CHAPTER 4

ESTIMATION AND SUBTRACTION OF THE BACKGROUND FROM VIDEO OBJECTS

4.1 INTRODUCTION

Tracking a moving object is one of the challenging tasks in computer vision. The conventional approach to object tracking is based on the difference between the current image and the background image. People counting includes background estimation, background subtraction, segmentation and tracking. This chapter is focused on the first two problems of the People Counting system, namely, background extraction and subtraction. To estimate the background in real time is the first important step for many video surveillance applications. Background extraction faces numerous challenges due to low resolution, more foreground color and scene which always contains some foreground elements and a complex background. Once a background model is established, the humans in the video frames can be detected as the difference between the current video frame and the background model.

4.2 BACKGROUND ESTIMATION

Recently, researchers both in India and abroad have presented some classical algorithms of background Estimation, including the mean algorithm, median algorithm, determination of stable interval algorithm, the detection of changes algorithm and the mode algorithm. In addition, there are some highly
complex methods that have been used in the background estimation algorithm, such as approximated median filtering, mixture of the Gaussians model, progressive background estimation method and histogram/group-based histogram approaches. The various background estimation methods are briefly described here.

The Temporal and Approximated Median Filtering (AMF) methods are based on the assumption, that pixels related to the background scene would be present in more than half the frames of the entire video sequence. In some cases the number of stored frames is not large enough (buffer limitations); therefore, the basic assumption will be violated and the median will estimate a false value, which has nothing to do with the real background model. The AMF (McFaralane and Schofield 1995) applies the filtering procedure, by simply incrementing the background model intensity by one, if the incoming intensity value (in the new input frame) is larger than the previous existing intensity in the background model. In fact this approach is most suitable for indoor applications.

A MOG model is designed such that the foreground segmentation (Stauffer and Grimson 1999) is done by modeling the background and subtracting it out of the current input frame, and not by any operations performed directly on the foreground objects (i.e. directly modeling the texture, colour or edges). The processing is done pixel by pixel rather than by region based computations, and finally the background modeling decisions are made based on each frame itself, instead of benefiting from tracking information or other feedbacks from the previous steps.

The Progressive Background Estimation Method (Chung et al 2002) constructs the histograms from the preprocessed images known as the partial backgrounds. Each partial background is obtained using two consecutive input frames. This method is applicable to both gray scale and
colour images, and is capable of generating the background in a rather short period of time, and does not need large space for storing the image sequences.

The Group-Based Histogram (GBH) algorithm constructs background models by using the histogram of the intensities of each pixel on the image. Since the maximum count (amplitude) of the histogram is much greater in comparison to the frequencies of the intensities related to the moving objects, there will not be any effects of slow moving objects or transient stops in the detected foreground.

From the above discussion, the characteristics of the various estimation methods are observed and summarized as follows:

- All the methods deal only with non moving pixels.
- It takes a longer time to create a background model.

In this work, it is decided to apply the improved mode method to have a unified and common framework for all kinds of images. To achieve this, moving pixels in the frame are to be handled.

4.3 BACKGROUND SUBTRACTION

Background subtraction is a general term for a process which aims to separate foreground objects from a relatively stationary background. Background Subtraction generates a foreground mask for every frame. This step is simply performed by subtracting the background image from the current frame. When the background view excludes the foreground objects, it becomes obvious that the foreground objects can be obtained by comparing the background image with the current video frame. By applying this approach to each frame, one can effectively achieve the tracking of any moving object.
A background image can be used to determine the foreground objects, by comparing the input frame with the background image and marking the differences as foreground objects. This technique is commonly known as background subtraction or change detection. It is the most popular approach in video surveillance applications, because it is a computationally efficient technique, and is relatively easy to obtain background images from static surveillance cameras. In practice, camera noise and regions in which the object has the same color as the background make the separation of the foreground objects and the background more difficult. In addition, there are some highly complex methods that have been used in background subtraction, such as the pfinder, codebook and MOG. Background subtraction algorithm was intended to sample values over long periods, without making parametric assumptions. Mixed backgrounds can be modeled by multiple codewords.

The Pfinder uses a simple scheme, where the background pixels are modeled by a single value, and the foreground pixels are explicitly modeled by a mean and covariance, which are updated recursively. It requires an empty scene at start-up (Wren et al 1997).

Over time, different background objects are likely to appear at the same pixel location. When this is due to a permanent change in the scene's geometry, all the models reviewed so far will, more or less promptly, adapt so as to reflect the value of the current background object. However, sometimes the changes in the background object are not permanent, and appear at a rate faster than that of the background update.

The observations from Various Background Subtraction Methods are:

- Shadow effect is not discussed in all the methods.
- Illumination variations can occur in the scene and generate a false classification of image pixels.
Here, the proposed system is designed to estimate the background from a sequence of frames, and a pure background image is obtained. Using this background image the current frame is subtracted to get a segmented foreground image. To design a common framework for background subtraction from current images, appropriate methods should be chosen for subtracting. The major contributions of the background estimation / subtraction method are as follows:

- Background estimation using the improved mode algorithm for getting a pure background.
- Background subtraction using a fuzzy based system for getting the foreground mask by separating the foreground from the background.

### 4.4 SYSTEM DESCRIPTION

In this section, the processing steps of the proposed background estimation/subtraction approach are presented. The aim is to build an automatic background estimation and subtraction, which is able to accept different type of an image sequence.

Instead of using the trial-and-error procedure, a common preliminary step in any motion-based vision application is used, to obtain image frames from a stationary or a moving camera, or from multiple cameras. After the acquisition of the image frames, image segmentation can be performed using background subtraction, depending on the frame rate of the video sequences.

For high frame rate sequences, the adjacent frame subtraction is used since the change of motion between the consecutive frames is very small. This method eliminates the stationary background, leaving only the
desired motion regions. In certain video systems, the acquisition process may be susceptible to erratic changes in illumination, reflection, and noise. The block diagram of the proposed methodology is shown in Figure 4.1.

![Block diagram of Background Estimation and subtraction Method](image)

**Figure 4.1** Block diagram of Background Estimation and subtraction Method

The proposed method consists of the following stages: Background estimation, and background subtraction.

### 4.4.1 Background Estimation

Background estimations is the process of distinguishing novel (foreground) from non-novel (background) elements, in a scene from a video sequence. The improved mode algorithm is applied to an input video. This method consists of two steps: First, the input video is converted into frames. Second, the background is estimated automatically from the successive frames.

Current approaches for background estimation consider the pixel to be a non-moving object. The motivation for the improved mode algorithm is to use the distinguishing pixel to separate the unchanging background and the moving object. The frame differences method classifies the pixels into two: an
unchanging background and the moving objects, and then it calculate the unchanged background pixels through the mode algorithm.

In the Frame difference algorithm, if there is a moving object in the video, its grey level will change significantly between two frames. \( I_n(x, y) \) is the value of pixel at \((x, y)\) in frame \( t=t_n \). Likewise, \( I_{n+1}(x, y) \) is the value of pixel at \((x, y)\) in frame \( t=t_{n+1} \). The simple difference image \( D(x, y) \) between these two frames is:

\[
D(x,y)=| I_{n+1}(x,y) – I_n(x,y) | \quad \forall (x,y) \in [1,N]X[1,M] \tag{4.1}
\]

Where \( N \times M \) is the image frame dimension. Applying a suitable threshold \( T \) on \( D(x, y) \) results in a binary image, that classifies all pixels into two classifications: an unchanged background and moving objects.

\[
BW_n(x,y) = \begin{cases} 0= \text{unchanged background} & \text{if } D(x,y)<T \\ 1= \text{moving object} & \text{otherwise} \end{cases} \tag{4.2}
\]

After image binarization, applying the opening and closing of the mathematical morphology on \( BW_n(x,y) \) results in a devoicing image and still saves the result in \( BW_n(x,y) \).

\[
B_{\text{back}}_n(x,y) = \begin{cases} 1 & \text{if } BW_n(x,y)=0 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}
\]

\( B_{\text{back}}_n(x,y) \) is taken to mark the value of pixel at \((x, y)\) whether it is valid or not. If the value is available, then \( B_{\text{back}}_n(x,y) =1 \); this implies that the value of pixel at \((x, y)\) can be used in the calculation of the mode algorithm. Otherwise, \( B_{\text{back}}_n(x,y) \) should be equal to 0(a flag of unavailable value), and it should be avoided in the calculation.
The Mode Algorithm is used in 2-D image sequences. For every pixel at \((x, y)\), the corresponding point values in the previous \(N\) frames are: \(B_{t-N}(x, y), B_{t-N+1}(x, y), B_{t-N+2} \ldots\)

\(B_{t-2}(x, y), B_{t-1}(x, y)\) the sequence of values through the mode algorithm is calculated and the background value of the current image is taken as the result. The computing formula of the background value is:

\[
B(x, y) = \text{mode}(I_{t-N}(x, y), I_{t-N+1}(x, y) \ldots I_{t-2}(x, y), I_{t-1}(x, y)) \tag{4.4}
\]

When a moving object is stationary for a long time, the mode algorithm fails in estimating the background so the moving object is identified as background. To eliminate the deficiency of the mode algorithm, an improvement is done on it. The computing formula of the new method is:

\[
B_{G}(x, y) = \text{mode}(B_{t-N}(x, y) X\alpha_{i-N}, B_{t-N+1}(x, y) \alpha_{i-N+1}, B_{t-N+2}(x, y) X\alpha_{t-N+2} \ldots B_{t-2}(x, y) X\alpha_{t-2}, B_{t-1}(x, y) X\alpha_{t-1}) \tag{4.5}
\]

\[\alpha_n = \begin{cases} 1 & \text{if } B_{\text{back}}(x, y) = 1 \\ 0 & \text{otherwise} \end{cases}\]

\(\alpha_n\) determine the pixel (moving or background).

### 4.4.2 Background Subtraction

Some approaches achieve detection using only background subtraction, and predicting the background through the next update interval. In these approaches the background is not estimated but detected. Here, the background image was generated by the improved mode algorithm. In the standard background method, it is hard to determine whether a pixel is a moving object or not. As shadows, illumination variations can occur in the scene and generate a false classification of the image pixels. The fuzzy logic
inference system fuses multiple sources of information together for decision-making. To determine the foreground object, the proposed binarization of the fuzzy background subtraction is done, after passing through the median filter. The proposed method detects moving objects even if their gray level is similar to the background gray level. In addition, it can remove small noise because of the median filter.

In the proposed adaptive background modeling and classification scheme, the image data in the past frames is used to compute the joint distribution of the images to build a background model. Based on the background model, the image block is classified as the foreground or the background. If a new object is introduced into the background or a background object is moved, before it is updated, this object will be classified as a foreground object and hence becomes part of the silhouette. To solve this problem, high-level knowledge about the object motion is utilized to guide the adaptive update of the background model.

The main contribution of this thesis is a methodology to detach moving objects. Detaching objects from the adaptive background is a challenging task because of the lack of automated reasoning about what is, and what is not, a target object, as evidenced in a video sequence.

In this case, sophisticated object recognition and identification algorithms could be employed. However, these algorithms are often computationally intensive and not robust in a dynamic environment. Therefore, a simple and efficient algorithm for object segmentation is proposed. The scheme is based on a fuzzy logic inference system, which fuses multiple sources of information together for decision making. Suppose that one is working on frame $n$ and the object in frame $n-1$ has been correctly extracted. Let the foreground image region in frame $n$ be $O_n$, which might
contain the human body and moving objects. The fuzzy logic inference system is based on the following observations:

1. If an image block in $O_n$ belongs to the object, it should have a high possibility of finding a good match in $O_{n-1}$. The sum of absolute difference (SAD) is used to measure the “goodness” of matching.

2. If many of the blocks in its neighborhood have good matches in $O_{n-1}$, it is highly possible that this block also belongs to the object.

3. If this block is far from the predicted position of the object centroid, the possibility that this block belongs to the object is low.

4. SAD in motion matching. For each block in $O_n$, the best match in frame $n-1$ is found.

5. The SAD between this block and its best match forms the first feature variable is the distance between the new block and the predicted object centroid.

This algorithm is based on image differencing techniques. It is mathematically represented using the following equation:

$$D(t) = \frac{1}{N} \sum_{i} |I(t_i) - I(t_j)|$$

(4.6)

where $N$ is the number of pixels in the image used as a scaling factor, $I(t_i)$ is the image $I$ at time $i$, $I(t_j)$ is the image $I$ at time $j$ and $D(t)$ is the normalized SAD for that time.

In an ideal case, when there is no motion
\[ I(t_i) = I(t_j) \]  \hspace{1cm} (4.7)

and \( D(t) = 0 \). However, noise is always present in images and a better model of the images in the absence of motion will be

\[ I(t_i) = I(t_j) + n(p) \]  \hspace{1cm} (4.8)

where \( n(p) \) is a noise signal.

The value \( D(t) \), which represents the normalized SAD can be used as a reference to be compared with a threshold value. To determine the membership functions of these feature variables, a set of membership functions were defined from these distributions. A fuzzy inference system has the following parameters:

Rule 1 : If SAD is medium, AND the Neighborhood is small, THEN the Object is low

Rule 2 : If SAD is high, AND the Neighborhood is small, THEN the Object is high

Rule 3 : If SAD is high AND the Neighborhood is high, THEN the Object is high

According to the above rules, if an object is recognized as the foreground, it is observed that using the fuzzy inference system to detach the moving objects will erode the object due to misclassification. The proposed binarization of the fuzzy background subtraction, after passing morphological operations and neighborhood information, would find out these missing parts. Morphological operations are performed on the motion image. Such noise reduction is a typical pre-processing step to improve the results. The median filter can eliminate the effect of input noise values with extremely large magnitudes. The removal of the motion at the boundary is also accomplished.
Pixels of the motion region located along the boundary are eliminated to avoid ambiguity of the region as belonging to a possible moving object.

Finally, the segmented object is detected in the block. There is no moving object in the scene. Therefore, this object can be treated as an optimal object for background subtraction. Let the pixel coordinates of the reference background image and the scene image at frame, be gray levels.

4.4.3 Algorithm: Background Estimation/Subtraction

The automatic background estimation starts by processing the estimated background from the current image as in the following steps: According to the discussion given above, the various steps of the new method are as follows:

**Step 1** A movie is taken and converted into a number of successive frames (images).

**Step 2** Two frames are obtained at fixed intervals from the video and saved in $I_n(x, y)$ and $I_{n+1}(x, y)$. For the purpose of simple calculation and real-time speed, these two frames should be converted into grey images.

**Step 3** Using $I_n(x, y)$ and $I_{n+1}(x, y)$ the frame difference image can be obtained as $D_n(x, y)$. And then, the opening and closing operation of the mathematical morphology on $D(x, y)$ is applied and the computed result using equation 4.2 is saved in $BW_n(x, y)$.

**Step 4** According to the value of pixels in $BW_n(x, y)$, the pixels are classified into either background moving objects or unchanging objects. According to equation 4.3, make a flag as the pixel whether it is the moving objects or background. If the flag value is 1, it implies that the pixel belongs
to a moving object and it’s unavailable in mode calculation; otherwise its background is available. These values of the flag are saved in $B_{\text{back}}(x, y)$.

**Step 5)** If $n$ reaches the maximum set up, the procedure goes on to step 6; else the procedure should go to step 1.

**Step 6)** $B_n(x, y)$ should be calculated including the video and saved as $B(x, y, z)$, namely, the values of the background of all frames. Through the steps above, the background image can be estimated accurately. As a result of the pixels of moving objects being removed, even if the pixel background value just emerges once, it can be estimated accurately by the new method.

**Step 7)** The estimated background should be subtracted from the current image and the resulting image is filtered.

**Step 8)** Fuzzy based background subtraction is applied and the segmented foreground image is obtained from the input image.

**Step 9)** The resulting foreground image is compared with the ground truth image and the accuracy of the detected image using different metrics, is evaluated.

**Step 10)** The fuzzy based background subtraction method is compared with the other methods.

4.5 **RESULTS AND PERFORMANCE ANALYSIS**

4.5.1 **Experimental Setup**

The proposed algorithm was implemented in MATLAB and various metrics have been evaluated from the test results. These data set images have been gathered from the sites of several research groups.

- EC Funded CAVIAR Project, IST 2001
For sequences belonging to Toyama et al (1999), the ground truth is available as a binary detection mask for one reference frame.

http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

4.5.2 Ground Truth

This concept is used as the base for the quantitative comparisons. Ground truths are a special kind of video sequences, which contain only the desired moving objects of the scene (ideal foreground detection). Animated moving objects are superimposed manually on the recorded background scenes. Therefore the exact location of the pixels related to the foreground items is known. Another advantage of using this kind of sequences is that since the super imposed objects do not contain shadows, they can only focus on the performance of background detection instead of dealing with shadow removal algorithms.

4.5.3 Performance Analysis

The metrics used to evaluate the performance of the system are Precision, Recall, F-score and Similarity. Precision, recall and similarity have been computed, based on the number of correctly detected pixels in an image, in order to evaluate the efficiency and robustness of the algorithm. The next step would be taking advantage of practical scales in order to help us compare the achieved results with the ground-truths.

4.5.3.1 Classification of Pixels

Prior to the details of recall and precision definitions, certain concepts should be explained. These concepts include classifying the pixels into 4 different groups:
1. True Positive (TP): which represents the number of foreground pixels, correctly detected by the algorithm.

2. False Positive (FP): is responsible for the number of pixels, which are incorrectly classified as foreground objects.

3. True Negative (TN): indicating the number of background pixels, which are correctly detected as the background scene by the algorithm.

4. False Negative (FN): stands for the number of pixels corresponding to the foreground objects which are misclassified as part of the background image.

Definition 1: The recall is defined as the number of true positives divided by the total number of elements that actually belong to the foreground objects. (i.e, the sum of both the true positives and false negatives).

\[
\text{Recall} = \frac{tp}{tp + fn} \quad (4.9)
\]

Definition 2: Precision can be considered as a measure of exactness or fidelity, and is evaluated through dividing the number of items (foreground objects) correctly detected by the total number of pixels classified as the foreground by the algorithm.

\[
\text{Precision} = \frac{tp}{tp + fp} \quad (4.10)
\]

Definition 3: The figure of Merit or F-measure, that is the weighted harmonic mean of Precision and Recall

\[
F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4.11)
\]
Definition 4: Such a measure allows us to obtain a single measure that can be used to “rank” the different methods. Finally, the pixel-based Similarity measure, is defined as

\[
\text{Similarity} = \frac{tp}{tp + fn + fp} \tag{4.12}
\]

4.5.3.2 Experimental Results

The experimental results on several sequences show that the proposed fuzzy inference system for moving objects detachment not only preserves the advantage of the low-level feature-based object extraction algorithm but also verifies the robustness of the proposed detaching algorithm. The fuzzy system also has the added benefit of removing some of the dark shadows that the colour-based algorithm fails to eliminate. However, some of the hardened silhouettes appear slightly eroded.

Since our goal is to perform an activity analysis, the eroded object should not cause serious problems, as the proposed algorithm fills out the fuzzy shape. While the fuzzy logic system performs better than the corresponding crisp one, a good portion of the adaptive background is lost.

Each pixel in a background subtraction method classification was determined to be as follows: \( tp \) for a correctly classified foreground pixel, \( fp \) for a background pixel that was incorrectly classified as the foreground, \( tn \) for a correctly classified background pixel, and \( fn \) for a foreground pixel that was incorrectly classified as the background.

Results obtained by the proposed fuzzy inference system have been compared with those obtained by other existing algorithms, in terms of
different metrics. The compared methods, the adopted metrics and the accuracy results will be briefly described in the following sections. This metric has been adopted with the only objective of further comparing the results achieved by the proposed fuzzy inference system with the other methods (Culibrk et al 2007).

### 4.5.4 Comparison with other Background subtraction Approaches

For the sequences (Toyama et al 1999), the ground truth is available as a binary detection mask for one reference frame. Here, the fuzzy method attains accuracy values generally higher than those of all the other methods. For sequence WT, the Foreground masks obtained by the fuzzy method are shown in Figure 4.2, and the metrics values are reported in Table 4.1 and Figure 4.3. Specifically, the fuzzy method performs slightly better than all the other methods. Finally, for sequence IR, the ground truth is available as a binary detection mask for 113 reference frames. The foreground masks obtained by the fuzzy method are reported in Figure 4.4. The Average pixel based accuracy values over all reference images are reported in Table 4.2 and Figure 4.5, where it can be observed that the fuzzy method attains accuracy.

![Figure 4.2](image)

**Figure 4.2** Segmentation of sequence WT: (a) background image generated by the proposed extraction method; (b) test image; (c) ground truth; (d) Fuzzy result.
Table 4.1 Pixel-Based Accuracy Values For Sequence WT

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOBs</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>Pfinder</td>
<td>0.67</td>
<td>0.97</td>
<td>0.79</td>
<td>0.66</td>
</tr>
<tr>
<td>VSAM</td>
<td>0.99</td>
<td>0.94</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>CB</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>0.98</td>
<td>0.93</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Figure 4.3  Precision, Recall, F-score and similarity bar chart for Fuzzy logic and various methods for sequence WT

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

Figure 4.4 Segmentation of sequence IR: (a) background image generated by the proposed extraction method b) test image; (e) ground truth; (d) Fuzzy result.
Table 4.2 Average Pixel-Based Accuracy Values for Sequence IR

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOBs</td>
<td>0.84</td>
<td>0.95</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>Pfinder</td>
<td>0.82</td>
<td>0.86</td>
<td>0.83</td>
<td>0.72</td>
</tr>
<tr>
<td>VSAM</td>
<td>0.81</td>
<td>0.91</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>CB</td>
<td>0.75</td>
<td>0.93</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>0.84</td>
<td>0.96</td>
<td>0.89</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Figure 4.5 Precision, Recall, F-score and similarity bar chart for Fuzzy logic and various methods for Sequence IR

Further comparisons have been made with the background subtraction results (Culibrk et al 2007). Employ a feed-forward neural network to achieve background subtraction, named BNN, that is a combination of a probabilistic neural network and a winner-take-all neural network, with the addition of rules for the temporal adaptation of the network weights, based on a Bayesian formulation of the segmentation problem. They compared the results achieved by the BNN with those obtained by a MOG and their method (Li et al. 2004). The MOG method uses multiple Gaussian
distributions as a model for the values of the background pixels. The Gaussian distributions are then evaluated to determine which are the most likely to result from a background process.

The model (Li et al. 2004) of the probability density function detects the pixel values by a filter-based initial segmentation, in order to distinguish between the errors in the initial segmentation and the true foreground pixels. The probability density functions are updated in time and a Bayes’rule-based decision framework is formulated, based on the assumption that the pixel values observed more often at a single pixel are more likely to be due to the background object movement.

For comparison, consider the image sequences (Li et al 2004) publicly available at http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html. For all the sequences, the hand-segmented background is given for twenty test frames, randomly chosen along the sequence. The qualitative results in terms of the foreground masks obtained by the fuzzy algorithm on one of the test frames for each sequence, are shown in Figures. 4.6–4.10, and can be readily compared with those methods (Culibrk et al 2007). The fuzzy method parameter values that are not common to all the other reported experiments are detailed in Table 4.3 and Figure 4.11.

![Figure 4.6](image)

Figure 4.6 Segmentation of sequence CAM. ((a) background image generated by the proposed extraction method b) test image; (c) ground truth; (d) Fuzzy result.)
Figure 4.7  Segmentation of sequence FT. (a) background image generated by the proposed extraction method b) test image; (c) ground truth; (d) Fuzzy result.

Figure 4.8  Segmentation of sequence WS. (a) background image generated by the proposed extraction method b) test image; (c) ground truth; (d) Fuzzy result

Figure 4.9  Segmentation of sequence MR. (a) background image generated by the proposed extraction method b) test image; (c) ground truth; (d) Fuzzy result.
Figure 4.10  Segmentation of sequence LB. (a) background image generated by the proposed extraction method b) test image; (c) ground truth; (d) Fuzzy result.

Table 4.3  Comparison Of the Similarity Values for the fuzzy and other methods (Maddalena and Petrosino 2008)

<table>
<thead>
<tr>
<th></th>
<th>CAM</th>
<th>FT</th>
<th>WS</th>
<th>MR</th>
<th>LB</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOBs</td>
<td>0.6960</td>
<td>0.6554</td>
<td>0.8247</td>
<td>0.8178</td>
<td>0.6489</td>
</tr>
<tr>
<td>BNN</td>
<td>0.5256</td>
<td>0.4636</td>
<td>0.7540</td>
<td>0.7368</td>
<td>0.6276</td>
</tr>
<tr>
<td>MOG</td>
<td>0.0757</td>
<td>0.6854</td>
<td>0.7948</td>
<td>0.7580</td>
<td>0.6519</td>
</tr>
<tr>
<td>Li et al</td>
<td>0.1596</td>
<td>0.0999</td>
<td>0.0667</td>
<td>0.1841</td>
<td>0.1554</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>0.7010</td>
<td>0.6710</td>
<td>0.9140</td>
<td>0.8168</td>
<td>0.6561</td>
</tr>
</tbody>
</table>

Figure 4.11  Similarity bar chart for Fuzzy logic and various methods for sequence CAM, FT, WS, MR, LB
The performance measures for comparison are shown in Tables 4.1, 4.2 and 4.3. It is observed from these tables that

1) **PIXEL-BASED ACCURACY VALUES FOR SEQUENCE WT:** The proposed method produces 98% of recall, 93% of precision, 0.98 of F-score and 97% of similarity. When compared to the SOB method, the precision value is of poor quality; this is shown in Figure 4.2.

2) **PIXEL-BASED ACCURACY VALUES FOR SEQUENCE IR:** The proposed method produces 84% of recall, 96% of precision, 89% of F-score, and 81% of similarity. When compared to the SOB method, the precision and similarity values are of good quality. This is shown in Figure 4.3.

3) **Similarity values for the sequence of perception:** The proposed method produces better similarity values compared with the other methods; this is shown in Figures 4.4-4.10.

So, it is concluded from the results that, in all cases, the proposed method is found to be better than the other methods.

### 4.6 SUMMARY

In this section, a novel method of video background estimation, based on the improved mode algorithm and fuzzy approach-based background subtraction for video object segmentation is presented. The algorithm has been applied on several images. The experimental results testify that the method proposed, can estimate background images faster and better. It also includes a comprehensive accuracy testing and can be performed with both pixel-based and frame-based metrics. The Experimental results, using different sets of data and comparing different methods, have demonstrated the effectiveness of the proposed approach. Good segmentation results have been
obtained for these complex sequences. The proposed approach represents an improvement in the segmentation ability when compared to a well-known pure probabilistic approach. During the experiments, it was observed that the improved mode background estimation method works on different kinds of videos. The fuzzy based background subtraction method produces better segmented objects.