IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-12 E-ISSN: 2347-2693

An Efficient Human Recognition Using Background Subtraction and Bounding Box Technique for Surveillance Systems

Amrut C. Saindane[1*], Pravin S. Patil[2]

[1*]Department of Electronics, North Maharashtra University, Jalgaon
[2]Department of Electronics and communication, North Maharashtra University, Jalgaon
[1]SSVPS BSD College of Engineering, Dhule, Maharashtra, India

Available online at: www.ijcseonline.org

Received: 19/Nov/2016 Revised: 01/Dec/2016 Accepted: 16/Dec/2016 Published: 31/Dec/2016

***Abstract*—Visual surveillance has been a very active research topic in the last few years due to its growing importance in security, law enforcement, and military applications. The project presents moving object detection based on background subtraction for video surveillance system. In all computer vision system, the important step is to separate moving object from background and thus detecting all the objects from video images. The main aim of this paper is to design a bounding box concept for the human detection and tracking system in the presence of crowd. The bounding box around each object can track the moving objects in each frame and it can be used to detect crowd and the estimation of crowd. This paper gives the implementation results of bounding box for detecting objects and its tracking. In order to remove some unwanted pixels, morphological erosion and dilation operation is performed for object edge smoothness. The simulated result shows that used methodologies for effective object detection has better accuracy and with less processing time consumption rather than existing methods.**

Keywords-:Input video, Frame separation, Background subtraction, Morphological Filtering, Performance measurement.

I. INTRODUCTION

There is an increasing desire and need in video surveillance applications for a proposed solution to be able to analyze human behaviours and identify subjects for standoff threat analysis and determination. The main purpose of this survey is to look at current developments and capabilities of visual surveillance systems and assess the feasibility and challenges of using a visual surveillance system to automatically detect abnormal behaviour, detect hostile intent, and identify human subject. Visual (or video) surveillance devices have long been in use to gather information and to monitor people, events and activities. Visual surveillance technologies, CCD cameras, thermal cameras and night vision devices, are the three most widely used devices in the visual surveillance market.

Systems surveillance is the process of monitoring the behaviour of people, objects or processes within systems for conformity to expected or desired norms in trusted systems for security or social control. The word surveillance is commonly used to describe observation from a distance by means of electronic equipment or other technological means.

Fig. 1 Example of CCTV camera

At a basic level, computers are a surveillance target because large amounts of personal information are stored on them. Anyone who can access or remove a computer can retrieve information. If someone is able to install software on a computer system, they can turn the computer into a surveillance device. CCTV is a collection of video cameras used for video surveillance. CCTV is generally used in areas where there is an increased need for security, such as banks, airports and town centers. A basic CCTV system comprises of the Camera, lens and power supply. Recording device, VCR or a digital video recorder and monitor. Closed-circuit television (CCTV) is the use of video cameras to transmit a signal to a specific place, on a limited set of monitors.

The main tasks in visual surveillance systems include motion detection, object classification, tracking. Our focus here is on the detection phase of a general visual surveillance system using static cameras. The usual approach for moving object detection is through background subtraction that consists in maintaining an up-to date model of the background and detecting moving objects as those that deviate from such a model. The background image is not fixed but must adapt to: Illumination changes, sudden (such as clouds), Motion changes, camera oscillations, high-frequencies background objects (such as tree branches, sea waves, and similar) Changes in the background geometry.

II. LITERATURE SURVEY

Several problems arise while segmenting the video sequences because of changing background, clutter, occlusion, varying lighting conditions, automatic operation, adverse weather conditions such as fog, rain, snow, camera angle, and real time processing requirements etc. [1-7]. Zhang [4] divided the segmentation techniques into six groups: - Threshold based techniques, Pixel classification based techniques, Range image segmentation, Color image segmentation, Edge detection based segmentation and techniques based on fuzzy set theory. According to Cheung and Kamath [6], background adaptation techniques could also be categorized as: 1) no recursive and 2) recursive. A non-recursive technique estimates the background based on a sliding-window approach. Various interesting video object segmentation techniques can be found in literature [1-5] such as Running Gaussian Average, Temporal Median Filter, and Mixture of Gaussians.

A. Frame Differencing:

Detection of moving object from a sequence of frames captured from a static camera is widely performed by frame difference method. The objective of the approach is to detect the moving objects from the difference between the existing frame and the reference frame. The frame difference method is the common method of motion detection. This method adopts pixel-based difference to find the moving object.

Difference of Two Consecutive Frames

Ik is supposed to be the value of the kth frame in image sequences. Ik+1is the value of the (k+1)th frame in image sequences. The absolute differential image is defined as follows:

$$Id(k, k+1) = |Ik+1 - Ik|$$

Limitations

The proposed method also detects the motion due to the movement in air. As the air moves, the camera not remains in the position of static so when there is no movement of object then also it results motion and shows holes in the binary output image.

Fig. 2 Example of frame difference method

The Frame difference is arguably the simplest form of background subtraction. Frame differencing, also known as temporal difference, uses the video frame at time t-1 as the background model for the frame at time t. This technique is sensitive to noise and variations in illumination, and does not consider local consistency properties of the change mask. This method also fails to segment the non-background objects if they stop moving. Since it uses only a single previous frame, frame differencing may not be able to identify the interior pixels of a large, uniformly-colored moving object. This is commonly known as the aperture problem. It has strong adaptability, but it is generally difficult to obtain a complete outline of moving object, liable to appear the empty phenomenon, as a result the detection of moving object is not accurate.

III. PROPOSED METHOD

The proposed method is Background subtraction based effective moving object detection.

The process algorithm is described as follow:

1. Input video
2. Frame Separation
3. Image Sequence
4. Separation of Image Sequence in Current Frame Image and Background Frame Image
5. Background initiation.
6. Background Subtraction
11. Morphological filtering
12. Detection of Moving Object
13. Performance measurement (Similarity/MSE/PSNR/Correlation)

A. Proposed Method Block diagram:

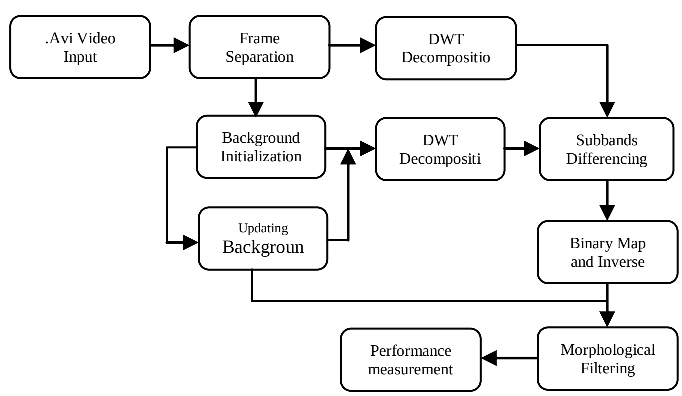

Fig.3block diagram of proposed method

3.1 Pre-processing:

In Pre-processing of the proposed system the following steps namely Gray scale conversion, Noise removal is involved. In computing, a gray scale digital image is an image in which the value of each pixel is a single sample, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. Gray scale images are distinct from one-bit bi-tonal black-and-white images, which in the context of computer imaging are images with only the two colors, black, and Gray scale images have many shades of gray in between. Gray scale images are often the result of measuring the intensity of light at each pixel in a single band of the electromagnetic spectrum, and in such cases they are monochromatic proper when only a given frequency is captured. And the gray scale conversion of image is given by [17].

$$gray(i,j) = \{0.29 * rgb(:,:,1) + 0.59 * rgb(:,:,2) + 0.11 * rgb(:,:,3)\};$$

Generally we are using median filter to suppress the noise. The procedures are

(i) Arranging matrix pixel value in the form of ascending order.

(ii) Find the median value of that matrix.

(iii) Replace that value into that noisy pixel location.

3.2 Background subtraction

The basic scheme of background subtraction is to subtract the image from a reference image that models the background scene.Background modeling constructs a reference image representing the background. Threshold selection determines appropriate threshold values used in the subtraction operation to obtain a desired detection rate. Subtraction operation or pixel classification classifies the type of a given pixel, i.e., the pixel is the part of background (including ordinary background and shaded background), or it is a moving object.

After background image B(X, Y) is obtained, subtract the background image B(X, Y) from the current frame FK(X, Y). If the pixel difference is greater than the set threshold T, then determines that the pixels appear in the moving object, otherwise, as the background pixels. The moving object can be detected after threshold operation.

Its expression is as follows:

$$DK(X,Y) = \begin{cases} \text{if } (|FK(X,Y) - B(X,Y)| > T) \\ 0 \text{ others} \end{cases} \quad (1)$$

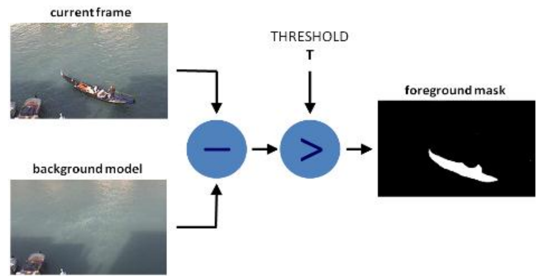

Fig. 4 Example of Background subtraction method

Background subtraction method is very sensitive to the changes in the external environment. The methods with a background model based on a single scalar value can guarantee adaptation to slow illumination changes, but cannot cope with multi-valued background distributions. As such, they will be prone to errors whenever those situations arise. Processing time required to detect the object using this technique is low but accuracy may not be good enough.

IV(1) EXPERIMENTS AND RESULTS

In this work the aim is to build such a surveillance system, which will detect motion even if the moving background, gradual illumination variations and camouflage and shadow into the background, thus achieves robust detection for different types of videos taken with stationary cameras. To fulfill this aim, strong computing software called Matlab is used. Matlab provides image Acquisition and Image Processing Toolboxes which facilitate us in creating a good code.

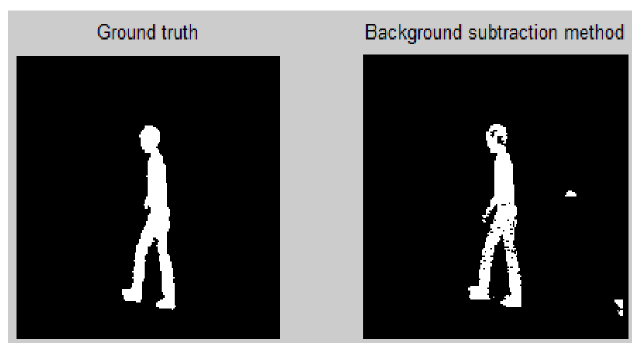

Fig. 5 Result of Background subtraction methodfor standard video

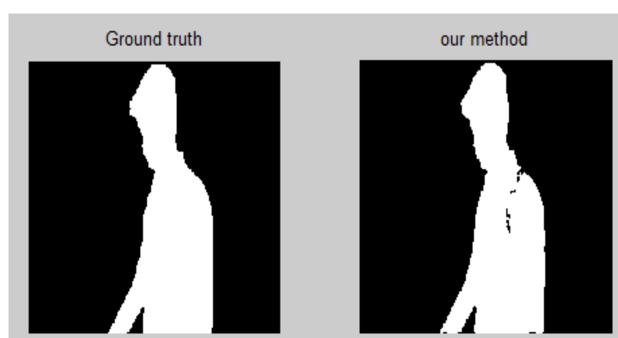

Fig.6 Result of Background subtraction method for Real video

Table 1.1 Validation for Standard video

S.No	Parameters	Methods	
		Frame difference	Background subtraction
1	MSE	0.0487	0.0115
2	PSNR	61.2563	67.5392
3	Entropy	0.0917	0.295
4	correlation coefficient	0.278302	0.8832
5	recall	0.955581	0.9946
6	precision	0.9949	0.9939
7	F1	0.975047	0.9942
8	Similarity	0.951309	0.9885

Table 1.2 Validation for Real video

S.No	Parameters	Methods	
		Frame difference	Background subtraction
1	MSE	0.208	0.002
2	PSNR	54.9503	75.0246
3	Entropy	0.1648	0.7745
4	correlation coefficient	0.2714	0.9942
5	recall	0.7929	0.998
6	precision	0.9986	1
7	F1	0.8839	0.999
8	Similarity	0.792	0.998

Graphical Representation of Real and Standard video Results:

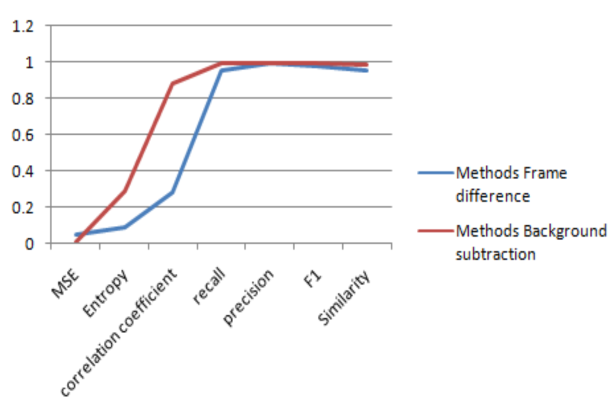

Fig. 7 Graphical Representation of Background subtraction method for Standard video

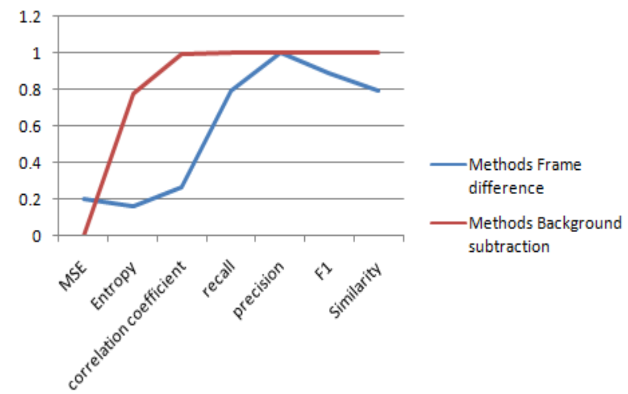

Fig. 8 Graphical Representation of Background subtraction method for Real video

V. PERFORMANCE EVALUATION

It can be observed from the results that none of the previously proposed segmentation algorithms give accurate segmentation result as compared to ground truth frames. In this paper, the performance of the proposed method has been compared quantitatively with other state-of-the-art methods. The proposed system is experimented with different settings of adjustable parameters which can be used for performance evaluation. This section outlines the set of performance evaluation metrics that have implemented in order to quantitatively analyze the performance of object detection methods. The ground truth information is represented in left side of output figure. Similarly, the results of object detection of different methods are shown in right side of output figure.

5.1 Frame-based Metrics

Starting with the first frame of the test sequence, frame based metrics are computed for every frame in the sequence. From each frame in the video sequence, first a few true and false detection and tracking quantities are computed.

True Negative, TN: Number of frames where both ground truth and system results agrees on the absence of any object.

True Positive, TP: Number of frames where both ground truth and system results agree on the presence of one or more objects, and the bounding box of at least one or more objects coincides among ground truth and tracker results.

False Negative, FN: Number of frames where ground truth contains at least one object, while system either does not contain any object or none of the system's objects fall within the bounding box of any ground truth object.

False Positive, FP: Number of frames where system results contain at least one object, while ground truth either does not contain any object or none of the ground truth's objects fall within the bounding box of any system object.
Once the above defined quantities are calculated for all the frames in the test sequence, in the second step, the following

metrics are computed: For measuring accuracy we adopted different metrics,

A. Mean Square Error (MSE) and Peak Signal to Noise ratio(PSNR)

The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image compression quality. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error. To compute the PSNR, the block first calculates the mean-squared error using the following equation:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N}$$

In this equation, M and N are the number of rows and columns in the input images, respectively. Now the block computes the PSNR using the following equation:

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right)$$

The PSNR between two images having 8 bits per pixel or sample in terms of decibels (dBs) is Generally when PSNR is 40 dB or greater, then the original and the reconstructed images are virtually indistinguishable by human observers . In this equation, *R* is the maximum fluctuation in the input image data type. For example, if the input image has a double-precision floating-point data type, then *R* is 1. If it has an 8-bit unsigned integer data type, *R* is 255, etc.

Correlation coefficient:

This gives statistical relationship between two or more random variable or observed data values. This computes the correlation coefficient between A and B, where A and B are matrices or vectors of the same size.

Similarity:

We considered the pixel-based *Similarity* measure as

$$\textbf{Similarity} = \frac{tp}{tp+fn+fp}$$

Greatest value of similarity shows accurate detection of moving object.

VI. Conclusion

In video surveillance, there are many interference factors such as target changes, complex scenes, and target deformation in the moving object tracking. In this paper moving objects detection using background subtraction have been proposed. The performance of the proposed method have been evaluated and compared with other standard methods in consideration in terms of various performance metrics. From the obtained results and their qualitative and quantitative analysis, it can be concluded that the proposed method is performing better in comparison to other methods.
Also we had applied our system for standard and real videos, and our algorithm performs very well in both input category. And we archived better results in real videos.

VII. Reference

[1]. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. Int. J. Comput. Vision 1, 321–332, 1988.

[2]. V.Caselles, R. Kimmel, and G. Sapiro.Geodesic active contours. In IEEE International Conference on Computer Vision .694–699, 1995.

[3]. N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, 266–280, 2000.

[4]. Comaniciu, D. And Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Analy. Mach. Intell. 24, 5, 603–619.

[5]. S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Patt. Analy. Mach. Intell. 18, 9, 884–900, 1996.

[6]. S.C. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," Video Communications and Image Processing, SPIE Electronic Imaging, UCRL Conf. San Jose, vol.200706, Jan 2004.

[7]. L. Maddalena and A. Petrosino, “A self-organizing approach to detection of moving patterns for real-time applications,” in Proc. 2nd Int. Symp. Brain, Vision, and ArtificialIntelligence, 2007, pp. 181–190, Lecture Notes Comput. Sci. 4729

[8]. N. McFarlane, C. Schofield, “Segmentation and tracking of piglets in images”, Machine Vision Application, Vol. 8, No. 3, pp. 187-193, 1995

[9]. C. Stauffer , W.E.L. Grimson, “Adaptive background mixture models for real-time tracking,” In Proc. Computer Vision and Pattern Recognition conference ,pp. 246-252, 1999.

[10]. Z. Zivkovic and F. van der Heijden,” Efficient adaptive density estimation per image pixel for the task of background subtraction”, Pattern Recognition Letters, Vol. 27 No.7, pp. 773–780,

[11]. KanokphanLertniphonphan , SupavadeeAramvith and Thanarat H,“The Region-Based Distance of Oriented Gradient and Motion Directionfor Human Action Classification”,

International Sumposium onCommunication and Information technologies, IEEE 2012.

[12]. Kairoek Choeychuent, PinitKumhomta nd and KosinChamnongthait,"An Efficient Implementation of the Nearest Neighbor Based VisualObjects Tracking", International Symposium on Intelligent SignalProcessing and Communication Systems (ISPACS2006), Japan,2006.

[13]. Smitha Suresh, Chitra K and Deepak P" Patch Based frame work forocclusion Handling", IEEE 2013.

[14]. Nan Lu, Jihong Wang, Q.H Wu, Li Yang, An Improved Motion Detection Method for Real-Time Surveillance, International Journal Of Computer Science, vol.1, Issue 6, 2008.

[15]. Guangping Hu, HuijieLian, Pei Wang, Design and Research on the Motion Target Detection and Tracking Algorithm, (978-1-4244-76183)Information Science and Engineering, 2nd International conference, pp 1114-1117, Dec 2010.

