISSN: 2005-4254 IJSIP

used for the general video compression may not be the efficient technique for the compression of surveillance video [2].

Two main approaches to video compression are block based and object based coding. The currently used block based approach has the advantages of simplicity, efficient and robustness because of which it has been successfully used and adopted in many video compression standards MPEG 1, MPEG2, H.263, H.264 [3]. In this technique, the use of fixed block partitioning on a fixed grid results in blocking artifacts and unnatural object motion specially at very low bit rates [4]. Object based video coding has been extensively researched in last 10 years and is also supported by MPEG4 standard [5]. This coding approach can achieve efficient compression by separating coherently moving objects from stationary background and compactly representing their shape, motion and the content. Some of the advantages of these approaches include a more natural and accurate rendition of the moving object motion and efficient utilization of available bitrate by focusing on the moving objects. Addition to this advantage, if suitably constructed the object base functionalities such as ability to selectively code, decode and manipulate specific objects in a video stream become possible [6].

In surveillance domain, videos usually consist of objects that are smaller in size as compared to the background. These objects are fast moving and have frequents occlusion with each other [2]. Therefore object based approach is appropriate for the compression of surveillance video. In this paper a new object based motion estimation and compensation technique for surveillance video compression has been proposed.

The paper is structured as follows. In Section II, moving object identification and segmentation is described. Section III describes the algorithm for object based fast motion estimation and compensation technique for the video compression. Experimental results are described in Section IV and finally conclusion is presented in Section V.

2. Moving Object Detection and Segmentation

The proposed compression technique is divided into two modules. First is the moving object segmentation and second is the object motion estimation, compensation and coding. A common method for extracting moving objects in video sequence is background subtraction. To detect the moving objects in video frames, initially, the model of background scene must be constructed (*i.e.*, the image without the moving objects), then current frame is subtracted from the background model and eventually, the difference, determines the moving objects. This background subtraction effectively extracts the correct shape of moving objects [7]. Therefore it is proposed to use the background subtraction technique for extracting the moving objects from the video signal. Here it is assumed that the background image has been constructed using any of the methods used in [8, 9, 10]. For segmenting moving object, a new background differencing and summing technique (BDST) has been proposed.

Firstly first eight consecutive frames from a video signal are considered and each of them is subtracted from the background image. The difference images are them added together to obtain a resultant image. The resultant image represents the maximum movement of the objects in the given eight frames. Let $f_i(x,y)\ where\ i\ varies\ from\ 1\ to\ 8$, be the sequence of frame from a video signal then the equation 1 represent the difference of the image sequence with the background image f(x,y) and equation 2 represent the sum of difference image.

$$O_i(x,y) = f(x,y) - f_i(x,y) \qquad (1)$$

$$O(x,y) = \sum_{i=1}^{8} O_i(x,y) \qquad (2)$$

Here $O_i(x, y)$ represent the moving object in each frame. The summation of the entire difference image results in an image which shows the maximum displacement of an object in the given sequence of frames. The block diagram of this proposed method is shown in Figure 1. The noise from the resultant image is removed using morphological operations. The final image is a black and white image that consists of black background with the objects along with its displacement denoted by white connected pixels.

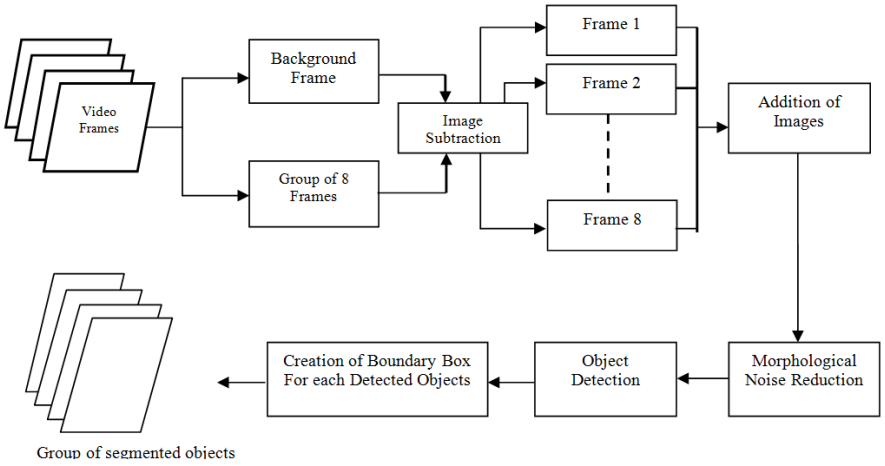

Figure 1. Block diagram for the segmentation of moving objects from the frames

Now in order to segment the identified moving object, pixel labeling technique based on the 8 neighbors connectivity is used. The labeling procedure starts by scanning the individual pixel. Scanning is done one row at a time starting from the top left corner of the image i.e. from left to right of an image. Once the scan reaches the second row the scanning is done in the reverse manner i.e. from right to left, and so on till all the rows of the image matrix are scanned. This means the odd rows are scanned from left to right and even rows are scanned from right to left. As the image is in the binary format the object is denoted by connected pixels with value 1 and the remaining background is denoted by 0. During scanning labeling the pixels will not start until the first '1' is encountered. Once the first '1' is encountered it is assumed that the first object is found and now the labeling process can be started. Let *p* denote the encountered pixel during the initial scan when no object has been detected, if *p*=1, the pixel *p* is labeled with a new label followed by the labeling of its entire eight neighborhood pixels with the same label as that of *p*. This labeling is done to differentiate the different objects. After the first pixel labeling it is know that the object has been detected but the area of the object is yet to be known. To know the exact dimension of the object, pixels on the same row are scanned, if the pixel *p*=1 again the same labeling process is followed.

If the next pixel p=0 and if any of the surrounding eight neighboring pixels contains a white pixel (p=1), then the same procedure is followed, *i.e.*, eight neighborhood of 0 is scanned and the label that is assigned to any one of its eight neighborhoods is assigned to all its non-labeled neighborhood pixels. If in case the eight neighborhood of 0 pixel contains only black pixels (p=0) whether labeled or non- labeled, skip the current pixel and move scanning to the next pixel. By doing this we skip labeling the black pixels which may be the empty spaces inside the frame and which is not the part of the object itself.

A condition may arise that if the current pixel is a black pixel or p=0 and the surrounding eight neighborhood pixel may contain combination of white and black pixels, if these pixels are not labeled than labeling is simply avoided, else if any one of the 8 neighbors is 1 and its labeled than as mentioned above the same level is assign to all the surrounding eight neighborhood and the current pixel itself. This condition applies only if the current pixel scanned is 0.

All the connected pixel of the first detected object is labeled as 'v' as shown in the Figure 2 as per the proposed scanning process. Here the labeling stops at fifth row first column where the current pixel value is p=0. Here since all the neighbors of the p is 0, and as stated above for this condition the labeling of the pixel is skipped. Since all the remaining pixel fall under same condition the labelling process is skipped.

0	0(A)	0(A)	1(A)	1(A)	0(A)	0(A)	0(A)	0
0	0(A)	0(A)	0(A)	1(A)	1(A)	0(A)	0(A)	0
0	0(A)	0(A)	1(A)	1(A)	0(A)	0(A)	0(A)	0
0	0(A)	0(A)	0(A)	0(A)	0(A)	0(A)	0(A)	0
0	0(A)	0(A)	0(A)	0(A)	0(A)	0(A)	0(A)	0
0	0	0	0	0	0	0	0	0
0(B)	0(B)	0(B)	0(B)	0(B)	0(B)	0	0	0
0(B)	0(B)	1(B)	1(B)	1(B)	0(B)	0(B)	0(B)	0
0(B)	0(B)	0(B)	1(B)	0(B)	1(B)	0(B)	0(B)	0
0(B)	1(B)	1(B)	0(B)	0(B)	0(B)	0(B)	0(B)	0

Figure 2. Pixel Labeling Process

Once the labeling is complete, all the connected pixels of same object will have the same label assigned. We neglect the labels whose pixel count is less than 400 pixels because those labeled pixels can be treated as noise. At last based on the coordinates of the maximum and minimum horizontal and vertical coordinates of the object, a bonding box is created on the entire eight frames to segment the moving objects. This also defines the maximum displacement of the moving object.

3. Object based motion estimation and compensation

In [11] objet based coding using pixel state analysis has been proposed. This results in bit saving only for certain class of video sequence only and because of overhead in coding pixel state information there is high computation complexity. In [12], the moving object is segmented from the background by using image attributes-object pixels, object edge map and object difference edge map and the motion estimation is performed within the edge map of the object. However determining edges for each and every object is quite complex as the object keep on changing it shape in the video signal. In this section we propose a simple but robust object based motion estimation and compensation technique in order to exploit the temporal redundancy. Once the object is segmented from the frames and bonding box is created. The bonding box indicates the maximum movement of the object in the eight consecutive frames. Therefore motion estimation is performed only on the bonding box region.

The discrete wavelet transform has received considerable attention in the field of image and video processing. This is because of its flexibility in representing nonstationary signals and its ability in adapting to human visual characteristics [13]. It perfectly provides a multiresolution representation of a signal with localization in both time and frequency which is desired property in image and video coding [14]. A wavelet transform corresponds to two sets of analysis/synthesis digital filters, a high pass filter and low-pass filter. Thus wavelet transforms deco-relate the pixel values in video frame and result in frequency and spatial orientation separation. The two dimensional discrete wavelet transform of image f(x,y) with resolution depth M can be represented as a sequence of subimages

$$\{S_M f, [W_M^j f]_{j=1,2,3} \ldots . [W_1^j f]_{j=1,2,3} \qquad (3)$$

The sequence $\{S_m f : m = 1 \ldots . M\}$ represent the approximation of a given video frame at different resolution and $[W_M^j f]_{j=1,2,3}$ represent three detail subimages at various orientation known as horizontal, vertical and diagonal subimages. The Figure 3 depicts pyramid structure of 3 level wavelet decomposition of an image f(x,y). The pyramid consists of a total of 10 subimages with three subimages at each layer and one low pass subimages at the top. Each subimage of frames also represents the global motion structure of the particular video signal at different scales.

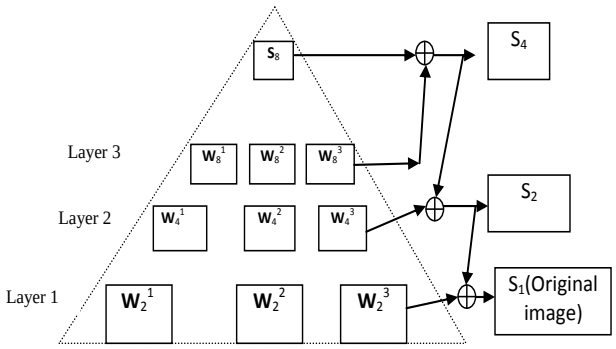

Figure 3. Pyramid Structure of 3 level decomposition of an image

The motion activities for particular subimages at different resolution may be different but are highly correlated. The hierarchical search for the motion activities result in fast estimation of motion vectors with reduced computational complexity [15]. Therefore, in this paper, the hierarchical multiresolution motion estimation technique used in [13] has been used for the estimation of motion vector and compensation. The inverse wavelet transform is calculated in the reverse manner, *i.e.*, starting from the lowest resolution subimage, the higher resolution images are calculated recursively.

Once the object is segmented from the frames and bonding box is created. Three level discrete wavelet transform is applied to the current frame and the reference frame objects for the motion estimation and compensation After the motion compensation, the error image is obtained which is then encoded along with the motion vectors using Huffman coding. The block diagram of the proposed algorithm is shown in Figure 4 .

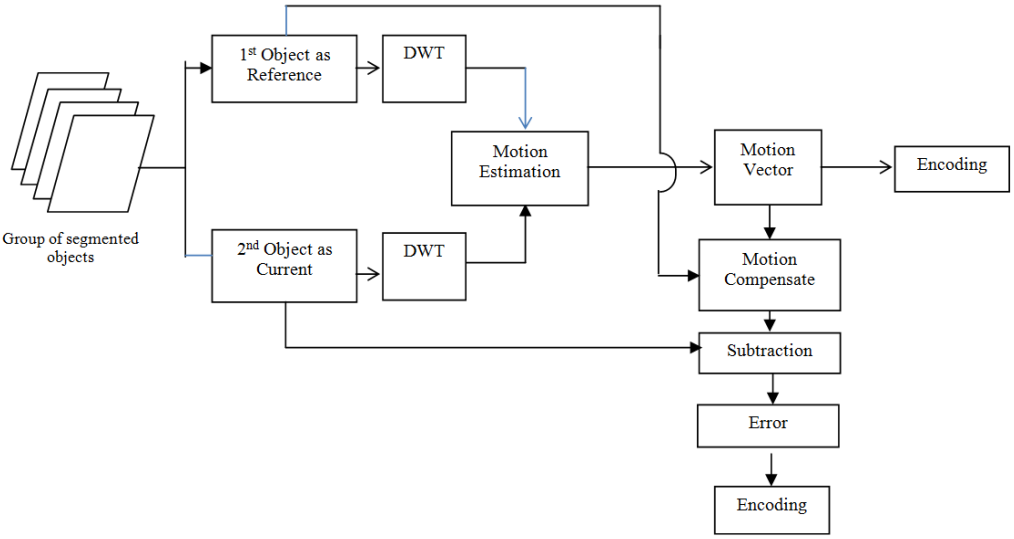

Figure 4. Block diagram for object based motion estimation and compensation

Mean absolute difference is used as the measure for motion estimation performance. If the size of a video frame is X x Y, the MAD is given by

$$MAD = \frac{1}{XY}\sum_{i=O}^{X-1}\sum_{j=0}^{Y-1}\left[W(i,j) - \widehat{W}(i,j)\right] \qquad (4)$$

where W(i,j) and $\widehat{W}(i,j)$ respectively represent the pixel value of the original frame and the reverence frame. In our experiment, a block size of 4 pixels and search parameter (SP) of 2 pixels and MAD=10 has been used for the block matching. For simplicity, two frames are considered for implementation of the proposed algorithm. The above mentioned algorithm for motion estimation and compensation is summarized as follows:

STEP 1: Segment the moving objects using algorithm mentioned in Section 2.

STEP 2: Calculate the height and width of that object, and find the total no. of macro blocks of that object.

STEP 3: For block: =1 to total blocks in Frame 1, do

3.1: MAD_min: = 10

3.2: Search for matching block in Frame 2 within the search area, i.e. For i: = -SP to SP, increment by 1, do

For j: = -SP to SP increment by 1, do

Calculate MAD;

If MAD<MAD_min

MAD_min: = MAD

[End if]

[End For j]

[End For i]

3.3: Store (i , j) of the block from Frame 2as motion vector whose MAD is minimum.

3.4: If no block match found within the search area then intra code that block.
[End For block]

STEP 4: Compensate the object using the stored motion vectors and the object from Frame 1.

STEP 5: Compute, Error: = Compensated Object – Original Object

STEP 6: Apply quantization on the Error image.

STEP 7: Huffman encoding of Motion Vectors and Error.

4. Experimental results

To test the proposed algorithm, three video namely 'hall_monitor.avi', 'outdoor.avi' and 'road_side.avi' has been used. Figure 5 shows the background and first eight consecutive frame for 'hall_monitor.avi' and the result of subtraction of each of the eight frame of the video signal from the background frame. The first frame of the video is chosen as the background image as it do not contain any moving object or it can be obtained by any method used in [8, 9, 10]. Figure 6(a) is the resultant frame obtained after the addition of all the eight difference frames after the removal of noise. This frame shows the maximum displacement of the moving object between first and the eighth frame. With the help of the maximum and minimum coordinated of this moving object a bonding box is created on each and every object in all the eight frames. Therefore this represent region with the movement of the object and hence will be used for the motion estimation between the frames. For the motion estimation and compensation, only the region within the bonding box is used. Figure 6(b) shows the segmented object with the bonding box. For the motion estimation we used a bock size of 4 x 4 and search parameter of 2 pixels. Figure 9 shows the motions compensated objects with the black spots. The black spots are generated because of the non matching of the block and these non-matched blocks are intra coded. Finally, the motion compensated frame is shown in Figure 10(b).

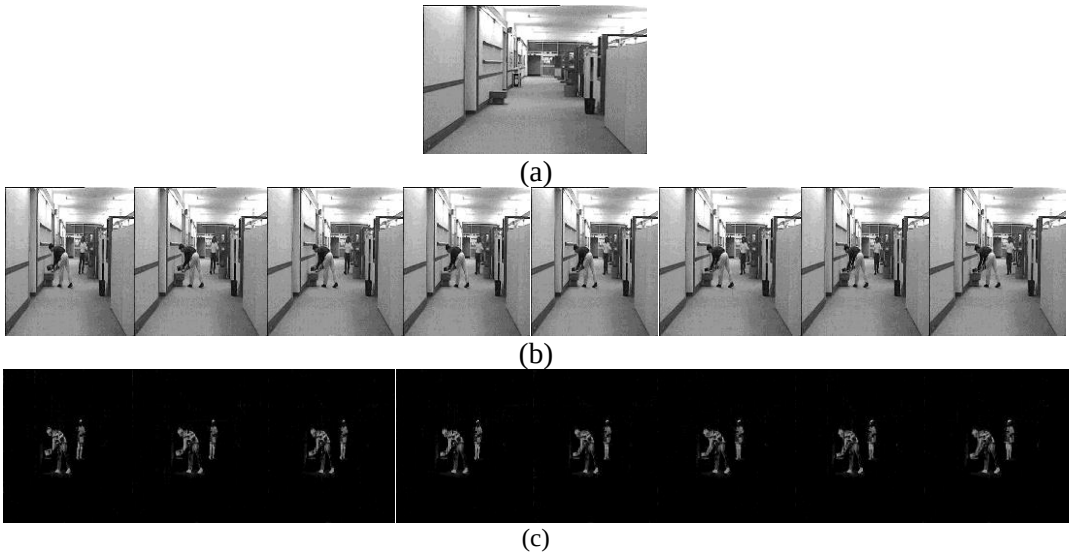

Figure 5. (a). Background frame; (b). Group of eight consecutive frames; (c). Difference of frames and the eight frames

(a) (b)

Figure 6. (a) Image showing object's maximum displacement within 8 frames; (b) Creation of rectangular bounding boxes covering the region of maximum displacement by the object

Figure 7. Two objects being segmented from the image for motion estimation

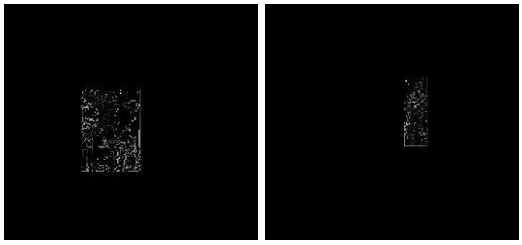

Figure 9. Images showing the errors for each object

(a) (b)

Figure 10. (a) Original frame; (b)Corresponding Motion

Similarly the results for other test videos "outdoor.avi" and "road_side.avi" are shown in Figure 11 and Figure 12.

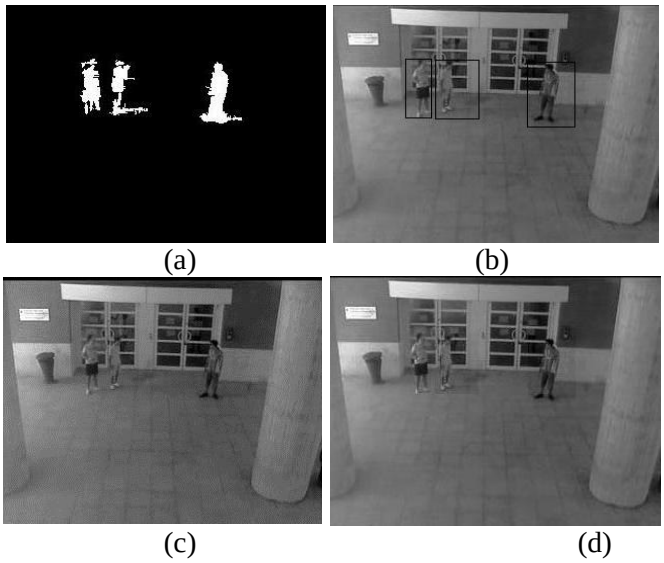

Figure 11. (a) Object's maximum displacement within 8 consecutive frames; (b) Creation of rectangular bounding boxes covering the region of maximum displacement by the object; (c) Original 4th frame; (d) Motion compensated 4th frame

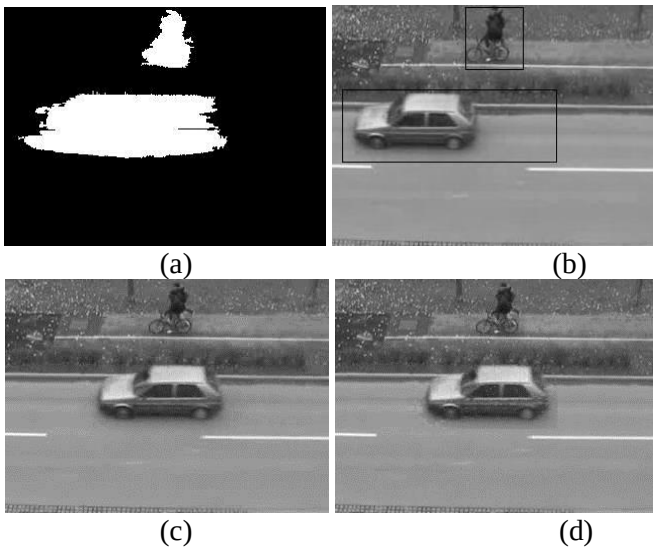

Figure 12. (a) Object's maximum displacement within 8 consecutive frames; (b) Creation of rectangular bounding boxes covering the region of maximum displacement by the object; (c) Original 4th frame; (d) Motion compensated 4th frame

The quality of the reconstructed frames is measured in terms of peak signal to noise ratio (PSNR). The variation of PSNR values for the reconstructed frames of three video sequence is shown in the Figure 13.

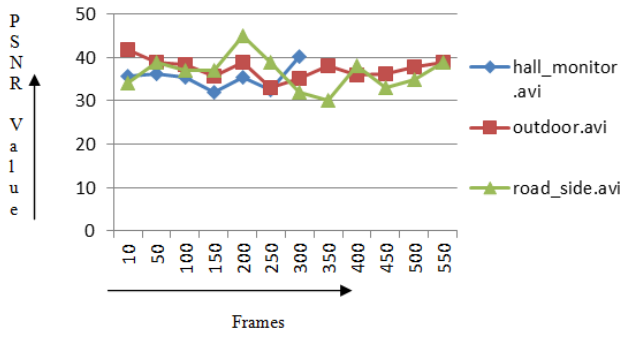

Figure 13. Variations of PSNR values

The average PSNR for each of the video is given in Table 1.

Table 1. Comparisons PSNR value: Proposed object base coding vs MPEG-2

Sl. No.	Video Name	PSNR Value	
		MPEG-2	Object based
1	hall_monitor.avi	32.24	35.32
2	outdoor.avi	33.46	37.62
3	road_side.avi	33.28	37.16

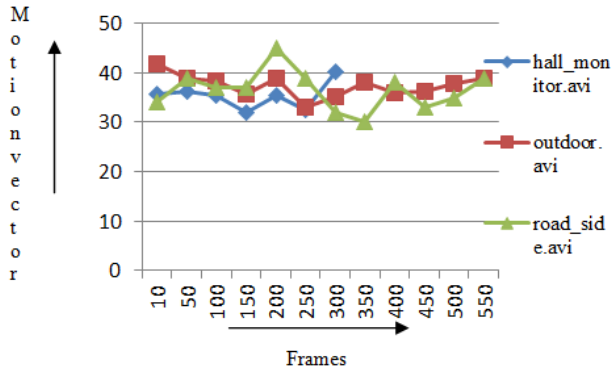

Figure 14. Number of motion vectors per frame

The number of motion vectors per frame in all test videos are shown in Figure 14. It can be seen that the variation in PSNR value as show in Figure 13 is because of the variation in motion vectors. Video 'roadside.avi' has the highest PSNR variation as it has the highest variation of motion vectors. This shows that with the variation in motion there is slight variation in the quality of the reconstructed signal. That is why a group of eight frames are considered for the proposed algorithm so that the movement of the object is as small as possible. The PSNR obtained by the proposed method is compared with MPEG-2 and the result is shown in Table 1. From the table it is clear that with the proposed method there is

improvement in the quality of the video signal. Similarly, we compared the compression ratio for three test video and are shown in Table 2.

Table 2. Comparison of compression ratio between proposed algorithm and MPEG-2 algorithm

Sl. No	File Name	No. Of Frames	Size (MB)	MPEG-2 (MB)	Object based (MB)	Compression factor
1	hall_monitor.avi	300	31.8	5.2	3.27	1.59
2	outdoor.avi	820	60	7.25	4.75	1.52
3	road_side.avi	550	40.3	10.7	6.2	1.72

From the table it is evident that high compression can be achieved by using object based coding technique. The variation in the compression ratio is because of the variation of motion vector, *i.e.*, higher the motion in the frame less is the compression ratio.

5. Conclusion

In this work, we have described an object based compression technique for surveillance video. We have proposed a new moving object identification algorithm based on differencing and summing technique. The motion estimation and compensation has been applied only on the moving part of the frame. The performance of the proposed compression technique is compared with that of the standard MPEG-2 compression technique. The quality of the reconstructed video signal in terms of PSNR value is better than that of MPEG-2 by a factor of 1.1 and the compression ration higher than the MPEG-2 compression technique by a factor of 1.5.

References

[1] R. Girisha and S. Murali, "Object Segmentation from Surveillance Video Sequence", IEEE International Conference on Integrated Intelligent Computing, **(2010)**, pp. 146-152.
[2] A. Hakeem, K. Shafique and M. Shah, "An Object-based Video coding framework for Video Sequences Obtained From Static Cameras", pp. 608-617.
[3] A. Vetro, T. Haga, K. Sumi and H. Sun, "Object-based coding for long-term archive of surveillance video", in IEEE International Conf. on Multimedia and Expo, vol. 2, **(2003)** July, pp. 417-420.
[4] C. Kim and J. -N. Hwang, "Object-based video abstraction for for video surveillance system", IEEE Trans. Circuits and Systems for VideoTechnology, vol. 12, no. 12, **(2002)** December, pp. 1128-1138.
[5] K. Belloulate, A. Belalia and S. Zhu, "Object Based Stereo Video Compression Using Fractals and Shape Adaptive DCT", International Journal of Electronics and Communication, vol. 68, **(2014)** July, pp. 686-697.
[6] R. Talluri and K. Oehler, "A Robust Scalable, Object-Based Video Compression Techniuqe For Very Low Bitrate coding", Circuits and Systems for Video Technology, vol. 7, **(1996)**, pp. 221-233.
[7] P. Spagnolo babu, T. D. Orazio, M. Leo and A. Distante, "Moving Object Segmentation by background Subtraction and Temporal Analysis", ELSEVIER, Image and Vision Computing, vol. 24, **(2006)**, pp. 418-425.
[8] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proceedings of the IEEE, vol. 90, no. 7**, (2002)** July, pp. 1151-1163.
[9] C. Stauffer and W. E. L Grimson, "Adaptive background mixture models for real-time tracking", IEEE International Conference on Computer Vision, **(1999)**, pp.255-261.

[10] J. Gallego *et al.*, "Segmentation and tracking of static and moving objects in video surveillance scenarios", IEEE ICIP, **(2008)**, pp. 2716-2719.
[11] T. Nishi and H. Fujiyoshi, "Object-based Video Coding Using Pixel State Analysis", IEEE Int. Conference on Pattern Recognition, **(2004)**, pp. 306-309.
[12] R. V. Babu and A. Makur, "Object –based Surveillance /video Compression using Foreground Motion Compensation", ICARCV, **(2006)**, pp. 1-6.
[13] J. Zan, M. O. Ahmad and M. N. S Swamy, "Wavelet Based Multiresolution Motion Estimation Through Median Filtering", Acoustics, Speech and Signal processing, vol. 4, **(2002)**, pp. 3273-3276.
[14] S. Zafar and Y. –Z. Zhang, "Multiresolution Video Representation using Multiresolution Motion Compensation and Wavelet Decomposition", IEEE Journal of selected area in Communication, vol. 11, **(1993)** January, pp. 24-35.
[15] E. Mohammed, Al-Mualla, C. N. Canagarajah and D. R. Bull, "Reduced Complexity Motion estimation Techniques: Review and comparative Study", Electronics, Circuits and Systems, vol. 2, **(2003)**, pp. 607-610.

Authors

Gopal Thapa: He received M.SC (Electronics) and M.Tech(IT) from Sikkim Manipal University. He is currently pursuing PhD in Video compression using Discrete Wavelet Transform. His areas of interest include image, video processing and Steganography.

Dr. Kalpana Sharma: She is Professor of the Department of Computer Science & Engineering at Sikkim Manipal. Institute of Technology, Majitar, Sikkim, India since August, 1998. She did her BE from National Institute of Technology, Silchar, India and M.Tech from IIT Kharagpur, India. She completed her PhD in the field of Wireless Sensor Network Security from Sikkim Manipal University in 2012. Her areas of research interest are Wireless Sensor Networks, Steganography, Network & Information Security, Real Time Systems and image processing. She has published a number of technical papers in various national and international journals in addition to presentation/ publication in several international/ national conferences.

M. K. Ghose: A former Senior Scientist of Indian Space Research Organization, Department of Space, Govt of India, joined as Prof. and Head, Dept. of Computer Science and Engineering, SMIT, Sikkim, India in June 2006. He is currently the Dean (Academic). His areas of research interest are Data Mining, Simulation & Modeling, Network, Sensor Network, Information Security, Optimization & Genetic Algorithm, Digital Image processing, Remote Sensing & GIS and Software Engineering.

