A NOVEL APPROACH BASED ON 2D - DWT AND VARIANCE METHOD FOR HUMAN DETECTION AND TRACKING IN VIDEO SURVEILLANCE APPLICATIONS

[1]V. R. Satpute, [2]Dr. K. D. Kulat, [2]Dr. A. G. Keskar

[1]Assistant Professor, [2]Professor, Department of ECE,

Visvesvaraya National institute of Technology, Nagpur (India).

ABSTRACT

Moving object detection is very important in modern world for fast video surveillance. There are various methods used for detecting moving objects out of which frame differencing method is widely used and is most efficient method. In this paper we focus on the surveillance at the most secured areas such as airports, defense establishments, power stations etc. Similarly, the area where no human is allowed without authority to enter such as bank locker rooms, restricted military area etc. plays a vital role. In real time surveillance system, storing the captured video and detecting object are two most important issues. Storing such videos needs more memory and the detection of the object is also need to be fast. To solve these problems compression and fast object detection is required. To detect the moving object, detection of its edges and location in the frame are important steps. In this paper we propose a mechanism to use discrete wavelet transform (DWT) for purpose of edge detection, whereas to locate the object we propose variance method on to the 2-D DWT outputs of video frames. For this analysis HAAR wavelet is used as reference due to its easy of implementation and having inherent properties.

***Keywords*: DWT, Variance, Object Detection, Tracking, Frame Differencing, Haar Wavelet.**

I. INTRODUCTION

Intelligent and automated surveillance system has become an essential area of research in these days as its demand has increased rapidly for surveillance at airports, railways stations, power stations etc. The availability of high quality video cameras and increasing need of automated video surveillance has generated great interest in the areas like motion detection, object detection, tracking and analysis [14,15]. So the possible major steps in video surveillance are detection of interesting moving objects, tracking of detected objects in the video sequences and analysis of these tracked objects. In some conventional objects detection approaches such as optical flow algorithm or statistical algorithm, computational complexity and the time are very high which is not suitable for live video surveillance applications. So the frame differencing algorithm or background subtraction algorithms are much suitable for such a kind of application. Background subtraction technique is a most suitable choice to detect stationary foreground objects as it works good when camera is stationary as well as change in ambient lighting is also very gradual. Hence it is a most popular technique to separate out foreground i.e. object of interest in video (frames). In video surveillance frame size of the video is maintained higher for the sec of clear visualization of security personal. This requires large amount of storage memory. The major hurdle in video surveillance is resolution of video frames and storage of such videos. As we know that when one goes for more resolution then for storing one need more memory. The next hurdle is the time required for processing for such videos for object detection and tracking. It needs very large time to do such operations. To solve these problems, in this paper, the variance based object detection is proposed and is compared with existing well known mean shift method. Discrete wavelet transform (DWT) is used for solving the problem of edge detection of the object and finally variance based method for the moving object detection and tracking is used. Further some morphological operations are used for filtering out some noises available in the video. Variance method can be used iteratively to get more accuracy.

II. IMAGE DATA AQUISITION

The input videos are taken in the area where persons are not allowed without prior permission of the concern authority in the thermal power station area. These videos are acquired from static video surveillance camera mounted on wall (about 12 meters above the surface) for surveillance purpose. The camera (Nikon Coolpix s3300) was set on 6.0 mega pixels mode with frame size of 640-by-480 and having automatic gain and exposure control.

III. DETECTION FLOW

In this paper, an algorithm in which at first whole video is applied with Discrete Wavelet Transform (DWT) to the input video. Here we have used only level one DWT. Now except LL part, remaining three parts i.e. LH, HL and HH of the Discrete Wavelet Transform are used for object detection thereof is proposed. The HH, HL and LH parts are nothing but the high pass outputs of the transform are added together which results in a form in which most of the image pixels becomes dark except the object which reduces errors in final detecting process and also reduces processing time. Static background of whole video is subtracted from these resulted frames of whole video. This process is well known as background subtraction. While in this the previous frame is also subtracted from the current frame. This process is called frame differencing. These resulting frames are further binarized. The resulted frames still contain some unwanted noise or pixels other than moving object. These appear because of non standard lighting conditions in the input video. To reduce these noise morphological operations were carried out. The last but most important step to find the exact moving object area in spatial domain is done by calculating variance of these frames. Furthermore one iterative method is used to find centre of the object. After completion of process bounding box is placed around the object which indicates the presence of the moving object in the video and helps in tracking it. The above steps are summarized in a flow chart. (Refer figure – 1).

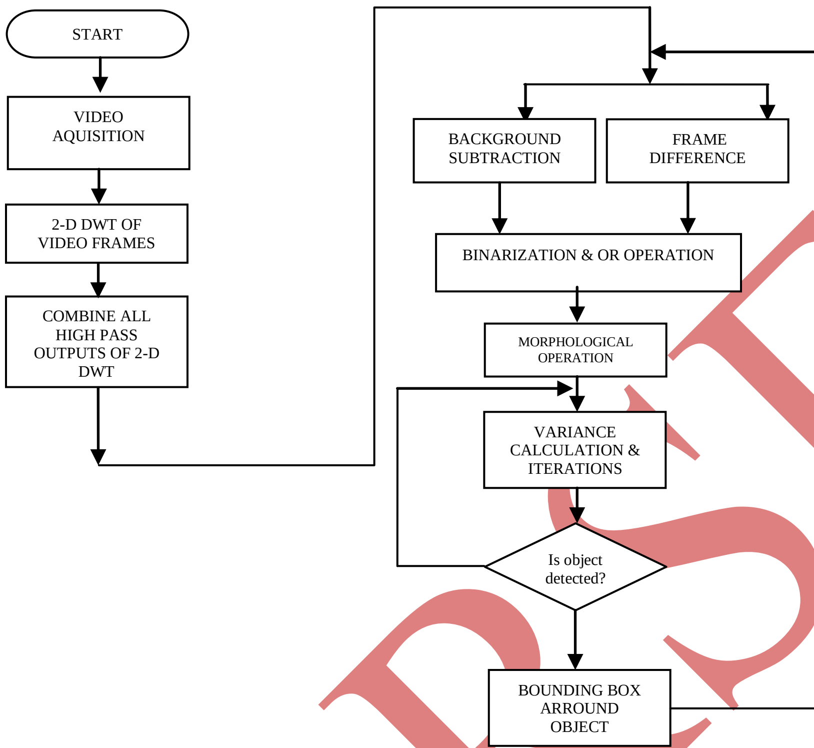


Fig1. Flow Chart of Algorithm.

IV. DISCRETE WAVELET TRANSFORM

Discrete wavelet transform (DWT) transforms discrete time signal to a discrete wavelet representation. It converts an input series X_0, X_1,… X_m into one high pass wavelet coefficient series and one low pass coefficient series[3]. This can be represented by

$$Hi = \sum_{m=0}^{k-1} X(2i-m).Sm(z) \quad (1)$$

$$Li = \sum_{m=0}^{k-1} X(2i-m).tm(z) \quad (2)$$

Where $S_m(Z)$ and $T_m(Z)$ are called wavelet filters, k is length of the filter, and i=0,….,[n/2]-1.[3] Practically on any image in spatial domain discrete wavelet transform is applied in two direction 1st level 2-D DWT, results in four parts called LL, LH, HL & HH [9]. Finally input image is decomposed into four components as shown as in fig 2. Then LL part can be used for further decomposing or higher level of 2D-DWT. The result of 2D DWT on one of the frame taken from video database is shown in figure 3.

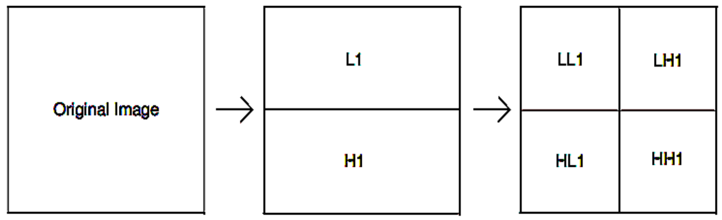


Fig. 2 Application of Discrete wavelet transform process for an image

High pass components i.e. LH, HL, HH are spatially added together forming a sub frame for object detection process. It is nothing but the sudden change in the intensity hence treated as edges of various objects present in the original frame, which can work similar to image segmentation [12]. For equation (3) i=1 to m/4 and j=1 to n/4.

$$Sub-frame(i,j) = LH(i,j) + HL(i,j) + HH(i,j) \qquad (3)$$

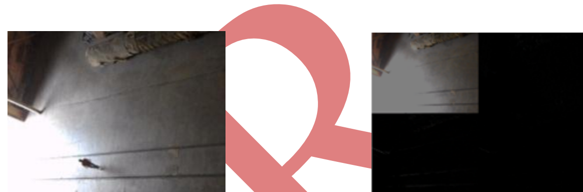


Fig. 3 (a) original frame (b) result of 2-D DWT of original frame.

(a) (b) (c) (d)

Fig. 4 (a)HL (b)LH (c)HH components (d)combination of a, b and c.

Figure 4 (a), (b) & (c) are LH, HL and HH parts of the frame referred in the figure 3, while figure 4(d) shows the result of combination of these parts. In this, the area or the edges of the input frame are present because as LH,HL and HH parts itself contains the pixels of the frame input frame which are passed from high pass filters which gives edge intensities.

V. BACKGROUND SUBSTRACTION AND FRAME DIFFERNECING

These are the two methods widely used in object detection mechanism. When the camera position is said to be steady and no object is moving then the frame is treated as reference or background frame. This frame is subtracted from each one of frames in the video to get unwanted objects or the objects other than available in the reference frame [2] [4]. This method is vastly used when steady objects are to be detected and is called as background subtraction (BS). On the other hand each frame is subtracted from its successive frame is called as frame differencing (FD). Object pixels which are moving are detected in this method.

$$\{BS(i,j)\}_k = \{F(i,j)\}_k - \{F(i,j)\}_{k=0} \qquad (4)$$

$$\{FD(i,j)\}_k = \{F(i,j)\}_k - \{F(i,j)\}_{k-1} \qquad (5)$$

Where i = 1,…,m and j = 1,….,n, k = 1,2,….,p; p is number of frames in the video

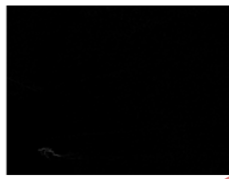


Fig. 5 Resultant frame: no 194.

VI. BINARIZATION (THRESHOLDING)

Binarization is the process in which the gray scale frame is converted into black and white form. i.e. gray scale image consists of 256 levels of intensities which are converted into equivalent 2 intensities, either dark or white. This conversion is done by taking one threshold level of the gray scale levels of image. If pixel intensity is above the threshold value then pixel value becomes 1 or white and if pixel value is less than threshold value then pixel is converted into dark or black i.e. 0. Selection of threshold value for binarization is very crucial process in image processing field. Value of the threshold may change from frame to frame. This occurs because average frame intensities changes frame to frame. It can affect detection process also. Hence threshold is calculated for each frame of resultant. To calculate global threshold for the particular resultant frame some parameters are calculated like mean and standard deviation. [1]

$$u = \frac{1}{r*c}\sum_{i=1}^{r}\sum_{j=1}^{c} FD(i,j), \qquad (6)$$

$$\sigma = \sqrt{\frac{1}{r*c}\sum_{i=1}^{r}\sum_{j=1}^{c}(FD(i,j) - u)^2} \qquad (7)$$

$$T = 0.06 * \sigma \qquad (8)$$

In which FD (i, j) is difference of pixels (i, j) between two successive frames, and [r, c] is size of frames. T is value of threshold used for the current frame only. The binarized form of differenced and background subtracted frames are undergone through OR operation so as to achieve object detection while moving or when objects stops suddenly.

$$\{RV(i,j)\}_k = \{BS(i,j)\}_k \; OR \; \{FD(i,j)\}_k \qquad (9)$$

Where RV is resultant video. Figure 6 shows frame from RV corresponding to figure 5.

Fig.6: Binarized frame

Fig. 7 Resultant frame after morphological operations.

VII. MORPHOLOGICAL OPERATIONS

Morphological operations are used for obtaining noise corrections in the binarized frame. No thresholding method is ideal yet enough which will avoid all noise pixels in binary frame as shown in fig 6. So to eliminate some noise observed in binarized frame opening and closing operations are used [3]. Opening operation removes noise pixels along with it some part of object pixels may also get removed. Hence to retrieve them closing operation is carried out. The resultant frame of figure 6 is shown in figure 7.

VIII. OBJECT DETECTION

Morphological operations eliminate small holes and fills gaps in object contour. Here we propose new idea of obtaining the object area by calculating variance of rows and columns of the frame [8]. This operation gives us variation in the image intensity i.e. non-zero value of variance at locations where moving object is getting detected. The variance is calculated based on the formula i.e. [10].

$$\sigma^2 = \frac{1}{n^2}\sum_{x-1}^{m}\sum_{y-1}^{n} I^2{}_{sub}(x,y) \qquad (10)$$

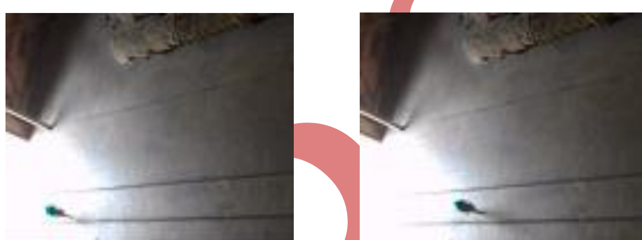
Where, I_{sub} is the sub image under considerations and n is the number of pixels in the sub – image. By taking intersection point of row and column containing maximum variance in the frame a window is formed around it. Size of the window is decided by finding out presiding and succeeding non zero value of variance to the selected row and column. Once again the same process is carried out in the window. New pixel point is obtained. Window is shifted around the new point. This process is carried out till it converges to a point. Finally bounding box is placed around the object to show the location of the object in the input frame for the security personal to get some alert as well as it will be helpful for tracking of the object in the video.

IX. EXPERIMENTAL RESULT

The system has been implemented using MATLAB and tests were executed on an Intel(TM) i3 processor with 2.8GHz clock frequency and 3GB of RAM. The system was tested on various videos downloaded from internet and captured at different places under non-standard conditions. The details about the videos are given in table I. Figure 8 shows moving object detected and framed by a blue square box which will alert the security person about the movement as well as keeps an automatic track of the moving object in the video.

TABLE I. VIDEO DETAILS

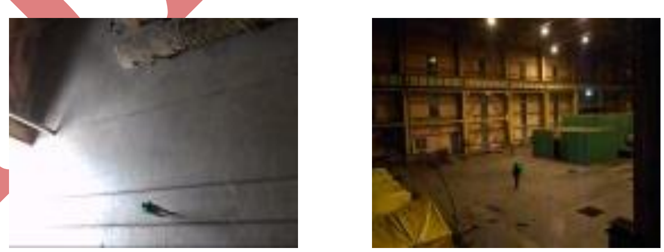
	Frame Size	Total Frames	Duration (in sec)
menearbfp.avi	480×640	543	49
Menearbfp1.avi	480×640	581	51
Magnet.avi	240×320	539	41
Turbfloor1.avi	480×640	983	64
Motinas_emilio_webcam.avi	240×320	448	38
Motinas_room105_audioviisual.avi	288×360	1077	98
Motinas_room160_audioviisual.avi	288×360	1073	102
Walk1.mpg	384×288	610	24
Walk2.mpg	384×288	1054	42
Walk3.mpg	384×288	1178	55
Browse1.mpg	384×288	1040	41



(a) (b)

Fig. 8 Captured moving object (a) 194th frame (b) 300th frame.

Figure 9(a) shows 240th frame of menearbfp1 video taken at the same place which also shows the detected object and a blue colored bounding box around the moving object. Whereas, figure 9(b) 500th frame of turbfloor1 video at another location showing the moving object in the green colored bounding box. Figure 10 and 11 shows the frames from the video database selected from [6,7] in which the moving object is also detected. Figure 13(a) and figure 13(b) are the results of the algorithm applied on the video frames of the video database selected from [8].



(a) (b)

Fig. 9 (a) 240th frame of menearbfp1 video (b) 500th frame of turbfloor video.

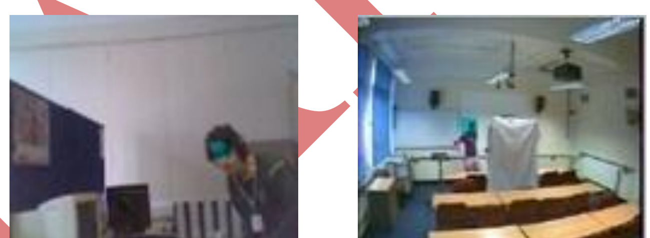
(a) (b)

Fig. 10 (a) 380th frame of walk1 (b) 360th frame of walk2.

For the detection of object and to get more denser area of its location there exist a method called mean shift method, which computes centre of mass of the window, whereas in our proposed method we calculate intersection of row and column where the variance is maximum. Table II shows detection statistics along with details of input video which are tested with our algorithm. The analysis includes four cases i.e. true positive (TP), true negative (TN), false positive (FP) and false negative (FN).

(a) (b)

Fig. 11 (a) 591st frame of walk3 (b) 515th frame of Browse1



(a) (b)

Fig. 13 (a) 214th frame & (b) 228nd frame of AVSS London database.

TABLE II. RESULTS FOR CORRESPONDING VIDEOS

Video	TP	FP	TN	FN
menearbfp.avi	446	97	0	0
Menearbfp1.avi	509	72	0	0
Magnet.avi	426	111	0	2
Turbfloor1.avi	598	369	12	4
Motinas_emilio_webcam.avi	220	167	7	4
Motinas_room105_audiovisual.avi	622	423	15	17
Walk1.mpg	377	198	14	21
Walk2.mpg	539	455	39	21

Walk3.mpg	698	464	12	4
Browse1.mpg	584	397	36	23
Total	**5019**	**2753**	**135**	**96**

Chart 1 shows percentagewise detection data of above videos using the proposed approach. True positive indicates that object is present in the frame and it is detected, false positive indicates that object is not present in the frame and not detected whereas true negative shows that object is present in the frame but not detected and false negative indicates that object is present but still detected. Further we compared our work with the work previously done by using mean shift method. The statistics of the comparison which includes computation time. When the videos are fed to algorithms in offline object detection method are shown in table III. Table III shows comparison of proposed method with well known mean shift method on basis of computation time required to detect the objects for entire video sequences the video datasets [12] [13].

TABLE III. COMPUTATIONAL TIME REQUIRED (IN SECONDS)

Video	Mean Shift Method	Variance Method
menearbfp.avi	32.223188	19.460825
Menearbfp1.avi	30.773116	19.809639
Magnet.avi	15.492027	7.307792
Turbfloor1.avi	224.56335	154.97628
Motinas_emilio_webcam.avi	9.418403	6.421233
Motinas_room105_audioviisual.avi	22.849072	17.104224
Motinas_room160_audioviisual.avi	28.623408	18.016701

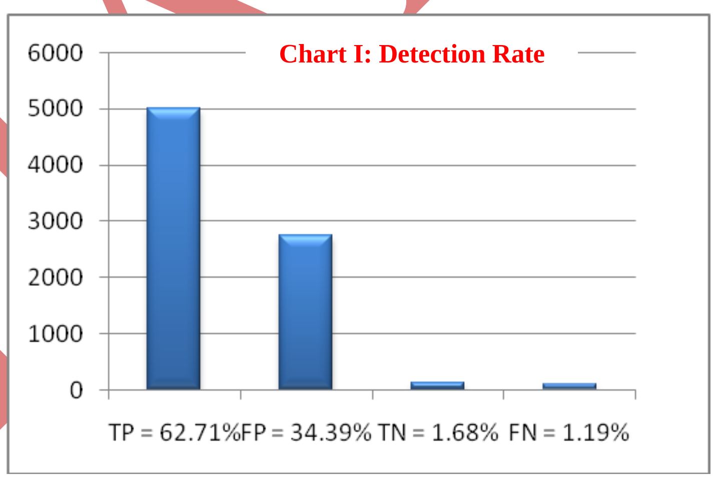


Chart I: Detection Rate

X. CONCLUSION

The proposed algorithm is tested on different videos, From the table II and chart 1 one can conclude that almost 97.11% of total video frames in which proper detection process is carried out successfully. Out of which for 62.71% object was present in the frame and got detected with our algorithm, on the other hand for 34.39% object was not present and not detected in the frames. The error part is only 2.88%. Further to conclude that only high pass components of the 2D-DWT outputs are used in this analysis which helps to detect and track the moving objects. As the object moves the edges are created and the details of which are available in the high pass components of the DWT. High pass components of the DWT output are very good edge detectors which comprises very less amount of noise in it. The comparative analysis of the variance based method for object detection and localization with commonly used mean-shift method concludes that, the proposed variance based method is faster when compared to the existing mean shift method and hence the object detection becomes faster.

REFERENCES

[1] Xiaoshi Zheng, Yanling Zhao, Na li, Huimin Wu," An Automatic Moving Object Detection Algorithm for Video Surveillance Applic ations", 2009 international Conferences on Embedded Software and Systems978-0-7695- 3678-1/09 © 2009 IEEE.

[2] Muyun Weng Guoce Huang Xinyu Da," A New Interframe Difference Algorithm for Moving Target Detection", 2010 3rd International Congress on Image and Signal Processing (CISP2010), 978-1-4244-6516-3/10 ©2010 IEEE.

[3] Arnab Roy, Sanket Shinde and Kyoung-Don Kang," An Approach for Efficient Real Time Moving Object Detection".

[4] Rafael C. Gonzalez & Richard E. Woods," Digital Image Processing" 2nd edition, Prentice Hall.

[5] Anil K. Jain, "Fundamentals-of-Digital-Image-Processing", Prentice Hall.

[6] For video database: (Database downloaded on 25th December, 2013)
Link:http://www.eecs.qmul.ac.uk/ ~andrea/avss2007_d. html

[7] For video database: (Database downloaded on 25th December, 2013)
Link: http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/

[8] Vishal R. Satpute, Kishor D. Kulat, Avinash G. Keskar, "A novel approach based on variance for local feature analysis of facial images", "IEEE-RAICS International Conference on Recent Advances in Intelligent Computational Systems (RAICS) 2011" on 22-24 Sep – 11, pp – 210-215.

[9] K. Sureshraju, Vishal R. Satpute, Kishor D. Kulat, Avinash G. Keskar, "Image Compression Using Wavelet Transform: Compression Ratio and PSNR Calculations", "National Conference on Computer Science and Informatics – NCCSI'12, Madurai" on 23-24 April – 12, pp – 195 -199.

[10] Vishal R. Satpute, Kishor D. Kulat, Avinash G. Keskar, "Variance Method for Finding Local Feature Points on Facial images", "International Conference on Signal, Image and Video Processing (ICSIVP)2012,IITPatna" on 13-15 Jan – 12, pp – 148 -153.

[11] Jerome M. shapiro, "An Embedded hierarchical coder using zero trees of wavelet coefficients.", "Data compression Conference, 1993", pp 214-223.

[12] Gao Tao, Liu Zhengguang & Zhang Jun, "Redundant discrete wavelet transforms based moving object recognition and tracking", "Journal of Systems Engineering and Electronics", Vol. 20, No. 5, 2009, pp.1115-1123.

[13] Ganesh Rakate, "human body tracking system based on DWT and mean shift algorithm on ARM-Linux platform", 978-1-467348058/12©2012 IEEE.

[14] Naveen Ch., Vishal R. Satpute, Kishor D. Kulat, Avinash G. Keskar, "Video encoding techniques based on 3D-DWT", IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2014, pp-1-6

[15] Naveen Ch., Vishal R. Satpute, Kishor D. Kulat, Avinash G. Keskar, "Fast and Memory Efficient 3D-DWT Based Video Encoding Techniques", International MultiConference of Engineers and Computer Scientists