SIViP (2015) 9:635–650
DOI 10.1007/s11760-013-0496-4

ORIGINAL PAPER

Moving object segmentation in Daubechies complex wavelet domain

Manish Khare · Rajneesh Kumar Srivastava · Ashish Khare

Received: 25 June 2012 / Revised: 31 January 2013 / Accepted: 30 April 2013 / Published online: 21 May 2013

Abstract Motion segmentation is a crucial step in video analysis and is associated with a number of computer vision applications. This paper introduces a new method for segmentation of moving object which is based on double change detection technique applied on Daubechies complex wavelet coefficients of three consecutive frames. Daubechies complex wavelet transform for segmentation of moving object has been chosen as it is approximate shift invariant and has a better directional selectivity as compared to real valued wavelet transform. Double change detection technique is used to obtain video object plane by inter-frame difference of three consecutive frames. Double change detection technique also provides automatic detection of appearance of new objects. The proposed method does not require any other parameter except Daubechies complex wavelet coefficients. Results of the proposed method for segmentation of moving objects are compared with results of other state-of-the-art methods in terms of visual performance and a number of quantitative performance metrics viz. Misclassification Penalty, Relative Foreground Area Measure, Pixel Classification Based Measure, Normalized Absolute Error, and Percentage of Correct Classification. The proposed method is found to have high degree of segmentation accuracy than the other state-of-the-art methods.

M. Khare · R. K. Srivastava · A. Khare (✉)
Department of Electronics and Communication,
University of Allahabad, Allahabad, India
e-mail: ashishkhare@hotmail.com
e-mail: khare@allduniv.ac.in

M. Khare
e-mail: mkharejk@gmail.com

R. K. Srivastava
e-mail: rkumarsau@gmail.com

Keywords Double change detection · Daubechies complex wavelet transform · Moving object segmentation

1 Introduction

Moving object segmentation is an important step for development of any computer vision system [1]. Segmentation of moving object is a process of isolation of foreground object from background of a video sequence. Segmentation is a complicated task due to following reasons [2].

1. Changing background.
2. Presence of noise and blur in video.
3. Shape and size of object may vary from frame to frame.
4. Varying lighting conditions.
5. Abrupt motion of object.
6. Presence of occlusion.

In comparison with the literatures available on image segmentation [3], limited literatures are available on moving object segmentation. As per available literatures, moving object segmentation techniques can be broadly classified into following four groups [4,5].

1. Segmentation of moving object based on motion information [6–8].
2. Segmentation of moving object based on motion and spatial information [9–12].
3. Segmentation of moving object based on learning [13, 14].
4. Segmentation of moving object based on change detection [15–20].

Motion information-based moving object segmentation algorithms are closely related with motion estimation. Kim et al. [6] have proposed a real-time foreground–background segmentation method using codebook. In this method, background values at each pixel are quantized into codebook, and the codebook represents a compressed form of background model for an image sequence. Xiaoyan et al. [7] have proposed a video object segmentation technique on the basis of adaptive change. This method is not able to remove noise from the video frames. Mahmoodi [8] has proposed a shape-based active contour method for video segmentation which is based on a piecewise constant approximation of the Mumford shah functional model. This method is slow as it is based on level set framework. Due to the lack of spatial information of object, these algorithms suffer from occlusion problem, and therefore, these algorithms are not able to detect accurate boundaries of segmented object.

Merging spatial information with motion information will provide more stable object boundary extraction. Javed et al. [9] have proposed a segmentation method that uses Gaussian mixture model (GMM) with color and gradient feature vectors to make the process of background subtraction more robust against illumination changes. GMM-based technique can be effectively used for scene modeling, but it suffers from the problems of slow convergence and false motion prediction. Reza et al. [10] have proposed a moving object segmentation technique, combining temporal and spatial features. This approach takes into account a current frame, ten preceding frames and ten next consecutive frames. All the twenty-one frames are processed to detect and remove the static part of motion window resulting in the segmented object. The method detects moving objects independent of their size and speed, but there is no provision for reduction in noise from frames, which may lead to inaccurate object segmentation. Ivanov et al. [11] have proposed an improvement over background subtraction method, which is faster than that proposed by Reza et al. [10] and is invariant to runtime change illuminations. Colombari et al. [12] have proposed a method for segmentation of multiple video objects by combining content-based representation with background modeling. This method uses projective transformation of objects from one frame to another.

Learning-based moving object segmentation algorithms require some predefined learning patterns. Kato et al. [13] have proposed a segmentation method for monitoring of traffic video based on hidden Markov model (HMM). In this method, each pixel or region is classified into three categories: shadow, foreground, and background. This method comprises of two phases: learning phase and segmentation phase. Stauffer et al. [14] have proposed a tracking method using learning patterns of observation. In this method, motion segmentation has been done by adaptive background subtraction that models each pixel as a mixture of Gaussians.

Strategy of change detection algorithm is based on computation of frame difference of two or more frames. The obtained frame difference is further processed to identify moving objects. Change detection method for segmentation gives feasible solutions because it enables automatic detection of new appearances [20]. Huang et al. [17] have proposed an algorithm for moving object segmentation using double change detection applied in wavelet domain. Baradarani [18,19] have refined the work of Huang et al. [17] using 9/7–10/8 dual tree complex filter bank in wavelet domain algorithm. A robust scene change detection method for video object segmentation that combines the intensity and motion information has been proposed by Huang and Liao [16].

All the methods discussed so far, for segmentation of moving objects, are suffering from the problem of either inaccurate segmentation due to non-removal of noise or failure to detect new appearance automatically. Motivated by these facts and work of Baradarani [18,19], we have proposed a new method for segmentation of moving object using double change detection method applied on Daubechies complex wavelet transform. The Daubechies complex wavelet transform is having advantages of shift invariance and better directional selectivity as compared to real valued discrete wavelet transform. The proposed method consists of the following six steps.

1. Wavelet decomposition of three consecutive frames using Daubechies complex wavelet transform.
2. Double change detection method applied in complex wavelet domain.
3. Denoising of frames
4. Canny edge detection applied on denoized frames
5. Detection of moving edges
6. Detection of moving object by using detected moving edges.

The proposed method has been compared with the following state-of-the-art methods.

1. Method proposed by Kim et al. [6] that uses codebook for segmentation.
2. Method proposed by Mahmoodi [8] that is based on active contour.
3. Method proposed by Huang et al. [17] that is based on change detection in wavelet domain.
4. Method proposed by Baradarani [18,19] that is based on change detection in wavelet domain.

Performance of the proposed method is found better in terms of visual performance and a number of quantitative performance metrics viz. Misclassification Penalty (MP), Relative Foreground Area Measure (RFAM), Pixel Classification based Measure (PCM), Normalized Absolute Error (NAE), and Percentage of Correct Classification (PCC).

Rest of the paper is organized as follows: Sect. 2 describes basic introduction of Daubechies complex wavelet transform and its properties useful for motion segmentation. Section 3 explains the proposed method in detail. Experimental results and performance evaluation are given in Sects. 4 and 5, respectively. Finally, conclusions are given in Sect. 6.

2 Daubechies complex wavelet transform

In the proposed work, we have used Daubechies complex wavelet transform with its reduced shift sensitivity and edge information properties.

2.1 Construction of Daubechies complex wavelet

The basic equation of multiresolution theory is the scaling equation

$$\phi(u) = 2\sum_i a_i \phi(2u - i) \tag{1}$$

where a_i's are coefficients, and $\phi(u)$ is the scaling function. The a_i's can be real as well as complex valued and $\sum a_i = 1$. Daubechies' wavelet bases $\{\psi_{j,k}(t)\}$ in one-dimension are defined through the above-mentioned scaling function and multiresolution of $L_2(\Re)$ [21]. For formulation of general solution, Daubechies considered a_i's to be real valued. If we relax the condition for a_i's to be real valued, it leads to Daubechies complex valued scaling function. Construction of Daubechies complex wavelet transform is reported in [21,22].

The generating wavelet $\psi(t)$ is given by

$$\psi(t) = 2\sum_n (-1)^n \overline{a_{1-n}} \phi(2t - n) \tag{2}$$

Any function $x(t)$ can be decomposed into complex scaling wavelet function as follows:

$$x(t) = \sum_k c_k^{j_0} \phi_{j_0,k}(t) + \sum_{j=j_0}^{j_{\max}-1} d_k^j \psi_{j,k}(t) \tag{3}$$

where, j_0 is a given resolution level, $\left\{c_k^{j_0}\right\}$ and $\left\{d_k^j\right\}$ are known as approximation and detailed coefficients, respectively.

Daubechies complex wavelets transform has following advantages, over real valued wavelet transform.

1. Daubechies complex wavelet transform is approximate shift invariant [23,24] in nature.
2. Daubechies complex wavelet transform has perfect reconstruction property [22].
3. Daubechies complex wavelet transform provides true phase information [24].
4. Daubechies complex wavelet transform has no redundancy [22].

2.2 Properties of Daubechies complex wavelet transform

All general properties of real Daubechies wavelet bases are derived from amplitude only [21]. Therefore, these properties also hold for complex Daubechies wavelet. Daubechies complex wavelets have some other important properties that directly influence video object segmentation algorithm. A brief description of these properties is given in Subsects. 2.2.1 and 2.2.2.

2.2.1 Reduced shift sensitivity

Daubechies complex wavelet transform has reduced shift sensitivity. A transform is said to be shift sensitive if shift in input signal causes an unpredictable change in transform coefficients. In discrete wavelet transform (DWT), the shift sensitivity arises due to downsampler in its implementation. Figure 1 shows the reduced shift sensitivity of Daubechies complex wavelet transform. Figure 1a, b show an input signal and shifted form of the input signal by one sample, respectively. Figure 1c, d show magnitude of high-pass wavelet coefficients of the original and the shifted signal using DWT. Figure 1e, f show magnitude of high-pass wavelet coefficients of the original and the shifted signal using Daubechies complex wavelet transform. From the Fig. 1, it is clear that DWT is highly shift sensitive, whereas the Daubechies complex wavelet transform is approximate shift invariant.

This property helps in segmentation of object with shifting orientation in different frames of video.

2.2.2 Edge detection property of Daubechies complex wavelet transform

Let $\phi(t) = l(t) + iv(t)$ be a scaling function and $\psi(t) = k(t) + iu(t)$ be a wavelet function. Let $\hat{v}(w)$ and $\hat{l}(w)$ are Fourier transforms of $v(t)$ and $l(t)$. Consider the ratio

$$\alpha(w) = -\frac{\hat{v}(w)}{\hat{l}(w)} \tag{4}$$

Clonda et al. [22] observed that $\alpha(w)$ is strictly real valued and behaves as w^2 for $|w| < \pi$. This experiment relates the imaginary and real components of scaling function as $v(t)$ accurately approximate another derivatives $l(t)$ up to some constant factor.

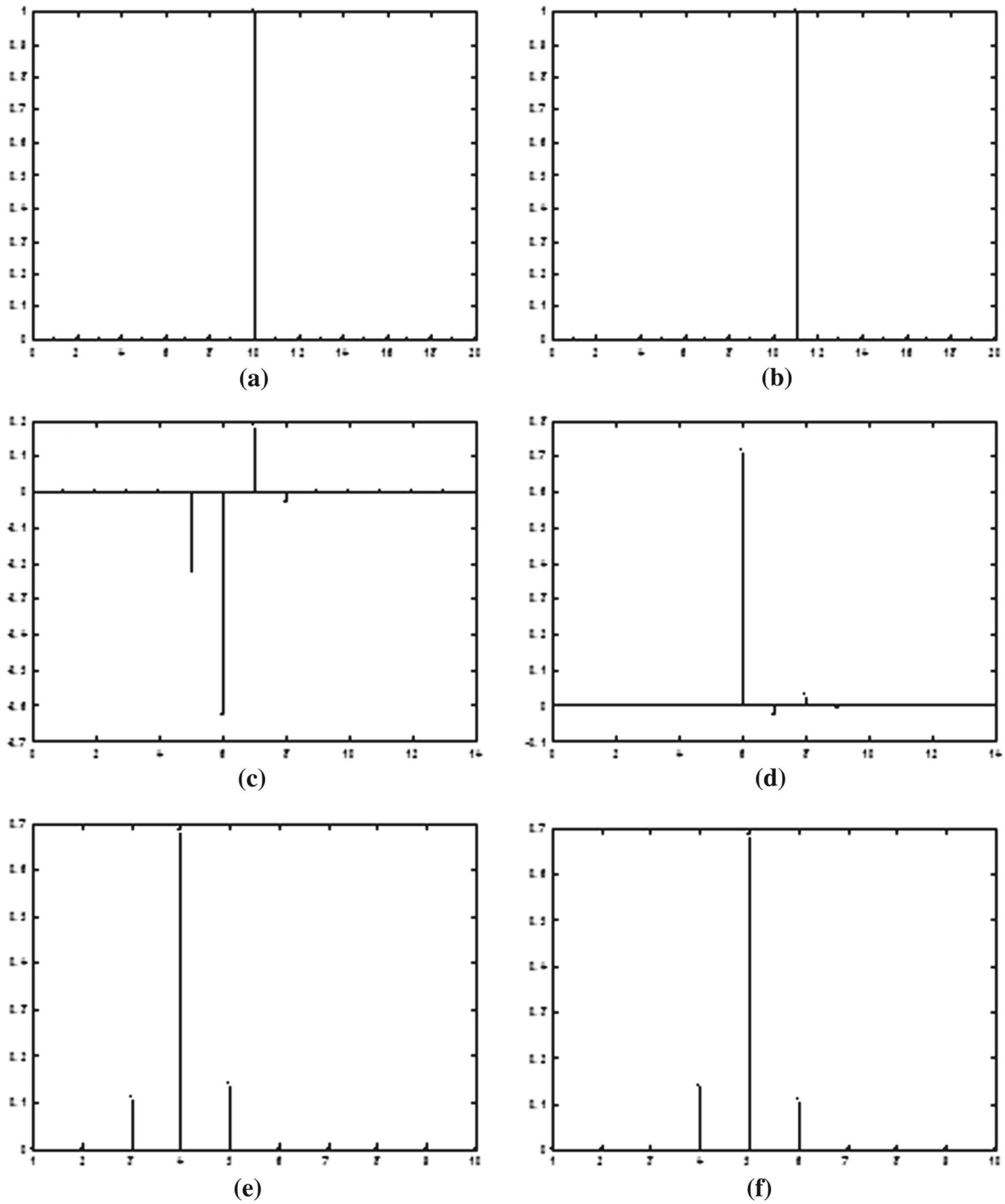


Fig. 1 **a** Original input signal, **b** input signal shifted by one sample, **c** magnitude of high-pass wavelet coefficient of original input signal using real db4 wavelet, **d** magnitude of high-pass wavelet coefficient of shifted input signal using real db4 wavelet, **e** magnitude of complex wavelet coefficient of original input signal using SDW6 wavelet, **f** magnitude of complex wavelet coefficient of shifted input signal using SDW6 wavelet

As a result, it can be derived from Eq. (4) that $v(t) \approx \alpha \Delta l(t)$, where Δ represents second-order derivative. This gives multiscale projection as

$$\begin{aligned}\left\langle f(t), \phi_{j,k}(t)\right\rangle &= \left\langle f(t), l_{j,k}(t)\right\rangle + i\left\langle f(t), v_{j,k}(t)\right\rangle \\ &\approx \left\langle f(t), l_{j,k}(t)\right\rangle + i\alpha\left\langle \Delta f(t), l_{j,k}(t)\right\rangle \qquad (5)\end{aligned}$$

From Eq. (5), it can be concluded that the real component of complex scaling function carries averaging information and the imaginary component carries edge information. Daubechies complex wavelet transform acts as local edge detector because imaginary components of complex scaling coefficient represent strong edges. This helps in preserving the edges and implementation of edge sensitive video object segmentation method.

3 The proposed method

This paper presents a new method for automatic segmentation of moving object. The proposed algorithm uses double change detection scheme in Daubechies complex wavelet transform domain.

Change detection is a method used for obtaining video object plane using frame difference. It is easy to implement for real-time systems [15–19]. In case of change detection method, difference of two consecutive frames, which also gives the changed pixel value from frame $n-1$ to n, can be obtained using following algorithm:

FOR every$(i, j)\epsilon$ the co-ordinate of frame
$\quad \mathrm{FD}_n(i, j) = |P_n(i, j) - P_{n-1}i, j)|$
$\quad$ IF $\mathrm{FD}_n(i, j) < V_{thr}$ THEN $\mathrm{FD}_n(i, j) = 0$
end FOR

Here, $P_n(i, j)$ is the value of (i, j)th pixel of frame n and $P_{n-1}(i, j)$ is the value of (i, j)th pixel of frame $n-1$, V_{thr} is a threshold value and $\mathrm{FD}_n(i, j)$ is the frames difference of two consecutive frames.

There are two types of change detection methods—single change detection (SCD) method and double change detection (DCD) method. Double change detection method is advantageous over SCD method as it uses three consecutive frames for obtaining frame difference.

The proposed method consists of the following six steps:

1. Wavelet decomposition of sequence of frames.
2. Application of double change detection method on wavelet coefficients.
3. Application of soft thresholding to remove noise.
4. Application of canny edge detector to detect strong edges in wavelet domain.
5. Detection of moving edges.
6. Detection of moving object.

Details of all these steps are described below:

Block Diagram of the proposed method is shown in Fig. 2.

3.1 Step 1: Wavelet decomposition of image frames

Wavelet transform converts an image into four subbands: *LH*, *HL*, *HH*, and *LL*. *LH* contains the high-frequency horizontal information, *HL* contains the high-frequency vertical information, *HH* contains the high-frequency diagonal information, and *LL* contains-low frequency diagonal information. In the proposed approach, we have computed wavelet transform of three consecutive frames—current frame, previous frame, and next frame.

3.2 Step 2: Double change detection method

After wavelet decomposition of three consecutive frames—current frame, previous frame, and next frame—we apply the double change detection method in wavelet domain. For extraction of moving objects with double change detection method, three consecutive frames F_{n-1}, F_n, and F_{n+1} are used. Moving objects in frame n can be obtained by detection of common regions of (F_{n-1}, F_n), and (F_n, F_{n+1}), where (F_{n-1}, F_n) represents frame difference between frame $n-1$ and frame n, and (F_n, F_{n+1}) represents frame difference between frame n and frame $n+1$.

Let $\mathrm{WI}_n(i, j)$, $\mathrm{WI}_{n-1}(i, j)$ and $\mathrm{WI}_{n+1}(i, j)$ are the wavelet coefficients of the current frame, previous frame and next frame, respectively. Wavelet domain change detection masks $\{\mathrm{FD}_n(i, j)\}$ and $\{\mathrm{FD}_{n+1}(i, j)\}$ can be calculated as

$$\begin{aligned} \mathrm{FD}_n(i, j) &= \mathrm{WI}_n(i, j) - \mathrm{WI}_{n-1}(i, j) \\ \mathrm{FD}_{n+1}(i, j) &= \mathrm{WI}_{n+1}(i, j) - \mathrm{WI}_n(i, j) \end{aligned} \tag{6}$$

3.3 Step 3: Soft thresholding

Change detection masks obtained after applying double change detection method may have noise. In the presence of noise, Eq. (6) is expressed as:

$$\begin{aligned} \mathrm{FD}_n(i, j) &= \mathrm{FD}'_n(i, j) + \eta_1 \\ \mathrm{FD}_{n+1}(i, j) &= \mathrm{FD}'_{n+1}(i, j) + \eta_2 \end{aligned} \tag{7}$$

where $\mathrm{FD}'_n(i, j)$ and $\mathrm{FD}'_{n+1}(i, j)$ are frame difference without noise, η_1 and η_2 represent corresponding noise components. We have applied wavelet domain soft thresholding technique [23,25] for estimation of frame difference $\mathrm{FD}'_n(i, j)$ and $\mathrm{FD}'_{n+1}(i, j)$ without noise.

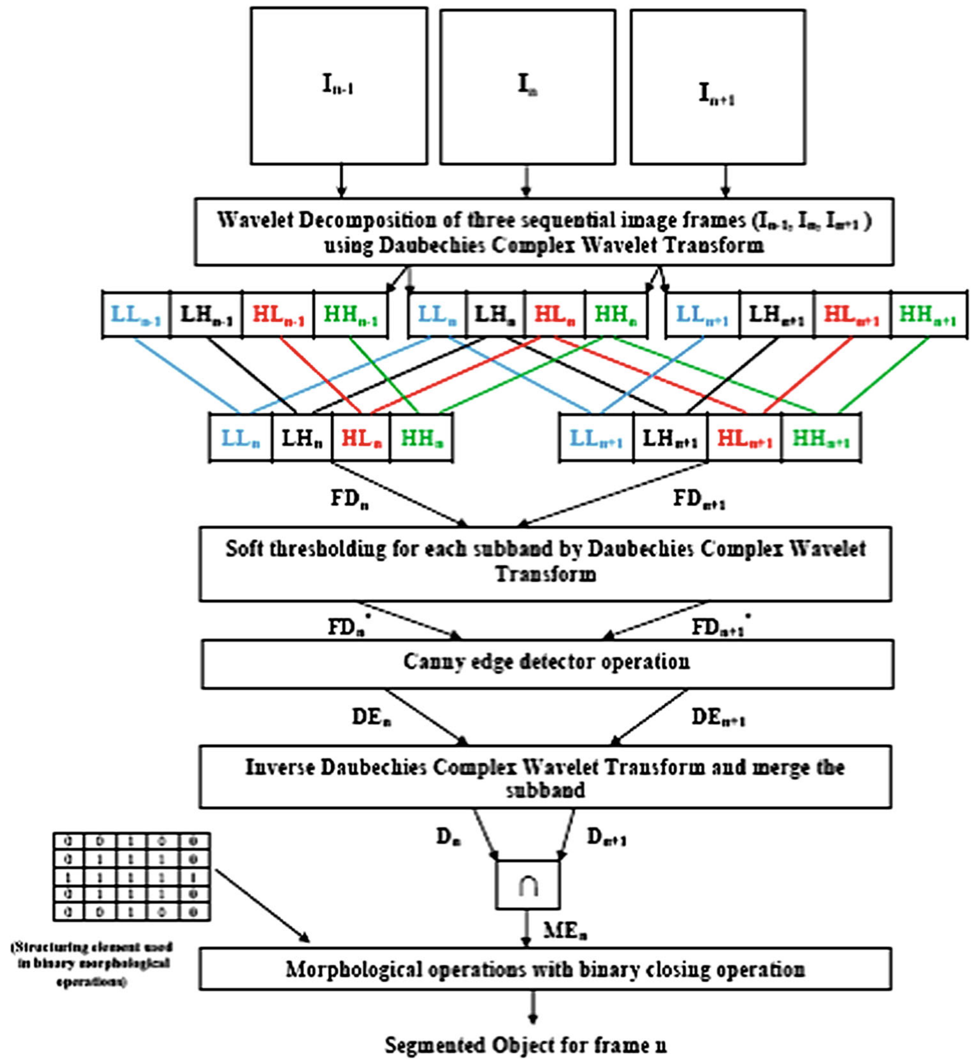
3.4 Step 4: Canny edge detection

Canny edge detection method [26] is one of the most useful and popular edge detection methods, because of its low error rate, well-localized edge points, and single edge point response.

Therefore, we applied Canny edge detection operator on $\{\mathrm{FD}'_n(i, j)\}$ and $\{\mathrm{FD}'_{n+1}(i, j)\}$ to detect the edges of significant difference pixels in all subbands., i.e.,

$$\begin{aligned} \mathrm{DE}_n(i, j) &= \mathrm{Canny}(\mathrm{FD}'_n(i, j)) \\ \mathrm{DE}_{n+1}(i, j) &= \mathrm{Canny}(\mathrm{FD}'_{n+1}(i, j)) \end{aligned} \tag{8}$$

Fig. 2 Block diagram of the proposed method



3.5 Step 5: Moving edge detection

Moving object edge map for frame $n(\mathrm{ME}_n)$ is derived as intersection of D_n and D_{n+1}, where D_n and D_{n+1} are inverse wavelet transform of DE_n and DE_{n+1}, respectively, to get the robust interframe difference edges D_n and D_{n+1}. The points corresponding to the moving object edge map (ME_n) are not necessarily restricted to be on the boundary of the moving object. Instead, these points may be part of interior of the boundary of the moving object.

3.6 Step 6: Moving object detection

Moving objects for frame n can be formed by the detected moving edges. Due to non-ideal segmentation of moving object edges, disconnected edges are found. Extractions of moving object using these disconnected edges are lead to inaccurate moving object segmentation. Therefore, some morphological operation is needed for post-processing of object edge map to generate connected edges. Here, we have used binary closing morphological operation as described in Gonzalez and Woods [27]. The structuring element of binary closing morphological operation is shown in Fig. 2.

4 Experiments and result

The proposed method for segmentation of moving object as described in Sect. 3 has been applied on a number of video clips. Results are being presented here for four representative video sequences viz. One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence. Hall Monitor video sequence is a low-quality video sequence. Campus video sequence is a low contrast video sequence. Shadows of different objects are also present in this video sequence. In Taichi video sequence, background is non-stationary.

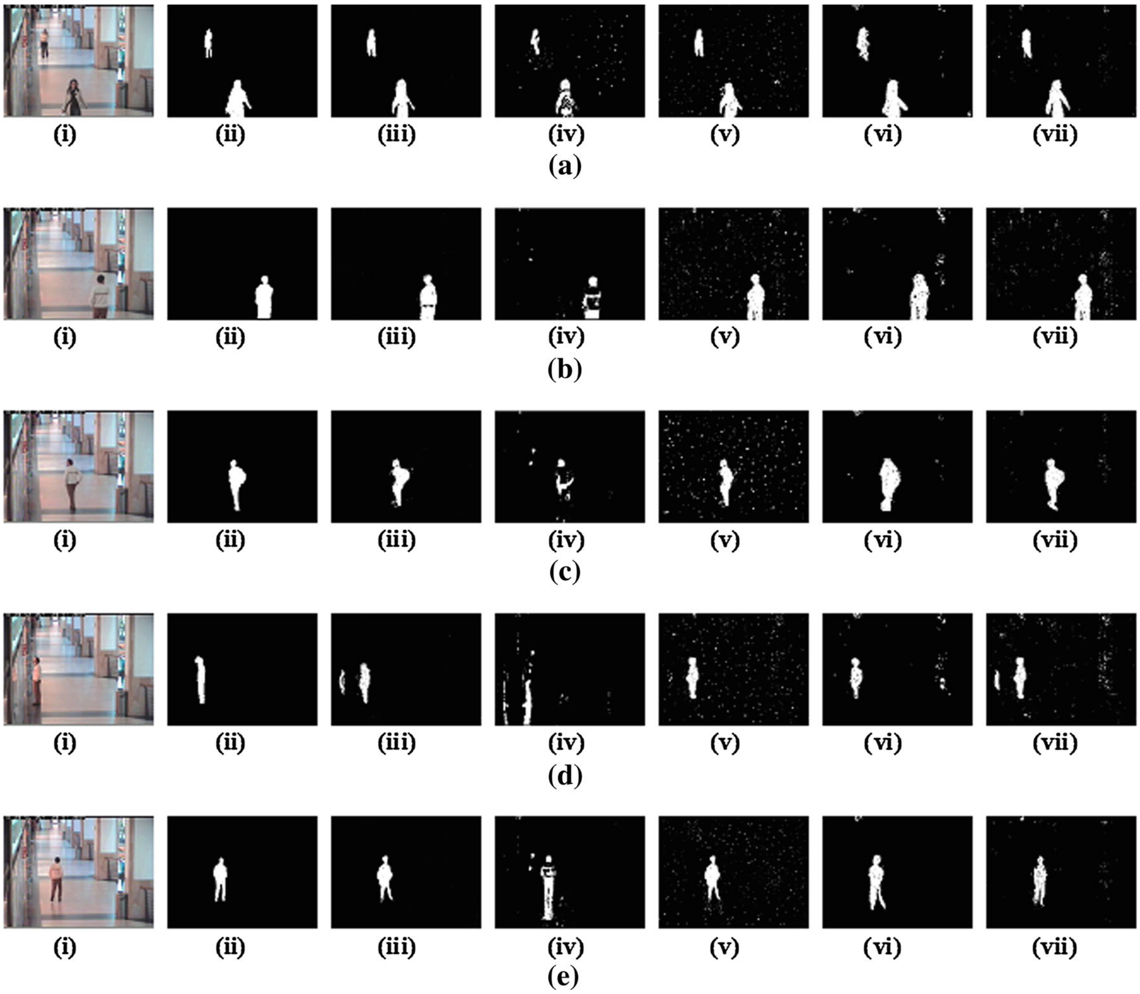


Fig. 3 Segmentation results for One Step video sequence corresponding to **a** Frame 2, **b** frame 100, **c** frame 200, **d** frame 300, **e** frame 400 (*i* Frame no. n, *ii* ground truth segmented frame, *iii* segmented frame obtained by use of the proposed method, *iv* segmented frame obtained by the method used by Kim et al. [6], *v* segmented frame obtained by the method used by Mahmoodi [8], *vi* segmented frame obtained by the method used by Huang et al. [17] and *vii* segmented frame obtained by the method used by Baradarani [18,19])

To start with frames 1, 2, and 3 have been taken into account for segmentation of the moving object as the first frame for which a just previous frame and next frame are available is frame-2. The same is repeated for the frames following frame-2. The results obtained for all representative video clips have been compared with those obtained using methods used by Kim et al. [6] and Mahmoodi [8] falling under category of spatial domain segmentation methods; and methods used by Huang et al. [17] and Baradarani [18,19] falling under category of wavelet domain segmentation methods. Implementation regarding method used by Kim et al. [6] is taken from book entitled "Learning Open CV" authored by Gary Bradski and Adrian Kaehler [28]. Implementation regarding method used by Mahmoodi [8] is available on author's own website. Method used by Huang et al. [17] and Baradarani [18,19] has been implemented by the authors of this work.

4.1 Experiment-1

Figure 3 shows results for One Step video sequence which is downloaded from http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/. One Step video sequence contains 458 frames of frame size 384×288. Results for frame numbers 2, 100, 200, 300, and 400 are given in Fig. 3.

4.2 Experiment-2

Figure 4 shows results for Hall Monitor video sequence which is downloaded from http://trace.eas.asu.edu/yuv/. Hall monitor video sequence contains 287 frames of frame size 352×240. Results for frame numbers 25, 75, 125, 175, 225, and 275 are given in Fig. 4.

From Figs. 3 and 4, it is observed that better shape of moving object with least noise in segmented frame is

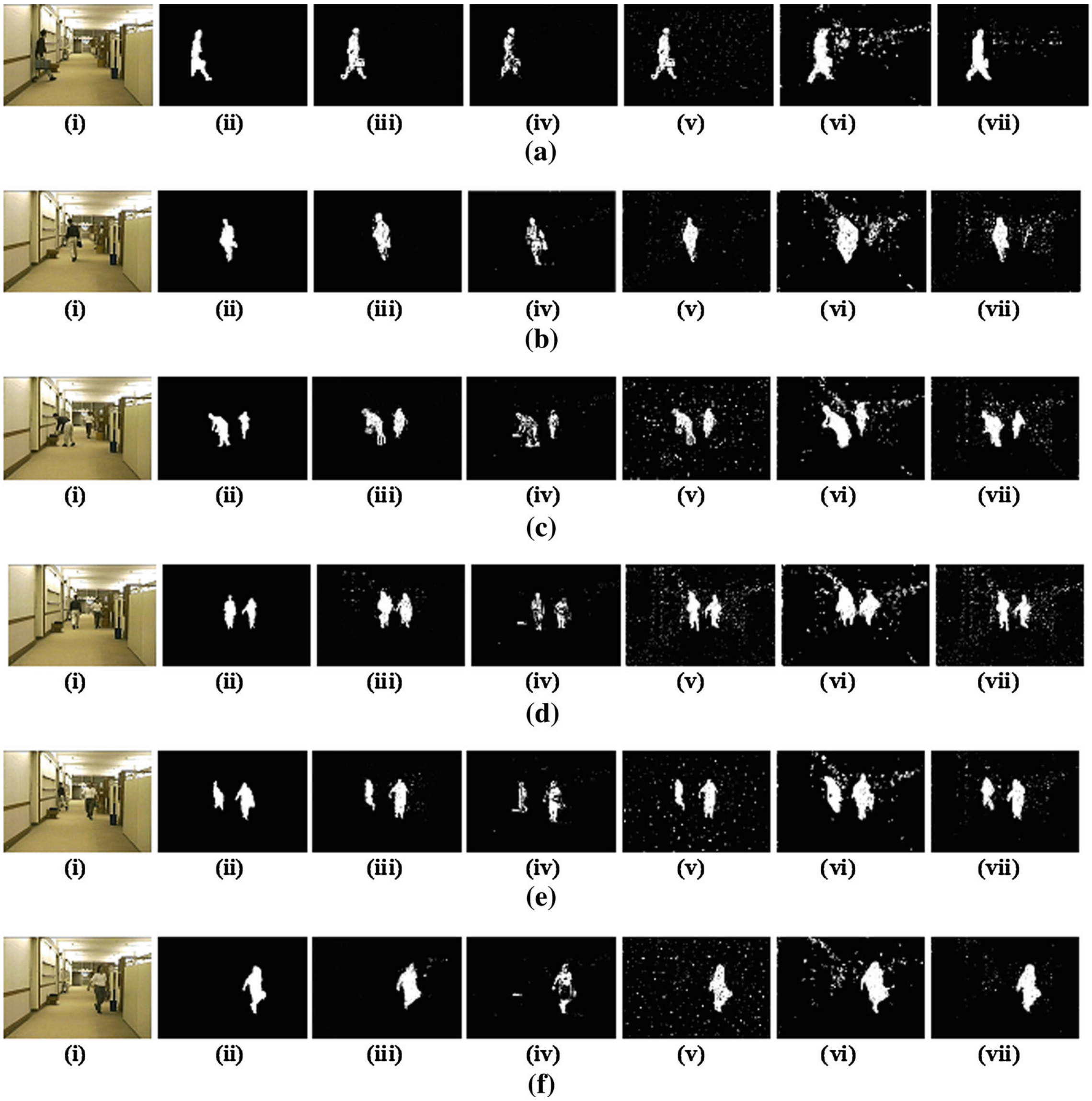


Fig. 4 Segmentation results for Hall Monitor video sequence corresponding to **a** Frame 25, **b** frame 75, **c** frame 125, **d** frame 175, **e** frame 225, **f** frame 275 (*i* Frame no. n, *ii* ground-truth segmented frame, *iii* segmented frame obtained by use of the proposed method, *iv* segmented frame obtained by the method used by Kim et al. [6], *v* segmented frame obtained by the method used by Mahmoodi [8], *vi* segmented frame obtained by the method used by Huang et al. [17] and *vii* segmented frame obtained by the method used by Baradarani [18,19])

obtained by using the proposed method as compared to all other methods [6,8,17–19] taken into consideration for comparison. Method used by Baradarani [18,19] results in comparable shape structure as compared to the proposed method but is poor in noise removal. Method used by Mahmoodi [8] also gives satisfactory shape information but have poor noise removal capability in comparison with the proposed method, and the method used by Baradarani [18,19]. Method used by Kim et al. [6] is poor in shape information than the proposed method, but is better in noise removal as compared to the method used by Baradarani [18,19] and Mahmoodi [8]. Method used by Huang et al. [17] is inferior in shape information of moving object as well as in noise removal as compared to the proposed method and method used by Kim et al. [6], Mahmoodi [8], and Baradarani [18,19].

4.3 Experiment-3

Figure 5 shows results for Campus video sequence which is downloaded from http://itee.uq.edu.au/~uqasanin/. Campus video sequence is of low-quality video with poor contrast. Shadows of different objects are also present in this video.

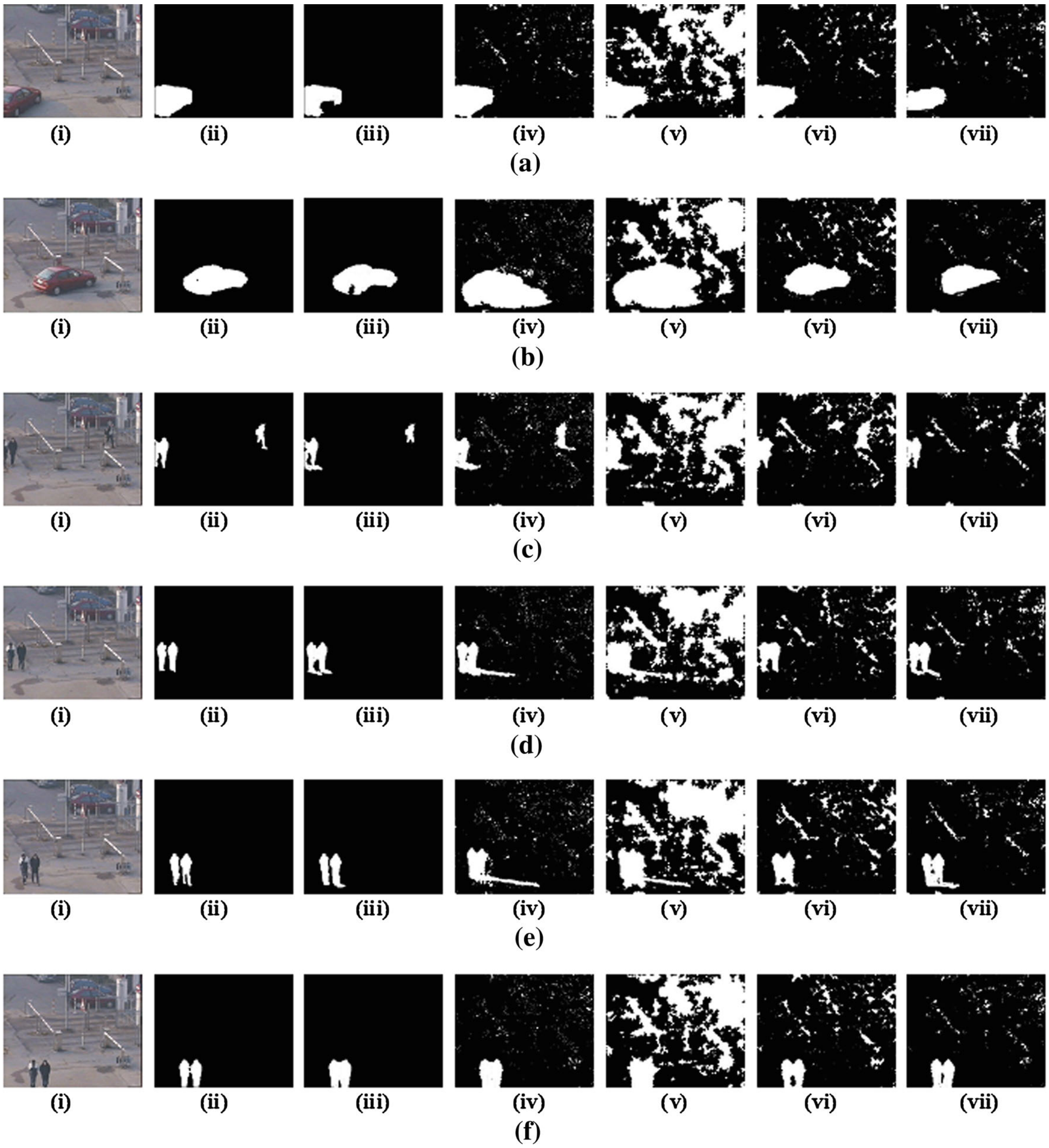


Fig. 5 Segmentation results for Campus video sequence corresponding to **a** Frame 100, **b** frame 150, **c** frame 200, **d** frame 250, **e** frame 300, **f** frame 350 (*i* Frame no. n, *ii* ground-truth segmented frame, *iii* segmented frame obtained by use of the proposed method, *iv* segmented frame obtained by the method used by Kim et al. [6], *v* segmented frame obtained by the method used by Mahmoodi [8], *vi* segmented frame obtained by the method used by Huang et al. [17] and *vii* segmented frame obtained by the method used by Baradarani [18,19])

It contains 410 frames of frame size 352×288. Results for frame numbers 100, 150, 200, 250, 300, and 350 are given in Fig. 5.

From Fig. 5, one can observe that the segmentation result obtained by the proposed method has better shape of object as compared to all other methods [6,8,17–19] for low contrast video as well as for video with shadow of moving objects.

4.4 Experiment-4

Figure 6 shows segmentation results for Taichi video sequence which is downloaded from http://crcv.ucf.edu/data/UCF101.php. This video sequence has non-stationary background. Taichi video sequence contains 335 frames of frame size 320×240. Results for frame numbers 50, 100, 150, 200, 250, and 300 are given in Fig. 6.

Fig. 6 Segmentation results for Taichi video sequence corresponding to **a** Frame 50, **b** frame 100, **c** frame 150, **d** frame 200, **e** frame 250, **f** frame 300 (*i* Frame no. n, *ii* ground-truth segmented frame, *iii* segmented frame obtained by use of the proposed method, *iv* segmented frame obtained by the method used by Kim et al. [6], *v* segmented frame obtained by the method used by Mahmoodi [8], *vi* segmented frame obtained by the method used by Huang et al. [17] and *vii* segmented frame obtained by the method used by Baradarani [18,19])

Double change detection method is suitable for video with stationary background. In case of non-stationary background, results of this method would not satisfactory because difference between three consecutive frames with moving object and non-stationary background would not be necessarily same as difference corresponding to stationary background. From Fig. 6, one can observe that segmentation results for a video with non-stationary background are poor using the proposed method as well as method used by Kim et al. [6], Mahmoodi [8], Huang et al. [17], and Baradarani [18,19]. However, it is worth mentioning here that the proposed method results in better segmentation of moving object with slowly varying background as compared to the other methods [6,8,17–19] as evidenced by Fig. 6f.

The proposed method has been compared with other mentioned methods in terms of computation time required. Computation has been done on a machine with Intel 1.73 GHz Dual core processor with 2 GB RAM using MATLAB 2012a software. In Table 1, we have shown total computation time (in second) for different methods for a video of frame size 384×288 with 500 frames.

Table 1 Total computation time of different methods

Sr. no.	Methods	Computation time (in second/frame)
1.	The proposed method	3.076
2.	Method used by Kim et al. [6]	0.574
3.	Method used by Mahmoodi [8]	1.262
4.	Method used by Huang et al. [17]	2.218
5.	Method used by Baradarani [18,19]	2.802

5 Performance evaluation

It can be observed from the results that none of the segmentation algorithm gives accurate segmentation result as compared to ground-truth frames. It is also very difficult to compare the segmentation results visually because human visual system can identify and understand scenes with different connected objects effortlessly. Therefore, quantitative performance metrics together with visual results are more appropriate. We have compared quantitative performance metrics of the proposed method with those of other state-of-the-art methods viz. the methods used by Kim et al. [6], Mahmoodi [8], Huang et al. [17], and Baradarani [18,19]. For quantitative comparisons, we have considered five different performance metrics: MP, RFAM, PCM, NAE, and PCC.

5.1 Misclassification penalty

The misclassified pixels in the segmentation results which are farther from the actual object boundary (ground-truth image) are penalized more than the misclassified pixels which are closer to the actual object boundary [29]. MP value lies in the range [0,1] and can be computed using following formula [29]:

$$\mathrm{MP} = \frac{\sum_{(m,n)} I(m,n) \cdot \mathrm{Chem}_{\mathrm{GT}}(m,n)}{\sum_{(m,n)} \mathrm{Chem}_{\mathrm{GT}}(m,n)} \quad (9)$$

Table 2 Values of misclassification penalty (MP)

Frame number	The proposed method	Method used by Kim et al. [6]	Method used by Mahmoodi [8]	Method used by Huang et al. [17]	Method used by Baradarani [18,19]
A: One Step video sequence					
2	0.0128	0.0860	0.0502	0.0511	0.0311
100	0.0052	0.4477	0.3274	0.2613	0.2432
200	0.0400	0.9069	0.3607	0.1926	0.0900
300	0.0122	0.0836	0.3146	0.3274	0.1503
400	0.0125	0.1359	0.9435	0.0199	0.0385
B: Hall Monitor video sequence					
25	0.0141	0.0420	0.1408	0.0821	0.1322
75	0.0092	0.0618	0.0137	0.0481	0.0279
125	0.0144	0.0231	0.0090	0.0382	0.0168
175	0.0040	0.0101	0.0080	0.0170	0.0018
225	0.0020	0.0035	0.0018	0.0093	0.0032
275	0.0019	0.0091	0.0025	0.0063	0.0067
C: Campus video sequence					
100	0.0412	0.0506	0.2669	0.1799	0.1669
150	0.0329	0.0560	0.0233	0.2436	0.0792
200	0.0865	0.2943	0.3187	0.2313	0.1444
250	0.0756	0.1012	0.1228	0.4873	0.9209
300	0.0336	0.1664	0.1001	0.4524	0.2707
350	0.0611	0.1522	0.0694	0.6522	0.2710
D: Taichi video sequence					
50	0.0156	0.4810	0.0310	0.0231	0.0212
100	0.2661	0.1421	0.4161	0.2139	0.0731
150	0.8086	1.7471	0.7283	1.0582	0.9582
200	0.4628	0.2381	0.5619	0.5423	0.3631
250	1.2565	1.5955	1.4021	1.2804	1.8047
300	0.1008	0.5822	0.0850	0.1951	0.1678

Table 3 Values of relative foreground area measure (RFAM)

Frame number	The proposed method	Method used by Kim et al. [6]	Method used by Mahmoodi [8]	Method used by Huang et al. [17]	Method used by Baradarani [18,19]
A: One Step video sequence					
2	0.9504	0.5602	0.9741	0.8588	0.9003
100	0.9754	0.5364	0.7262	0.7262	0.7634
200	0.9752	0.3914	0.7928	0.6385	0.7983
300	0.8702	0.8326	0.6000	0.7946	0.5582
400	0.9540	0.7186	0.6870	0.6807	0.8360
B: Hall Monitor video sequence					
25	0.9258	0.5850	0.7684	0.4134	0.6870
75	0.8488	0.8250	0.8411	0.3288	0.5628
125	0.8807	0.7027	0.6275	0.4379	0.6165
175	0.6893	0.6869	0.7515	0.3774	0.5081
225	0.9235	0.8032	0.5081	0.4473	0.6979
275	0.8986	0.4430	0.6927	0.5109	0.7619
C: Campus video sequence					
100	0.8536	0.7237	0.1520	0.5843	0.8392
150	0.9251	0.4868	0.1818	0.5777	0.9798
200	0.9167	0.4010	0.0713	0.2261	0.3604
250	0.7843	0.3991	0.0614	0.2746	0.3335
300	0.8755	0.3830	0.0657	0.2505	0.3926
350	0.8331	0.5126	0.0756	0.3042	0.4937
D: Taichi video sequence					
50	0.4640	0.5421	0.4896	0.7006	0.7251
100	0.3414	0.6291	0.4364	0.5665	0.6555
150	0.1992	0.6773	0.2034	0.3566	0.3551
200	0.3607	0.7338	0.2899	0.2802	0.4693
250	0.4341	0.7287	0.1653	0.5860	0.5793
300	0.6796	0.5405	0.2746	0.5854	0.6055

where Chem_{GT} denotes the Chamfer distance transform of the boundary of ground-truth object. Indicator function I can be computed as

$$I(m,n) = \begin{Bmatrix} 1, & \text{if} \quad I_{\text{GT}}(m,n) \neq I_{\text{Seg}}(m,n) \\ 0, & \text{if} \quad I_{\text{GT}}(m,n) = I_{\text{Seg}}(m,n) \end{Bmatrix} \tag{10}$$

where $I_{\text{GT}}(m,n)$ and $I_{\text{Seg}}(m,n)$ are ground-truth frame and segmented frame, respectively, with dimension $(m \times n)$. Zero value of MP represents perfect segmentation.

Table 2 shows the values of MP for the proposed method and other methods [6,8,17–19] for four video sequences: One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence.

From Table 2, it is obvious that the proposed method has less value of MP as compared to other methods [6,8,17–19].

5.2 Relative foreground area measure

Relative foreground area measure (RFAM) is calculated between ground-truth frame and segmented frame using following formula [30]:

$$\text{RFAM} = 1 - \frac{\left|\text{Area}(I_{\text{GT}}) - \text{Area}(I_{\text{Seg}})\right|}{\text{Area}(I_{\text{Seg}})} \tag{11}$$

where $\text{Area}(I_{\text{GT}})$ and $\text{Area}(I_{\text{Seg}})$ are area of objects in ground-truth frame and segmented frame, respectively. Value of RFAM will be in the range [0,1]. RFAM with value 1 represents perfect segmentation.

Table 3 shows the values of RFAM for the proposed method and other methods [6,8,17–19] for four video sequences: One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence.

Table 4 Values of pixel classification based measure (PCM)

Frame number	The proposed method	Method used by Kim et al. [6]	Method used by Mahmoodi [8]	Method used by Huang et al. [17]	Method used by Baradarani [18,19]
A: One Step video sequence					
2	0.9254	0.9002	0.8436	0.9144	0.8846
100	0.9314	0.8557	0.7606	0.9096	0.8038
200	0.9449	0.8771	0.8251	0.9287	0.9229
300	0.9514	0.8583	0.7977	0.9429	0.8274
400	0.9476	0.8392	0.8169	0.9455	0.9351
B: Hall Monitor video sequence					
25	0.8969	0.9000	0.7044	0.7833	0.8653
75	0.9248	0.8037	0.8168	0.7816	0.7411
125	0.8648	0.7837	0.6178	0.7763	0.7377
175	0.8118	0.8096	0.6660	0.7757	0.5705
225	0.8436	0.7678	0.5705	0.8006	0.8070
275	0.8818	0.8248	0.7281	0.8134	0.8484
C: Campus video sequence					
100	0.7921	0.7741	0.3706	0.7625	0.4638
150	0.6817	0.6271	0.3142	0.6902	0.4618
200	0.8675	0.6791	0.4079	0.7280	0.8148
250	0.8623	0.6136	0.3635	0.7566	0.4233
300	0.8643	0.7402	0.3417	0.7262	0.4611
350	0.8673	0.7604	0.4078	0.7575	0.5187
D: Taichi video sequence					
50	0.8378	0.4727	0.5227	0.8796	0.8751
100	0.2671	0.4981	0.4292	0.8252	0.8393
150	0.0155	0.5655	0.6340	0.7741	0.3232
200	0.6968	0.8129	0.5627	0.5261	0.1133
250	0.7908	0.4803	0.3957	0.7889	0.3964
300	0.7748	0.7578	0.6288	0.3448	0.8053

From Table 3, one can conclude that the proposed method has better relative foreground area as compared to other methods [6,8,17–19].

5.3 Pixel classification based measure

Pixel classification based measure (PCM) reflects the percentage of background pixels misclassified as foreground pixels and conversely foregrounds pixels misclassified as background pixels [30]. Value of PCM will be in the range [0,1]. PCM with value 1 represents perfect segmentation. PCM can be computed as follows: [30]:

$$\mathrm{PCM} = 1 - \frac{\mathrm{Cardi}(B_{\mathrm{GT}} \cap F_{\mathrm{Seg}}) + \mathrm{Cardi}(F_{\mathrm{GT}} \cap B_{\mathrm{Seg}})}{\mathrm{Cardi}(F_{\mathrm{GT}}) + \mathrm{Cardi}(B_{\mathrm{GT}})} \tag{12}$$

where B_{GT} and F_{GT} denote the background and foreground of the ground-truth frame, whereas B_{Seg} and F_{Seg} denote the background and foreground pixels of the achieved segmented frame. '$\cap$' is the logical AND operation. Cardi$(\cdot)$ is the cardinality operator.

Table 4 shows the values of PCM for the proposed method and other methods [6,8,17–19] for four video sequences: One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence.

From Table 4, it is obvious that the proposed method has better values of PCM as compared to other methods [6,8,17–19].

5.4 Normalized absolute error

Normalized absolute error (NAE) is calculated between ground-truth frame and segmented frame using following formula [3,31]:

$$\mathrm{NAE} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left| I_{\mathrm{GT}}(i,j) - I_{\mathrm{Seg}}(i,j) \right|}{\sum_{i=1}^{m} \sum_{j=1}^{n} I_{\mathrm{GT}}(i,j)} \tag{13}$$

Table 5 Values of normalized absolute error

Frame number	The proposed method	Method used by Kim et al. [6]	Method used by Mahmoodi [8]	Method used by Huang et al. [17]	Method used by Baradarani [18,19]
A: One Step video sequence					
2	0.3030	0.5817	0.5216	0.4468	0.3195
100	0.2588	0.7667	0.5757	0.5305	0.5068
200	0.3655	1.0289	0.8442	0.8523	0.4942
300	0.6808	1.4916	0.9869	0.8171	1.1001
400	0.4038	1.0155	0.8071	0.6551	0.5563
B: Hall Monitor video sequence					
25	0.7074	0.7214	0.9265	1.6992	0.8248
75	0.5830	1.1625	0.5583	2.1108	1.0015
125	0.7122	0.7077	1.1664	1.4959	0.8629
175	0.7161	0.7889	0.9287	1.9743	1.2130
225	0.6558	0.8113	1.2130	1.3924	0.8499
275	0.3365	0.6221	0.7050	1.1837	0.4402
C: Campus video sequence					
100	0.1317	0.3556	2.5015	0.6768	0.6467
150	0.0734	0.2894	2.4606	0.6981	0.3180
200	0.2313	0.4122	5.8966	3.3027	1.6920
250	0.2738	0.4365	4.1367	2.5240	1.9040
300	0.1425	0.0465	3.1050	2.8816	1.4752
350	0.2014	0.8021	3.0913	2.1819	0.9615
D: Taichi video sequence					
50	1.1606	0.8638	1.0755	1.9780	2.0054
100	1.9302	0.6139	1.2637	2.2374	2.0244
150	4.0286	0.5923	3.8716	4.7804	3.0799
200	1.7647	0.6449	2.4316	2.9296	1.8182
250	1.2776	0.6090	3.0123	1.4231	1.2434
300	0.5312	1.0246	2.6246	1.2459	1.1354

where I_{GT} and I_{Seg} are ground-truth frame and segmented frame, respectively, with dimension $(m \times n)$. Lower value of NAE means good segmentation while high value of NAE indicates poor segmentation.

Table 5 shows the values of NAE for the proposed method and other methods [6,8,17–19] for four video sequences: One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence.

From Table 5, it can be inferred that the proposed method has lower values of NAE as compared to other method [6,8,17–19].

5.5 Percentage of correct classification

Percentage of corrected classification is defined as [32]:

$$\mathrm{PCC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{14}$$

where TP is True Positives, i.e., the number of pixels, with change, detected correctly as pixels with change, FP is False Positives, i.e., the number of pixels, with change, detected incorrectly as pixels without change. TN is True Negatives, i.e., the number of pixels, without change, detected correctly as pixels without change and FN is False Negative, i.e., the number of pixels, without change, detected incorrectly as pixel without change. Higher value of PCC means good segmentation.

Table 6 shows the values of PCC for the proposed method and other methods [6,8,17–19] for four video sequences: One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence.

From Table 6, one can observe that the proposed method has higher value of PCC as compared to other methods [6,8, 17–19].

Table 6 Values of percentage of correct classification

Frame number	The proposed method	Method used by Kim et al. [6]	Method used by Mahmoodi [8]	Method used by Huang et al. [17]	Method used by Baradarani [18,19]
A: One Step video sequence					
2	0.9456	0.9120	0.9325	0.9146	0.9319
100	0.9824	0.8153	0.8665	0.8907	0.8920
200	0.9851	0.7549	0.9155	0.9615	0.9776
300	0.9071	0.6600	0.7165	0.7233	0.7909
400	0.9790	0.8541	0.8798	0.9767	0.9631
B: Hall Monitor video sequence					
25	0.9715	0.9514	0.8794	0.7682	0.9185
75	0.9272	0.8542	0.9498	0.8755	0.9116
125	0.9686	0.9946	0.9407	0.8996	0.9561
175	0.9497	0.9944	0.9592	0.8802	0.9529
225	0.9688	0.9706	0.9529	0.9112	0.9408
275	0.9733	0.9728	0.9463	0.8673	0.9761
C: Campus video sequence					
100	0.4996	0.4245	0.1333	0.3736	0.3947
150	0.4996	0.2341	0.1548	0.3706	0.4601
200	0.4603	0.2931	0.0671	0.1886	0.2709
250	0.4398	0.2910	0.0584	0.2210	0.2561
300	0.4667	0.2270	0.0621	0.2049	0.2878
350	0.4543	0.3458	0.0710	0.2391	0.3377
D: Taichi video sequence					
50	0.3164	0.3594	0.3250	0.3368	0.3325
100	0.2544	0.3836	0.3064	0.3015	0.3297
150	0.1659	0.3898	0.1703	0.1522	0.2343
200	0.2656	0.3870	0.2231	0.1838	0.2475
250	0.3051	0.3915	0.1426	0.2723	0.3276
300	0.3976	0.3303	0.2162	0.2992	0.3143

6 Conclusions

In the present work, we have proposed a new method for segmentation of moving object using double change detection together with Daubechies complex wavelet transform. Reduced shift sensitivity and better edge detection properties of Daubechies complex wavelet transform make the proposed method more suitable for segmentation of moving object as compared to methods using real valued wavelet transform. Double change detection method has been chosen as it provides automatic detection of appearance of new objects. The proposed method has been evaluated for a number of video sequences, and results for four representative video sequences viz. One Step video sequence, Hall Monitor video sequence, Campus video sequence, and Taichi video sequence have been presented and analyzed. The results after segmentation of moving object using the proposed method have been compared qualitatively and quantitatively with those methods used by Kim et al. [6], Mahmoodi [8], Huang et al. [17], and Baradarani [18,19]. The proposed method exhibits better noise removal and better shape preservation of segmented moving object as compared to other methods [6,8,17–19]. The shape information obtained by the proposed method as well as other methods [6,8,17–19] are not satisfactory for video sequence with non-stationary background. The reason behind non-satisfactory shape information from the proposed method is that the change detection method being used by the proposed method assumes stationary background. However, the shape information from the proposed method is better than those from other methods [6,8,17–19] for video sequence with comparatively slowly varying non-stationary background. We have used five performance metrics viz. MP, RFAM, NAE, PCM, and PCC for quantitative comparison. The proposed method is found better in terms of all these performance metrics as compared to other methods [6,8,17–19].

Contribution of the proposed method can be summarized as follows:

1. A new method for moving object segmentation which is based on Daubechies complex wavelet transform and double change detection method is developed.
2. This method needs only values of Daubechies complex wavelet transform coefficients.
3. This method needs no manual intervention.
4. The method has been tested on several video sequences and is found to have better performance in terms of a number of performance metrics as compared to representative state-of-the-art methods.

Acknowledgments This work was supported in part by the Department of Science and Technology, New Delhi, India, under Grant No. SR/FTP/ETA-023/2009 and the University Grants Commission, New Delhi, India, under Grant No. 36- 246/2008(SR).

References

1. Sonka, M., Hlavac, V., Boyle, R.: Image Processing Analysis and Machine Vision, 3rd edn. Thomson Asia Pvt. Ltd., Singapore (2008)
2. Hu, W., Tan, T.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man Cybern. Part C **34**(3), 334–352 (2006)
3. Khare, M., Srivastava, R.K.: Level set method for segmentation of medical images without reinitialization. J. Med. Imaging Health Inform. **2**(2), 158–167 (2012)
4. Meier, T.: Segmentation for Video Object Plane Extraction and Reduction of Coding Artifacts, PhD Thesis. Department of Electrical and Electronics Engineering, University of Western, Australia (1988)
5. Kim, C., Hwang, J.N.: Fast and automatic video object segmentation and tracking for content based applications. IEEE Trans. Circuits Syst. Video Technol. **12**(2), 122–129 (2002)
6. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real time forground background segmentation using codebook model. Real Time Imaging **11**(3), 172–185 (2005)
7. Xiaoyan, Z., Lingxia, L., Xuchun, Z.: An automatic video segmentation scheme. In: proceeding of IEEE International Symposium on Intelligent Signal Processing and Communication Systems, pp. 272–275. Xieman, China, Nov. 28-Dec. 1 (2007)
8. Mahmoodi, S.: Shape based active contour for fast video segmentation. IEEE Signal Process. Lett. **16**(10), 857–860 (2009)
9. Javed, O., Shafique, K., Shah, M.: A hierarchical approach to robust background subtraction using color and gradient information. In: Proceeding of International Workshop on Motion and Video Computing, pp. 22–27. Orlando, Florida, USA, Dec. 5–6 (2002)
10. Reza, H., Broojeni, S., Charkari, N.M.: A new background subtraction method in video sequences based on temporal motion windows. In: Proceeding of International Conference on IT to Celebrate S. Charmonman's 72 Birthday, pp. 25.1–25.7, March 2009 (2009)
11. Ivanov, Y., Bobick, A., Li, J.: Fast lighting independent background subtraction. In: Proceeding of IEEE Workshop on Visual Surveillance, pp. 49–55. Bombay, India, Jan. 2, 1998 (1998)
12. Colombari, A., Fusiello, A., Murino, V.: Segmentation and tracking of multiple video objects. Pattern Recognit. **40**(4), 1307–1317 (2007)
13. Kato, J., Watanabe, T., Joga, S., Rittscher, J., Blake, A.: An HMM based segmentation method for traffic monitoring movies. IEEE Trans. Pattern Recognit. Mach. Intell. **24**(9), 1291–1296 (2002)
14. Stauffer, C., Eric, W., Grimson, L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Recognit. Mach. Intell. **22**(8), 747–757 (2000)
15. Hunag, J.C., Hsieh, W.S.: Wavelet based moving object segmentation. Electron. Lett. **39**(19), 1380–1382 (2003)
16. Huang, C.L., Liao, B.Y.: A robust scene-change detection method for video segmentation. IEEE Trans. Circuits Syst. Video Technol. **11**(12), 1281–1288 (2001)
17. Huang, J.C., Su, T.S., Wang, L.J., Hsieh, W.S.: Double change detection method for wavelet based moving object segmentation. Electron. Lett. **40**(13), 798–799 (2004)
18. Baradarani, A., Wu, Q.M.J.: Wavelet based moving object segmentation: from scalar wavelets to dual-tree complex filter banks. In: Herout, A. (ed.) Pattern Recognition Recent Advances, ISBN 978-953-7619-90-9, InTech Publication (2010)
19. Baradarani, A.: Moving object segmentation using 9/7-10/8 dual tree complex filter bank. In: Proceeding of IEEE 19th International Conference on Pattern Recognition (ICPR), pp. 1–4. Tampa, Florida, USA, Dec. 8–11, 2008 (2008)
20. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image change detection algorithms: a systematic survey. IEEE Trans. Image Process. **14**(3), 294–307 (2005)
21. Daubechies, I.: Ten Lectures on Wavelets. SIAM (1992)
22. Clonda, D., Lina, J.M., Goulard, B.: Complex Daubechies wavelets: properties and statistical image modeling. Signal Process. **84**(1), 1–23 (2004)
23. Khare, A., Tiwary, U.S., Pedrycz, W., Jeon, M.: Multilevel adaptive thresholding and shrinkage technique for denoising using Daubechies complex wavelet transform. Imaging Sci. J. **58**(6), 340–358 (2010)
24. Khare, A., Khare, M., Jeong, Y.Y., Kim, H., Jeon, M.: Despeckling of medical ultrasound images using Daubechies complex wavelet transform. Signal Process. **90**(2), 428–439 (2010)
25. Khare, A., Tiwary, U.S.: Soft thresholding for denoising of medical images—a multiresolution analysis. Int. J. Wavelets Multiresolut. Inf. Process. **3**(4), 477–496 (2005)
26. Canny, J.F.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **6**(8), 679–698 (1986)
27. Gonzalez, R.C., Woods, R.E.: Digital Image Processing Ch. 9, 3rd edn, pp. 519–566. Pearson Education Asia Publication, India (2008)
28. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with OpenCv Library, 1st edn. Oreilly, Sharoff Publication, India (2008)
29. Erdem, C.E., Sankur, B., Tekalp, A.M.: Performance measures for video object segmentation and tracking. IEEE Trans. Image Process. **13**(7), 937–951 (2004)
30. Gao-bo, Y., Zhao-yang, Z.: Objective performance evaluation of video segmentation algorithms with ground-truth. J. Shanghai Univ. (English Edition) **8**(1):70–74 (2002)
31. Avcibas, I., Sankur, B., Sayood, K.: Statistical evaluation of image quality measures. J. Electron. Imaging **11**(2), 206–223 (2002)
32. Rosin, P., Ioannidis, E.: Evaluation of global image thresholding for change detection. Pattern Recognit. Lett. **24**(14), 2345–2356 (2003)