A Survey on Human Detection Techniques used in Video Surveillance System

Aditya H. Karad[1], R. U. Shekokar[2]

[1]Department of Electronics & Telecommunication

[2]Professor, Department of Electronics & Telecommunication

R.M.D Sinhagad School of Engineering, Pune, India

***Abstract*— Detecting moving objects accurately in a visual surveillance system is crucial for diverse application areas including abnormal event detection, human gait characterization, congestion analysis, person identification, gender classification and fall detection for elderly people. At present, object detection methods widely used are background subtraction and Frame difference. The core of background subtraction is background modeling. Moving Object detection could be performed using background subtraction, optical flow, frame differencing technique, background subtraction using neural network etc. The first step of the detection process is to detect an object which is in motion and then we can classify them as per our requirement. A comprehensive review with comparisons on available techniques for moving object detection in surveillance videos is presented in this paper. The future research directions on detection have also been discussed.**

***Index Terms*— Background Subtraction, Optical flow, Frame differencing and template matching, BS with neural network, BS with wavelet transform.**

I. Introduction

Over the recent years, detecting objects in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in a dense crowd, person identification, gender classification, fall detection for elderly people, etc.

The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field.

An intelligent system detects and captures motion information of moving targets for accurate object classification. The classified object is being tracked for high-level analysis. In this study, we focus on moving object detection and do not consider recognition of their complex activities. Human detection is a difficult task from a machine vision perspective as it is influenced by a wide range of possible appearance due to changing articulated pose, clothing, lighting and background, but prior knowledge on these limitations can improve the detection performance [1].

The detection process generally occurs in two steps: object detection and object classification. Moving Object detection could be performed by background subtraction, optical flow, frame differencing, background subtraction using neural network, background subtraction using wavelet transform etc. Background subtraction is a popular method for object detection where it attempts to detect moving objects from the difference between the current frame and a background frame in a pixel by-pixel or block-by-block fashion [1, 2].

Background subtraction is a popular method to detect an object as a foreground by segmenting it from a scene of a surveillance camera. The camera could be fixed, pure translational or mobile in nature. Background subtraction attempts to detect moving objects from the difference between the current frame and the reference frame in a pixel-by-pixel or block-by-block fashion. The reference frame is commonly known as 'background image', 'background model' or 'environment model'. A good background model needs to be adaptive to the changes in dynamic scenes. Updating the background information in regular intervals could do this, but this could also be done without updating background information. Few available approaches have been discussed in this section:

a) Mixture of Gaussian model:

Stauffer and Grimson introduced an adaptive Gaussian mixture model, which is sensitive to the changes in dynamic scenes derived from illumination changes, extraneous events etc. Rather than modeling the values of all the pixels of an image as one particular type of distribution, they modeled the values of each pixel as a mixture of Gaussians. Over time, new pixel values update the mixture of Gaussian (MoG) using an online K-means approximation. In the literature, many approaches are proposed to improve the MoG. In an effective learning algorithm for MoG is proposed to overcome the requirement of the prior knowledge about the foreground and background ratio. In authors presented an algorithm to control the number of Gaussians adaptively in order to improve the computational time without sacrificing the background modeling quality. Each pixel is modelled by support vector regression. Kalman filter is used for adaptive background estimation. In a framework for hidden Markov Model (HMM) topology and parameter estimation is proposed. In colour and edge information are used to detect foreground regions. In normalized coefficients of five kinds of orthogonal transform (discrete cosine transformation, discrete Fourier transformation (DFT), Haar transform, single value decomposition and Hadamard transform) are utilized to detect moving regions. In each pixel is modelled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel.

b) Non-parametric background model:

Sometimes, optimization of parameters for a specific environment is a difficult task. Thus, a number of researchers introduced non-parametric background modeling techniques. Non-parametric background models consider the statistical behavior of image features to segment the foreground from the background. A non-parametric model is proposed for background modeling, where a kernel-based function is employed to represent the colour distribution of each background pixel. The kernel-based distribution is a generalization of MoG, which does not require parameter estimation. The computational requirement is high for this method. Kim and Kim proposed a non-parametric method, which was found effective for background subtraction in dynamic texture scenes (e.g. waving leaves, spouting fountain and rippling water). They proposed a clustering-based feature, called fuzzy colour histogram (FCH) to construct the background model by computing the similarity between local FCH features with an online update procedure. Although the processing time was high in comparison with the adaptive Gaussian mixture model, the false positive rate of detection is significantly low at high true positive rates.

c) Warping background:

Ko presented a background model that differentiates between background motion and foreground objects. Unlike most models that represent the variability of pixel intensity at a particular location in the image, they modeled the underlying warping of pixel locations arising from background motion. The background is modeled as a set of warping layers where at any given time, different layers may be visible due to the motion of an occluding layer. Foreground regions are thus defined as those that cannot be modeled by some composition of some warping of these background layers.

d) Hierarchical background model:

Chen proposed a hierarchical background model, which is based on region segmentation and pixel descriptors to detect and track foreground. It first segments the background images into several regions by the mean-shift algorithm. Then, a hierarchical model, which consists of the region models and pixel models, is created. The region model is one kind of approximate Gaussian mixture model extracted from the histogram of a specific region. The pixel model is based on the co-occurrence of image variations described by HOG of pixels in each region. Benefiting from the background segmentation, the region models and pixel models corresponding to different regions can be set to different parameters. The pixel descriptors are calculated only from neighbouring pixels belonging to the same object. The hierarchical models first detect the regions containing foreground and then locate the foreground only in these regions, thus avoid detection failure in other regions and reduce the time and cost. A similar two-stage hierarchical method has been introduced earlier by Chen where the block-based stage provides a course foreground segmentation followed by the pixel-based stage for finer segmentation. The method showed promising results when compared with MoG. Recent application of this approach can be seen in the study of Quan where the hierarchical background model (HBM) is combined with the codebook technique [1].

II. HUMAN MOTION DETECTION TECHNIQUES

A moving object i.e. human is generally detected by segmenting motion in a video image. Most conventional approaches for object detection are background subtraction, optical flow, frame differencing etc. They are outlined in the following subsections.

1. BACKGROUND SUBTRACTION

The background subtraction method is the common method of motion detection. It is a technology that uses the difference of the current image and the background image to detect the motion region, and it is generally able to provide data included object information. The key of this method lies in the initialization and update of the background image. The effectiveness of both will affect the accuracy of test results. Therefore, this paper uses an effective method to initialize the background, and update the background in real time.

A. Background image initialization

There are many ways to obtain the initial background image. For example, with the first frame as the background directly, or the average pixel brightness of the first few frames as the background or using a background image sequences without the prospect of moving objects to estimate the background model parameters and so on. Among these methods, the time average method is the most commonly used method of the establishment of an initial background. However, this method can't deal with the background image (especially the region of frequent movement) which has the shadow problems. While the method of taking the median from continuous multi-frame can resolve this problem simply and effectively. So the median method is selected in this paper to initialize the background. Expression is as follows:

$$B_{init}(x,y) = median\ f_k(x,y)\ k = 1,2,\dots n \qquad \dots(1)$$

Where $B_{init}$is the initial background, n is the total number of frames selected.

B. Background Update

For the background model can better adapt to light changes, the background needs to be updated in real time, so as to accurately extract the moving object. In this paper, the update algorithm is as follows:

In detection of the moving object, the pixels judged as belonging to the moving object maintain the original background gray values, not be updated. For the pixels which are judged to be the background, we update the background model according to following rules:

$$B_{k+1}(x,y) = \beta B_k(x,y) + (1-\beta)F_k(x,y) \qquad \dots(2)$$

Where βϵ (0,1) is update coefficient, in this paper β=0.004. $F_k(x,y)$ is the pixel gray value in the current frame. $B_k(x,y)$ And B_{k+1} are respectively the background value of the current frame and the next frame. As the camera is fixed, the background model can remain relatively stable in the long period of time. Using this method can effectively avoid the unexpected phenomenon of the background, such as the sudden appearance of something in the background which is not included in the original background. Moreover by the update of pixel gray value of the background, the impact brought by light, weather and other changes in the external environment can be effectively adapted.

C. Moving object detection

Now we will discuss steps involved in background subtraction method for moving object detection.

1. Moving object extraction

After the background image $B(x,y)$ is obtained, subtract the background image $B(x,y)$ from the current frame $F_k(x,y)$. If the 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:

$$D_k(x,y) = \begin{cases} 1 & |F_k(x,y) - B_{k-1}(x,y)| > T \\ 0 & others \end{cases} \qquad \dots(3)$$

Where $D_k(x,y)$ is the binary image of differential results. T is gray scale threshold, its size determines the accuracy of object identification.

As in the algorithm T is a fixed value, only for an ideal situation, is not suitable for complex environment with lighting changes. Therefore, this paper proposes the dynamic threshold method, we dynamically changes the threshold value according to the lighting changes of the two images obtained. On this basis, add a dynamic threshold ΔT to the above algorithm. Its mathematical expression is as follows:

$$\Delta T = \lambda \cdot \frac{1}{M \times N} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} |F(i,j) - B(i,j)| \qquad \dots(4)$$

Then

$$D_k(x,y) = \begin{cases} 1 & |F_k(x,y) - B_{k-1}(x,y)| > T + \Delta T \\ 0 & others \end{cases} \qquad \dots(5)$$

Where λ is the inhibitory coefficient, set it to a value according to the requirements of practical applications, and the reference values is 2. $M \times N$ is the size of each image to deal with. $M \times N$ numerical results indicate the number of pixels in detection region. ΔT reflects the overall changes in the environment. If small changes in image illumination, dynamic threshold ΔT takes a very small value. Under the premise of enough pixels in the detection region, ΔT will tend to O. If the image illumination changes significantly, then the dynamic threshold ΔT will increase significantly. This method can effectively suppress the impact of light changes.

2. Reprocessing

As the complexity of the background, the difference image obtained contains the motion region, in addition, also a large number of noise. Therefore, noise needs to be removed. This paper adopts median filter with the 3×3 window and filters out some noise.

After the median filter, in addition the motion region, includes not only body parts, but also may include moving cars, flying birds, flowing clouds and swaying trees and other non-body parts. Morphological methods are used for further processing. Firstly, corrosion operation is taken to effectively filter out non-human activity areas. Secondly, using the expansion operation to filter out most of the non-body motion regions while preserving the shape of human motion without injury. After expansion and corrosion operations, some isolated spots of the image and some interference of small pieces are eliminated, and we get more accurate human motion region.

3. Extraction of Moving Human Body

After median filtering and morphological operations, some accurate edge regions will be got, but the region belongs to the moving human body could not be determined. Through observation, we can find out that when moving object appears, shadow will appear in some regions of the scene. The presence of shadow will affect the accurate extraction of the moving object. By analyzing the characteristics of motion detection, we combine the projection operator with the previous methods. Based on the results of the methods above, adopting the method of combining vertical with horizontal projection to detect the height of the motion region. This can eliminate the impact of the shadow to a certain degree. Then we analyze the vertical projection value and set the threshold value (determined by experience) to remove the pseudo-local maximum value and the pseudo-local minimum value of the vertical projection to determine the number and width of the body in the motion region, we will get the moving human body with precise edge. This article assumes that people in the scene are all in upright-walking state. The flow chart of moving human body extraction is shown in Figure 2.l.

Human body detection is to identify the corresponding part of human from the moving region. But the extracted moving region may correspond to different moving objects, such as pedestrians, vehicles and other such birds, floating clouds, the swaying tree and other moving objects. Hence we use the shape features of motion regions to further determine whether the moving object is a human being. Judging criteria are as follows: (1) The object area is larger than the set threshold (2) The aspect ratio of the object region should conform to the set ratio. If these two conditions are met, the moving object is the moving human body, or is not a human body [3].

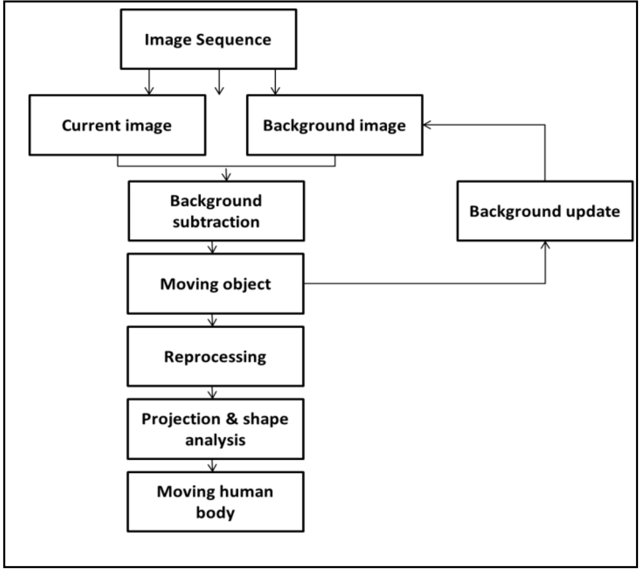

Figure 2.1 The flow chart of moving human body extraction

2. OPTICAL FLOW

Optical flow is a vector-based approach that estimates motion in video by matching points on objects over image frame(s). Under the assumption of brightness constancy and spatial smoothness, optical flow is used to describe coherent motion of points or features between image frames. Optical flow-based motion segmentation uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence. One key benefit of using optical flow is that it is robust to multiple and simultaneous cameras and object motions, making it ideal for crowd analysis and conditions that contain dense motion. Optical flow-based methods can be used to detect independently moving objects even in the presence of camera motion. Apart from their vulnerability to image noise, colour and non-uniform lighting, most of flow computation methods have large computational requirements and are sensitive to motion discontinuities. A real-time implementation of optical flow will often require a specialized hardware due to the complexity of the algorithm and moderately high frame rate for accurate measurements [1, 2].

Rowley and Rehg also focused on the segmentation of optical Flow fields of articulated objects. Its major contributions were to add kinematic motion constraints to each pixel, and to combine motion segmentation with estimation in expectation maximization (EM) computation.

Bregler’s work, each pixel was represented by its optical Flow. These Flow vectors were grouped into blobs having coherent motion and characterized by a mixture of multivariate Gaussians.

Friedman and Russell implemented a mixture of Gaussian classification model for each pixel. This model attempted to explicitly classify the pixel values into three separate predetermined distributions corresponding to background, foreground and shadow. Meanwhile it could also update the mixture component automatically for each class according to the likelihood of membership. Hence, slow-moving objects were handled perfectly; meanwhile shadows were eliminated much more effectively.

Barron's work, In addition to the basic methods described above, there are some other approaches to motion segmentation. Using the extended EM algorithm, Stringa also proposed a novel morphological algorithm for scene change detection. This proposed method allowed obtaining a stationary system even under varying environmental conditions. From the practical point of view, the statistical methods are better choice due to their adaptability in more unconstrained applications.

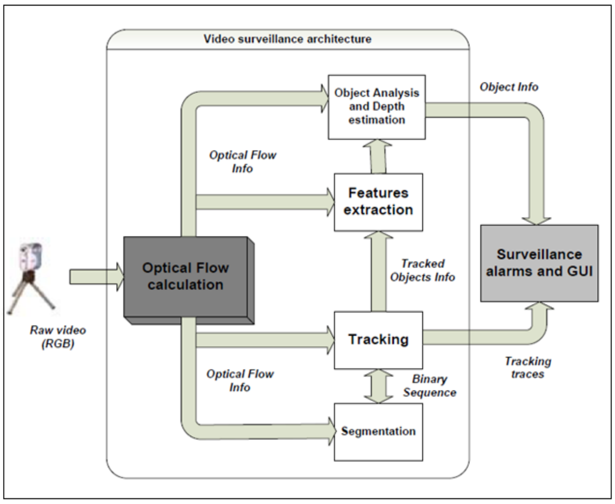

Figure 2.2 Optical Flow System Architecture

3. FRAME DIFFERENCING AND TEMPLATE MATCHING

Frame Differencing is a technique where the computer checks the difference between two video frames. If the pixels have changed there apparently was something changing in the image (consider frame). The Frame Differencing Algorithm (Jain and Nagel, 1979; Haritaoglu, 2000) is used for this purpose, which gives the position of object as output. This extracted position is then used to extract a rectangular image template (size is dynamic depending upon the dimension of object) from that region of the image. The sequence of templates is generated as object changes its position.

The generated templates from each frame are passed on to the tracking module, which starts tracking the moving object with an input reference template. The module uses template-matching (Richard, 2004; Comaniciu, 2003) to search for the input template in the scene grabbed by the camera. A new template is generated if the object is lost while tracking due to change in its appearance and used further. Generations of such templates are dynamic which helps to track the object in a robust manner. The main objective of this study is to provide a better and enhanced method to find the moving objects in the continuous video frame as well as to track them dynamically using template matching of the desired object. This method is effective in reducing the number of false alarms that may be triggered by a number of reasons such as bad weather or other natural calamity.

a) Object detection using frame differencing

The task to identify moving objects in a video sequence is critical and fundamental for a general object tracking system. For this approach Frame Differencing technique (Jain and Nagel, 1979) is applied to the consecutive frames, which identifies all the moving objects in consecutive frames. This basic technique employs the image subtraction operator (Rafael and Richard, 2002), which takes two images (or frames) as input and produces the output. This output is simply a third image produced after subtracting the second image pixel values from the first image pixel values. This subtraction is in a single pass. The general

Operation performed for this purpose is given by:

$$DIFF[i,j] = I_1[i,j] - I_3[i,j]$$

$DIFF[i,j]$ represents the difference image of two frames.

It seems to be a simple process but the challenge occurs is to develop a good frame differencing algorithm for object detection. These challenges can be of any type like:

1) Due to change in illumination the algorithm should be robust.
2) The detection of non-stationary object (like wind, snow etc.) is to be removed.

To overcome such challenges we need to pre-process the $DIFF[i,j]$ image. Pre-processing includes some mathematical morphological operations which results in an efficient difference image. $DIFF[i,j]$ image is first converted into a binary image by using binary threshold and the resultant binary image is processed by morphological operations. This proposed algorithm for object detection is to achieve these challenges and provide a highly efficient algorithm to maintain such task of object tracking. This algorithm provides the position of the moving object.

Steps involved in this algorithm are as follows:

Here we assume all previous frames are stored in a memory buffer and the current frame in video is F_i

1. Take i^{th} frame (F_i) as input.
2. Take $(i-3)^{th}$ frame (F_{i-3}) from the image buffer.

This image buffer is generally a temporary buffer used to store some of previous frames for future use.

3. Now, perform Frame Differencing Operation on the i^{th} and $(i-3)^{th}$frame. The resultant image generated is represented as:

$$DIFF_i = F_{i-3} - F_i$$

In earlier methods, Frame Differencing is to be performed on i^{th} and $(i-1)^{th}$ frame, which has limitation to detect slow moving objects. But this proposed method, the Frame Differencing between i^{th} and $(i-1)^{th}$ frame. This method removes the limitation to detect slow moving object, which makes it independent of speed of moving object and more reliable.

After the Frame Differencing Operation the Binary Threshold Operation is performed to convert difference image into a binary image with some threshold value and thus the moving object is identified with some irrelevant non-moving pixels due to flickering of camera. And some moving pixels are also there in binary image which corresponds to wind, dust, illusion etc., all these extra pixels should be removed in steps of pre-processing. The binary image (F_{bin}), in which the pixel corresponding to moving object is set to 1 and rest, is treated as background which sets to 0.

This threshold technique work as, a brightness Threshold (T) is chosen with the $DIFF[i,j]$ to which threshold is to be applied:

If $DIFF[i,j] \geq T$ then
$F_{bin[i,j]} = 1$ //for object
else
$F_{bin[i,j]} = 0$ //for background

This assumes that the interested parts are only light objects with a dark background. But for dark object having light background we use:

If $DIFF[i,j] \leq T$ then
$F_{bin} = 1$ //for object
else
$F_{bin} = 0$ //for background

The next operation is to calculate the Centre of Gravity (COG) of the binary objects in image F_{mor}. According to the centroid position a fixed sized rectangular box or a bounding box or perimeter is made for all the binary objects in i^{th} frame F_i. All of the centroid information is stored in a global array. By using threads and producer-consumer concept the centroid information of some objects is transferred to the object tracking module. Now, follow same procedure for other frame (starting from step 1).

b) Tracking of object using template matching:

The limitation with this tracking module is that all the centroid information received by motion detection module for tracking of objects should always be in camera view. The next operation is to track the only interesting moving object irrespective of other moving objects. The object tracker module is used for this purpose, which keep track of interesting objects over time by locating the position of moving object in every frame of the video. The proposed algorithm has flexibility to perform both task, object detection and to track object instances across frames simultaneously [4].

4. BACKGROUND SUBTRACTION USING NEURAL NETWORK

As we already know that the main problem of the background subtraction approach to moving object detection is its extreme sensitivity to dynamic scene changes due to lighting and extraneous events. Although these are usually detected, they leave behind “holes” where the newly exposed background imagery differs from the known background model (ghosts). While the background model eventually adapts to these “holes,” they generate false alarms for a short period of time. Therefore, it is highly desirable to construct an approach to motion detection based on a background model that automatically adapts to changes in a self-organizing manner and without a priori knowledge.

This system adopts a biologically inspired problem-solving method based on visual attention mechanisms. The aim is to obtain the objects that keep the user attention in accordance with a set of predefined features, including gray level, motion and shape features. Our approach defines a method for the generation of an active attention focus to monitor dynamic scenes for surveillance purposes. The idea is to build the background model by learning in a self-organizing manner many background variations, i.e., background motion cycles, seen as trajectories of pixels in time. Based on the learnt background model through a map of motion and stationary patterns, our algorithm can detect motion and selectively update the background model. Specifically, a novel neural network mapping method is proposed to use a whole trajectory incrementally in time fed as an input to the network. This makes the network structure much simpler and the learning process much more efficient.

The adopted artificial neural network is organized as a 2-D flat grid of neurons (or nodes) and, similarly to self-organizing maps (SOMs) or Kohonen networks, allows to produce representations of training samples with lower dimensionality, at the same time preserving topological neighbourhood relations of the input patterns (nearby outputs correspond to nearby input patterns). Each node computes a function of the weighted linear combination of incoming inputs, where weights resemble the neural network learning. Doing so, each node could be represented by a weight vector, obtained collecting the weights related to incoming links. In the following, the set of weight vectors will be called a model. An incoming pattern is mapped to the node whose model is "most similar" (according to a predefined metric) to the pattern, and weight vectors in a neighbourhood of such node are updated. Therefore, the network behaves as a competitive neural network that implements a winner take- all function with an associated mechanism that modifies the local synaptic plasticity of the neurons, allowing learning to be restricted spatially to the local neighbourhood of the most active neurons. For each colour pixel, we consider a neuronal map consisting of $n \times n$ weight vectors. Each incoming sample is mapped to the weight vector that is closest according to a suitable distance measure, and the weight vectors in its neighbourhood are updated. The whole set of weight vectors acts as a background model that is used for background subtraction in order to identify moving pixels.

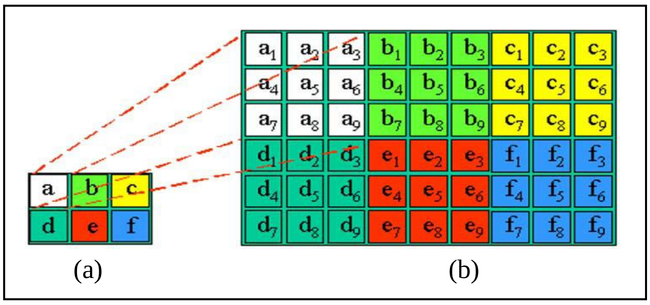
a
b
c
d
e
f
a_1 a_2 a_3 b_1 b_2 b_3 c_1 c_2 c_3
a_4 a_5 a_6 b_4 b_5 b_6 c_4 c_5 c_6
a_7 a_8 a_9 b_7 b_8 b_9 c_7 c_8 c_9
d_1 d_2 d_3 e_1 e_2 e_3 f_1 f_2 f_3
d_4 d_5 d_6 e_4 e_5 e_6 f_4 f_5 f_6
d_7 d_8 d_9 e_7 e_8 e_9 f_7 f_8 f_9
(a)
(b)

Figure 2.3 (a) Simple image and the (b) Neuronal map structure.

a) Initial Background Model

In the case of our background modeling application, we have at our disposal a fairly good means of initializing the weight vectors of the network: the first image of our sequence is indeed a good initial approximation of the background, and, therefore, for each pixel, the corresponding weight vectors are initialized with the pixel value. In order to represent each weight vector, we choose the HSV color space, relying on the hue, saturation and value properties of each color. Such color space allows us to specify colors in a way that is close to human experience of colors. Moreover, the intensity of the light is explicit and separated from chromaticity, and this allows change detection invariant to modifications of illumination strength. Let (h, s, v) be the HSV components of the generic pixel (x, y) of the first sequence frame I_0 , and let $C = (c_1, c_2, c_{3,} \ldots\ldots.c_{n^2})$ be the model for pixel (x, y). Each weight vector $c_i, i = 1, \ldots..n^2$, is a 3-D vector initialized as $c_i = (h, s, v)$.

The complete set of weight vectors for all pixels of an image I with N rows and M columns is represented as a neuronal map A with $(n \times N)$ rows and $(n \times M)$ columns, where the weight vectors for the generic pixel (x, y) of I are at neuronal map positions $(i, j), i = n \times x, \ldots\ldots, n \times (x + 1) - 1$ and $j = n \times y, \ldots, n \times (y + 1) - 1$. An example of such neuronal map structure for a simple image I with $N = 2$ rows and $M = 3$ columns obtained choosing $n = 3$ is given in Fig. 1. As depicted, the upper left pixel a of I (Fig. 1, left) has weight vectors $(a_1, \ldots\ldots, a_9)$ stored into the 3×3 elements of the upper left part of neuronal map A (Fig. 1, right). Analogous relations exist for each pixel I of and corresponding weight vectors storage. It appears evident that the neuronal map A can be seen as an initial background model, enlarged 3×3 times.

This configuration allows to easily taking into account the spatial relationship among pixels and corresponding weight vectors, as we shall see in the following section.

b) Subtraction and Update of the Background Model

After initialization, temporally subsequent samples are fed to the network. Each incoming pixel p_t of the t_{th} sequence frame I_t is compared to the current pixel C model to determine if there exists a weight vector that best matches it. If a best matching weight vector c_m is found, it means that p_t belongs to the background and it is used as the pixel encoding approximation, and the best matching weight vector, together with its neighbourhood, is reinforced. Otherwise, if no acceptable matching weight vector exists, we discriminate whether p_t is in the shadow cast by some object or not. In the first case, p_t should be still considered as background, but it should not be used to update the corresponding weight vectors, in order to avoid the reinforcement of shadow information into the background model; in the latter case p_t is detected as belonging to a moving object (foreground). The above described background subtraction and update procedure for each pixel can be sketched as in the following algorithm.

Algorithm SOBS (Self-Organizing Background Subtraction)

```
Input: pixel value $p_t$ in $I_t, t = 0, \ldots\ldots, LastFrame$
Output: background/foreground binary mask value $B(p_t)$
1. Initialize model $C$ for pixel $p_0$ and store it into $A$
2. for , $t = 1, LastFrame$
3. Find best match $c_m$ in $C$ to current sample $p_t$
```

```
4. **if** ($c_m$ found) **then**
5. $B(p_t) = 0$ //background
6. update $A$ in the neighborhood of $c_m$
7. **else if**( $p_t$ shadow) **then**
8. $B(p_t) = 0$ //background
9. **else**
10. $B(p_t) = 1$  //foreground
```

In practice, we distinguish the whole process into two phases: a calibration phase and an online phase. The calibration phase involves the neural network initial learning, and consists in steps 1–6 executed on the first $K + 1$ sequence frames (i.e., for $t = 1, K$ in step 2), with $K < LastFrame$. The online phase involves neural network adaptation and background subtraction, and consists in steps 2–10 executed on the last $LastFrame - K$ sequence frames (i.e., for $t = K + 1, LastFrame$ in step 2). The number K of sequence frames for the calibration phase basically depends on how many static (free of moving foreground objects) initial frames are available for each sequence [5].

5. BACKGROUND SUBTRACTION USING WAVELET TRANSFORM

Cheng proposed a discrete wavelet transform (DWT) based method for multiple objects tracking and identification. In this approach, inter-frame differencing method is used for segmentation of moving object in DWT domain. DWT based methods are shift-sensitive. Any shift sensitive method will not give good results for video applications because in video applications, objects are present in shifted form.

Motivated by these facts, a new method for video segmentation using Approximate Median Filter method in Daubechies complex wavelet domain is proposed in this section. The Daubechies complex wavelet transform have advantages of shift invariance and better directional selectivity as compared to DWT.

The proposed method is an approximate median filter based method in Daubechies complex wavelet domain. It uses frame differencing for obtaining video object planes which gives the changed pixel value from consecutive frames. First, we decompose two consecutive frames (In-1 and In) using complex wavelet domain and then apply approximate median filter based method to detect frame difference.

For every pixel location (i, j) the co-ordinate of frame

$$FD_n(i,j) = WI_n(i,j) - WI_{n-1}(i,j) \qquad \ldots \ (1)$$

Where $WI_n(i,j)$ and $WI_{n-1}(i,j)$ are wavelet coefficients of frame $I_n(i,j)$ and $I_{n-1}(i,j)$ respectively.

Obtained result may have some noise. Applying soft thresholding to remove noise. In presence of noise, equation (1) is expressed as:

$$FD_n(i,j) = FD_n(i,j) - \lambda \qquad \ldots (2)$$

Where $FD_n(i,j)$ is frame difference without noise, λ represents corresponding noise components. For de-noising, soft thresholding technique in wavelet domain is used for estimation of frame difference $FD_n(i,j)$. After noise removal, Sobel edge detection operator is applied on $FD_n(i,j)$ to detect the strong edges of significant difference pixels in all sub-bands as follows:

$$DE_n(i,j) = sobel\big(FD_n(i,j)\big) \qquad \ldots (3)$$

After finding edge map $DE_n(i,j)$ in wavelet domain, inverse wavelet transform is applied to get moving object segmentation in spatial domain i.e. E_n. The obtained segmented object may include a number of disconnected edges due to non-ideal segmentation of moving object edges. Therefore, some morphological operation is needed for post processing of object edge map to generate connected edges. Here, a binary closing morphological operation is used. After applying the morphological operator $M(E_n)$ is obtained which is the segmented moving object, and finally temporal updating of the background model is needed in order to adapt the changes in background and in lighting conditions as follows:

$$If\ \big(I_n(i,j) > I_{n-1}(i,j)\big) \qquad \ldots (4)$$

$$I_{n-1}(i,j) = I_{n-1}(i,j) + 1 \qquad \ldots (5)$$

$$else\ \ I_{n-1}(i,j) = I_{n-1}(i,j) - 1 \qquad \ldots (6)$$

Here, $I_n(i,j)$ is the value of $(i,j)^{th}$ pixel of n^{th} frame and $I_{n-1}(i,j)$ is the value of $(i,j)^{th}$pixel of $(n-1)^{th}$ frame [6].

6. COMPARISON

Table 2.1 Comparison of human detection techniques in terms of accuracy and computational time

Methods	Accuracy	Computational time	Comments
Background Subtraction	Moderate	Low to moderate	Its calculations are simple and easy to implement. It has strong adaptability for variety of dynamic environment. It is difficult to obtain a complete outline of moving objects. i.e. accuracy is less as compare to optical flow method.
Optical flow	Moderate	High	Large quantities of calculations are needed. It is sensitive to noise, poor anti-noise performance, not suitable for real time demanding occasions. In this method we get complete movement information and we detect the moving object.
Frame differencing and template matching	Moderate to high	Low to moderate	Motion detection using frame differencing method provides a better result for the object tracking and which can be easily applied to a number of fields. It helps to keep the track of movement of object which can be accessed later to analyse the motion.
Background subtraction using Neural mapping	High	High	Unlike existing methods that uses individual flow vectors as inputs, this method learns background motion trajectories in a self-organizing manner; this makes the neural network structure much simpler.
Background subtraction using wavelet transform	Moderate	Moderate	This method performs better in comparison to other methods. It also capable of alleviating the problems associated with other spatial domain methods such as ghosts, clutters, noises etc.

Ⅲ. CONCLUSION

In this paper, we have seen various moving object detection techniques in visual surveillance systems. The various Detection techniques in Video Surveillance system such as

a) Background subtraction
b) Optical flow
c) Frame differencing and template matching
d) Background subtraction using neural network
e) Background subtraction using wavelet transform

are studied and discussed in this paper. We compared these techniques in terms of accuracy and computational time, also we given a remark on that.

Although a large amount of work has been done in visual surveillance for detecting objects and their classification, many issues are still open and deserve further research.

REFERENCES

[1] Manoranjan Paul, Shah M E Haque and Subrata Chakraborty, "Human detection in surveillance videos and its applications - a review", EURASIP Journal on Advances in Signal Processing 2013.

[2] Weiming Hu, TieniuTan,Liang Wang and Steve Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors",1094-6977/04 © 2004 IEEE, pp. 334–352, 2004.

[3] Lijing Zhang, Yingli Liang, "Motion human detection based on background subtraction", 978-0-7695-3987-4 © 2010 IEEE, pp. 284-287, 2010.

[4] N. Prabhakar, V. Vaithiyanathan, Akshaya Prakash Sharma, Anurag Singh, Pulkit Singhal, "Object Tracking Using Frame Differencing and Template Matching", Res. J. Appl. Sci. Eng. Technol., 4(24): 5497-5501, pp. 5497-5501, 2012.

[5] Lucia Maddalena and Alfredo Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications", 1057-7149 © 2008 IEEE, pp. 1168-1177, 2008.

[6] Alok Kumar Singh Kushwaha, Rajeev Srivastava, "Complex Wavelet Based Moving Object Segmentation Approximate Median Filter Based Method For Video Surveillance", 978-1-4799-2572-8 © 2014 IEEE, pp. 973-978, 2014.