Multimedia Tools and Applications (2020) 79:5919–5940
https://doi.org/10.1007/s11042-019-08506-z

Moving object detection using statistical background subtraction in wavelet compressed domain

Sandeep Singh Sengar[1] · **Susanta Mukhopadhyay**[2]

Received: 19 March 2018 / Revised: 30 September 2019 / Accepted: 19 November 2019 /
Published online: 12 December 2019

Abstract
Moving object detection is a fundamental task and extensively used research area in modern world computer vision applications. Background subtraction is one of the widely used and the most efficient technique for it, which generates the initial background using different statistical parameters. Due to the enormous size of the video data, the segmentation process requires considerable amount of memory space and time. To reduce the above shortcomings, we propose a statistical background subtraction based motion segmentation method in a compressed transformed domain employing wavelet. We employ the weighted-mean and weighted-variance based background subtraction operations only on the detailed components of the wavelet transformed frame to reduce the computational complexity. Here, weight for each pixel location is computed using pixel-wise median operation between the successive frames. To detect the foreground objects, we employ adaptive threshold, the value of which is selected based on different statistical parameters. Finally, morphological operation, connected component analysis, and flood-fill algorithm are applied to efficiently and accurately detect the foreground objects. Our method is conceived, implemented, and tested on different real video sequences and experimental results show that the performance of our method is reasonably better compared to few other existing approaches.

Keywords Background subtraction · Moving object detection · Wavelet · Statistical parameters · Morphology

1 Introduction

Video surveillance [54], traffic monitoring [52], gesture recognition [45] are some of the possible important areas of research due to the purpose of security and surveillance at places

✉ Sandeep Singh Sengar
sandeep.iitdhanbad@gmail.com

[1] Department of Computer Science & Engineering, SRM University-AP, Amaravati, 522 502 Andhra Pradesh, India

[2] Department of Computer Science & Engineering, Indian Institute of Technology (ISM), Dhanbad, Jharkhand, India

like railway stations, airports, and other public as well as private places. Moving object segmentation [40, 41, 44] and tracking [15, 38, 51] are also crucial steps in these applications due to the detection of suspicious or unusual activities to perform further high-level processing. Shape, motion, or color features are mostly used by the various segmentation and tracking techniques for the above applications. However, it is difficult to segment accurately the foreground objects due to the reasons like illumination variation, changing background, fake or abrupt motion, and occlusion [31]. Optical flow [10], frame differencing [39] and, background subtraction [1] are the main techniques to perform the object detection task. Foreground objects in the optical flow technique are detected with the help of flow vectors of the moving objects. However, this technique is complex and due to the differential operations involved in this, it is susceptible to noise and illumination variations. While a simple technique called frame differencing is based on the concept of change detection across the successive frames, there are foreground aperture and ghosting problems in it. On the other side, one of the widely used significant and easy methods called background subtraction can detect the moving foreground by subtracting the background frame from the current frame. To remove the effects of lighting and inappropriate events, this method updates the background regularly. Due to the real-time performance and stability in dynamic situations, this method is wildly used for motion detection task. There is vast amount of data available for storage and transmission in the aforementioned application areas; therefore data are stored and transmitted in the compressed form. Several techniques are used for this purpose; and wavelet transform is one of them, which divides the frame sequences into detailed and approximate components and performs further operations on these components only [2, 27].

Numerous works have been proposed by the computer vision groups to segment the moving objects from the background. The change detection and Canny edge detection-based moving object segmentation algorithm in the wavelet domain was originally proposed by Huang et al. [22]. However, it leads to foreground aperture and ghosting problems. This method is further enhanced by Khare et al. [25] by using double change detection [39] and Daubechies complex wavelet coefficients [30]. Soft thresholding and Canny edge detection schemes [8] are employed in it to detect the potential edges in sub-bands. Discrete wavelet transform (DWT) [20] and variance method [34]-based object detection and tracking task is performed by the Gangal et al. [16]. This method combines both the background subtraction and frame differencing techniques to detect moving objects. However, the computational complexity of this method is high because it (1) employs two techniques viz., background subtraction and frame differencing for single task, and (2) combines all the sub-bands of 2D-DWT for object detection task. Complex wavelet and approximate median filter-based moving object segmentation algorithm are proposed for video surveillance environments [26]. Hybrid algorithm for moving object detection in the wavelet compressed domain is proposed by Töreyin et al. [46]. Inter-frame differencing and Daubechies complex wavelet transform-based motion segmentation algorithm is presented in [24]. Here, Daubechies complex wavelet transform is selected due to better directional selectivity and shift invariance property.

An unsupervised motion segmentation technique for extracting moving areas from compressed dataset with the help of *Markov random field classification* [28] and *global motion estimation* [13] is proposed by Chen et al. [9]. Hsia et al. [21] proposed a motion detection approach with the help of modified directional lifting-based 9/7 discrete wavelet transform. This technique preserves the fine shape information in low-resolution image and reduces the computational cost. An efficient moving object detection method based on normalized self-adaptive optical flow is presented by Sengar et al. [37]. Otsu's approach and self-adaptive window technique are employed by this method to compute the optimum threshold and to

select the moving object area respectively. However, slow moving small sized objects cannot be detected using this approach. Histogram-based frame differencing approach is combined with *W4* [19] method to remove the foreground aperture and ghosting problems in moving object detection [42]. Different convolutional neural network (CNN) based approaches have been proposed in [3, 4, 11, 33]. In the current scenario, these approaches are highly efficient to find the moving objects. However, the computational complexity of CNN architecture is very high. Bouwmans et al. [6, 48] suggested to use robust principal component analysis in the field of data sciences. The role and importance of features for background modeling and foreground detection have been shown by Bouwmans et al. [5]. Here, the author has shown the importance of color, texture, edge, streo, motion, and local histogram features in different environments. Background subtraction method using multi-scale structured low-rank and sparse factorization has been proposed by Zheng et al. [55]. Optimization technique has been employed in this method to effectively detect the moving objects. However, it has higher computational complexity.

Number of techniques have been proposed for the segmentation of moving objects in the compressed domain. However, still there are some deficiencies due to illumination variations, shaking camera, etc. For that, we propose a statistical background subtraction-based motion segmentation technique in the wavelet compressed domain. In our method, first we apply the dual-tree complex wavelet transform (DTCWT) on the video sequence to divide it into four sub-bands (LL, LH, HL and HH). Subsequently, initial background frame is estimated from all the detail coefficients (LH, HL, and HH) using median operation. In the next step, the statistical parameters: weighted-mean and weighted-variance for each detail coefficients are computed using the respective background frames. Successively, foreground objects for each detail sub-bands are computed with the help of statistical parameters-based background subtraction approach. Next, all the foreground objects are combined and super sampled to detect the moving object. Finally, the post-processing approach in the form of morphology, connected component analysis and flood fill are employed to efficiently and accurately segment the moving objects. Here, generated background frames from the detail sub-bands are updated regularly to adapt the change in environments. Experimental results and performance analysis on different publicly available benchmark video datasets evident that our approach is considerably better than other existing moving object detection techniques.

The main contributions of this work are summarized as below:

1. For obtaining the robust features, DTCWT based wavelet decomposition is employed. Here, DTCWT helps to reduce the frequency aliasing components and shift variant features.
2. Weighted mean and variance-based background model is generated to accurately detect the moving objects. Here, weights are generated in such a way that, they reduce the effect of outlier.
3. To obtain the accurate foreground, squared Mahalanobis distance between the current frame to weighted mean is obtained and compared with mean and standard deviation based threshold.
4. Target object is obtained by combining the output of foreground object of each sub-bands followed by super sample the video frames.
5. Finally, morphological operation, connected component analysis, and flood-fill algorithm are employed to suppress the noise, detect as well as label the connected components, and generate the silhouette of foreground objects respectively.

6. To address the issues of dynamic background (shaking camera and illumination variations), our background model has been updated for every consecutive frames.
7. Qualitative and quantitative analysis with the help of different benchmark datasets prove that our technique outperforms some state-of-the-art methods.

The rest of this paper is organized as follows. After this introductory section the background of the dual-tree complex wavelet transform is given in Section 2. The proposed statistical parameter based motion segmentation method in the compressed domain is elaborated in Section 3. Experimental results and performance analysis with the help of qualitative and quantitative evaluations are provided in Section 4. Finally, Section 5 concludes our work.

2 Dual-tree complex wavelet transform

The discrete wavelet transform (DWT) [20] has been extensively employed in numerous image and video processing applications such as de-noising, compression, and feature extraction, etc. [29]. However, there are some shortcomings of this method because of aliasing [50], shift variance [7], and lack of directionality. Due to the shift variance limitation of DWT, if there is any small shift in input data then wavelet coefficients are greatly distorted and each sub-band energy is changed. To overcome the above drawbacks and to get robust wavelet-based features, DTCWT [23, 36, 43] has been developed, which reduces the frequency aliasing components and has nearly shift-invariant property. As shown in Fig. 1, DTCWT is an enhancement over the ordinary DWT by employing two parallel DWTs with different low-pass $(h_0^i(n), g_0^i(n))$ and high-pass $(h_1^i(n), g_1^i(n))$ filters in each scale on an input signal. Here, first and second DWTs represent the real and imaginary part of the transform respectively. Second DWT's (imaginary part) wavelet function $\psi_g(t)$ is the Hilbert transform of the first DWT's (real part) wavelet function $\psi_h(t)$ i.e. $\psi_g(t) = H[\psi_h(t)]$ and this helps us to achieve the perfect reconstruction. Here, Hilbert transform is represented by $H[.]$. Hilbert transform pair condition is satisfied by the wavelet functions, if the associated low-pass filter of the second tree $g_0(n)$ is the half sample delayed version of the low pass filter of the first tree $h_0(n)$ [35, 53]. This condition is represented as below in time domain.

$$g_0(n) \approx h_0(n - 0.5) \implies \psi_g(t) \approx H(\psi_h(t)) \tag{1}$$

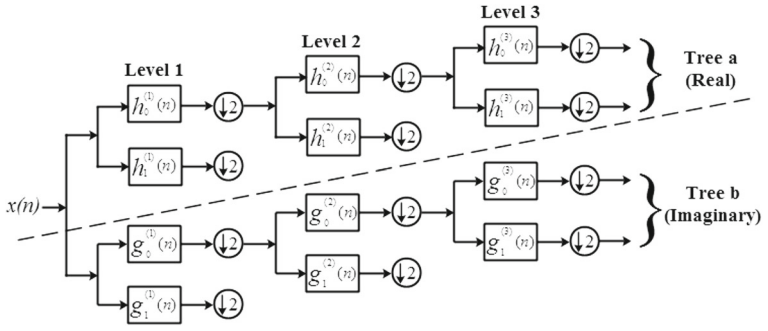

Fig. 1 Representation of dual-tree complex wavelet transform over 3 levels

It can be converted into the frequency domain as below:

$$G_0(w) = e^{-0.5jw} H_0(w) \ for \ |w| < \pi \tag{2}$$

Half sample delay condition cannot be satisfied by the FIR filters, and consequently perfect analytical condition is not fulfilled by the wavelet function pairs. Therefore, to remove the above limitations, approximation is made instead of employing half sample delay system [35]. Approximation can be made by employing different filters in the first stage from the following stages. The orthonormal perfect reconstruction filter pair is used for first stage, which satisfy the following equation:

$$g_0^i(n) = h_0^i(n-1) \tag{3}$$

where $h_0^{(i)}(n)$ and $g_0^{(i)}(n)$ are the low pass filter of real and imaginary tree respectively and i=1, 2, 3 denote the sub-band level for 3-level decomposition. The analytical DTCWT can be obtained approximately at every stage, if the condition of (3) is satisfied.

3 Proposed work

In this work, statistical parameter based background subtraction method is employed in the dual-tree complex wavelet transform domain to automatically segment the moving objects and the steps for our technique are as follows:

1. Suppression of noise by employing Gaussian filter on individual frame.
2. Conversion of frame sequences in wavelet domain using DTCWT prior to obtaining the approximate and detailed coefficients.
3. Application of statistical background subtraction method on each detail wavelet coefficients with the help of statistical parameters-based thresholding technique.
4. Detection of moving objects.
5. Execution of the post-processing steps with the help of morphological operation, connected component analysis, and flood-fill algorithm to generate the silhouette of target objects.
6. Updating the background model to adapt the change in dynamic environments.

The Schematic diagram for the proposed technique is displayed in Fig. 2 and the steps for the proposed work are elaborated below:

3.1 Noise smoothing

The smoothing operation on the individual frame is applied to suppress the noise. Several linear and/or non-linear operators [14, 32, 47] are available for noise suppression from the frame sequences. Among these Gaussian filter is one of the most efficient noise smoothing filters and we have applied it on the individual frame with the help of 1-D Gaussian masks corresponding to spatial dimension in a cascaded manner as per the following equation.

$$F_{smooth} = \frac{1}{\sigma\sqrt{2\pi}} \int \left[\int F(x, y) e^{-(x^2+y^2)/2\sigma^2} dx \right] dy \tag{4}$$

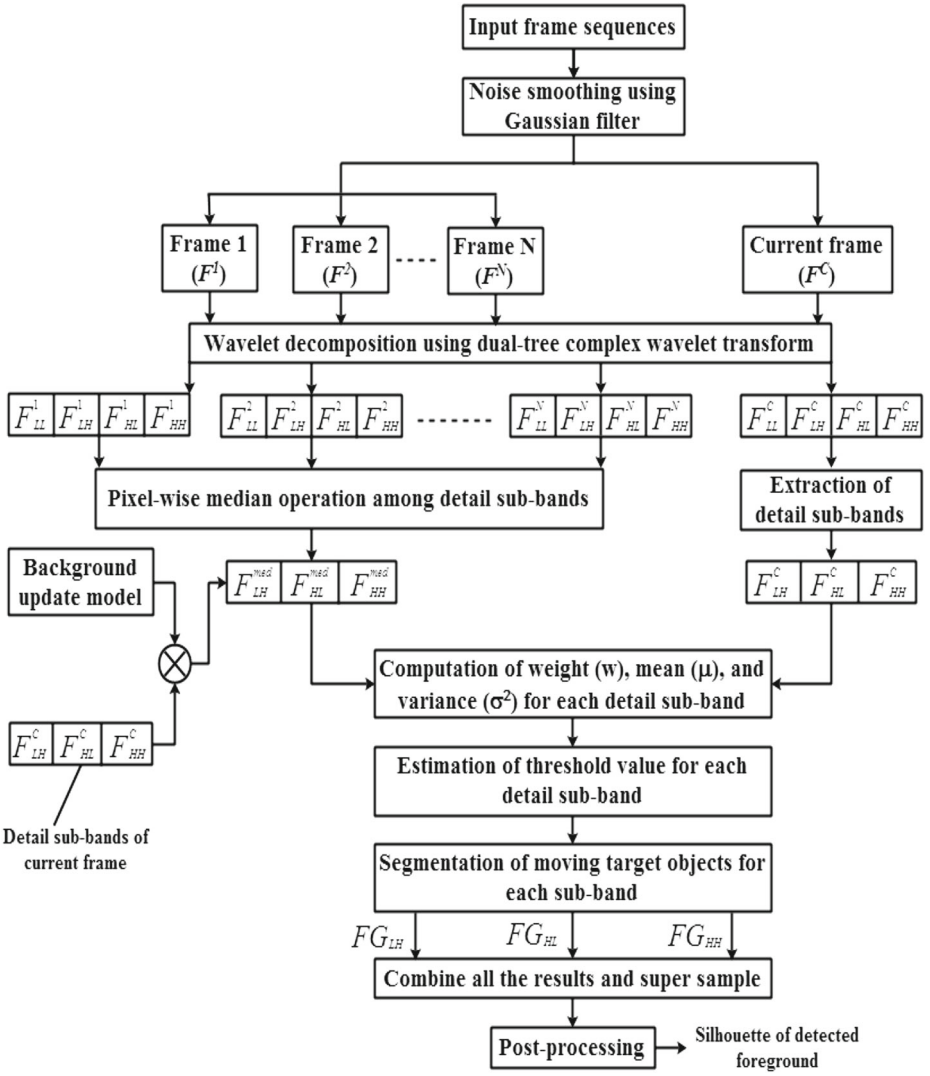

Fig. 2 Schematic diagram of the proposed approach

3.2 Wavelet decomposition

To find the detail of a signal in terms of lower and higher frequency components, wavelet-based band pass filter is employed. Here approximate coefficients (LL part) and detail coefficients (LH, HL, and HH parts) are represented for lower and higher frequencies respectively. Detailed coefficients produce the horizontal, vertical, and diagonal dimensions high frequency details, these details will effectively produce the motion information in further steps. Due to shift-invariant and anti-frequency aliasing properties of DTCWT, In our approach, first we estimate the wavelet coefficients using DTCWT (see Section 2) for N

numbers of frames to compute the initial background frame, here the value of N is 15 (selected experimentally). Next, the wavelet coefficients for current frame are estimated.

3.3 Statistical background model-based moving object detection

To accurately detect the moving objects, background frame should be constructed. In the first step of motion segmentation, statistical background model is generated to accurately represent the background. Pixel-wise median operation among frames is the simplest method to produce the background frame. Therefore, first a pixel-wise median operation is employed among the initial set of detail coefficients (F_{LH}, F_{HL}, and F_{HH}) of N frames to build a reference background for statistical background model construction, these are named as median coefficients F_{LH}^{med}, F_{HL}^{med}, and F_{HH}^{med} (shown in (5)–(7)).

$$F_{LH}^{med}(x, y) = median(F_{LH}^{i}(x, y)) \tag{5}$$

$$F_{HL}^{med}(x, y) = median(F_{HL}^{i}(x, y)) \tag{6}$$

$$F_{HH}^{med}(x, y) = median(F_{HH}^{i}(x, y)) \tag{7}$$

Here i=1, 2,.....,N and N is the total number of frames in the sequence used to build the reference background frame. $F_{LH}^{i}(x, y)$, $F_{HL}^{i}(x, y)$, and $F_{HH}^{i}(x, y)$ are the coefficient values at location (x, y) of detail sub-bands for the i^{th} frame.

The following weighted-mean (shown in (8)–(10)) and weighted-variance (shown in (11–13))) are used to form the statistical background model for each pixel of the detail coefficients of the Z frames:

$$\mu_{LH}(x, y) = \frac{\sum_{i=1}^{Z} W_{LH}^{i}(x, y).F_{LH}^{i}(x, y)}{\sum_{i=1}^{Z} W_{LH}^{i}(x, y)} \tag{8}$$

$$\mu_{HL}(x, y) = \frac{\sum_{i=1}^{Z} W_{HL}^{i}(x, y).F_{HL}^{i}(x, y)}{\sum_{i=1}^{Z} W_{HL}^{i}(x, y)} \tag{9}$$

$$\mu_{HH}(x, y) = \frac{\sum_{i=1}^{Z} W_{HH}^{i}(x, y).F_{HH}^{i}(x, y)}{\sum_{i=1}^{Z} W_{HH}^{i}(x, y)} \tag{10}$$

$$\sigma_{LH}^{2}(x, y) = \frac{\sum_{i=1}^{Z} W_{LH}^{i}(x, y).(F_{LH}^{i}(x, y) - \mu_{LH}(x, y))^{2}}{\frac{Z-1}{Z}\sum_{i=1}^{Z} W_{LH}^{i}(x, y)} \tag{11}$$

$$\sigma_{HL}^{2}(x, y) = \frac{\sum_{i=1}^{Z} W_{HL}^{i}(x, y).(F_{HL}^{i}(x, y) - \mu_{HL}(x, y))^{2}}{\frac{Z-1}{Z}\sum_{i=1}^{Z} W_{HL}^{i}(x, y)} \tag{12}$$

$$\sigma_{HH}^{2}(x, y) = \frac{\sum_{i=1}^{Z} W_{HH}^{i}(x, y).(F_{HH}^{i}(x, y) - \mu_{HH}(x, y))^{2}}{\frac{Z-1}{Z}\sum_{i=1}^{Z} W_{HH}^{i}(x, y)} \tag{13}$$

Here $F_{LH}^{i}(x, y)$, $F_{HL}^{i}(x, y)$, and $F_{HH}^{i}(x, y)$ are the coefficient values of pixels located at (x, y) in the LH, HL, and HH sub-band of i^{th} frame respectively. The weight parameters

$W^i_{LH}(x,y)$, $W^i_{HL}(x,y)$, and $W^i_{HH}(x,y)$ are used to minimize the effects of outliers (by subtracting background part from current frame) and computed as:

$$W^i_{LH}(x,y) = exp\left(\frac{(F^i_{LH}(x,y) - F^{med}_{LH}(x,y))^2}{-2SD^2}\right) \tag{14}$$

$$W^i_{HL}(x,y) = exp\left(\frac{(F^i_{HL}(x,y) - F^{med}_{HL}(x,y))^2}{-2SD^2}\right) \tag{15}$$

$$W^i_{HH}(x,y) = exp\left(\frac{(F^i_{HH}(x,y) - F^{med}_{HH}(x,y))^2}{-2SD^2}\right) \tag{16}$$

Where the value of parameter SD is 5.

The current frame is how much deviating from the mean of the whole video will produce the information of the foreground object. Using this concept, the foreground object for each sub-band is acquired with the help of generated mean and variance-based background models and squared Mahalanobis distance in (17)–(19). Here we consider the pixel value as 1 (foreground), if the Mahalnobis distance of any coefficient is greater than the specified threshold (Th); otherwise this is considered to be 0. This process is shown as follows:

$$FG_{LH}(x,y) = \begin{cases} 1 \text{ if } \frac{(F_{LH}(x,y)-\mu_{LH}(x,y))^2}{\sigma^2_{LH}(x,y)} > Th_1 \\ 0 \text{ otherwise} \end{cases} \tag{17}$$

$$FG_{HL}(x,y) = \begin{cases} 1 \text{ if } \frac{(F_{HL}(x,y)-\mu_{HL}(x,y))^2}{\sigma^2_{HL}(x,y)} > Th_2 \\ 0 \text{ otherwise} \end{cases} \tag{18}$$

$$FG_{HH}(x,y) = \begin{cases} 1 \text{ if } \frac{(F_{HH}(x,y)-\mu_{HH}(x,y))^2}{\sigma^2_{HH}(x,y)} > Th_3 \\ 0 \text{ otherwise} \end{cases} \tag{19}$$

Where the values of Th_1, Th_2, and $Th3$ is computed using following equations:

$$Th_1 = mean\left(\frac{(F_{LH}(x,y) - \mu_{LH}(x,y))^2}{\sigma^2_{LH}(x,y)}\right) + \gamma * std\left(\frac{(F_{LH}(x,y) - \mu_{LH}(x,y))^2}{\sigma^2_{LH}(x,y)}\right) \tag{20}$$

$$Th_2 = mean\left(\frac{(F_{HL}(x,y) - \mu_{HL}(x,y))^2}{\sigma^2_{HL}(x,y)}\right) + \gamma * std\left(\frac{(F_{HL}(x,y) - \mu_{HL}(x,y))^2}{\sigma^2_{HL}(x,y)}\right) \tag{21}$$

$$Th_3 = mean\left(\frac{(F_{HH}(x,y) - \mu_{HH}(x,y))^2}{\sigma^2_{HH}(x,y)}\right) + \gamma * std\left(\frac{(F_{HH}(x,y) - \mu_{HH}(x,y))^2}{\sigma^2_{HH}(x,y)}\right) \tag{22}$$

Here, effective threshold is computed based on of pixel's means and standard deviation. Mean will produce the average intensity and standard deviation will give the deviated intensity value from the mean. However, we need to control on standard deviation value. For that, we used parameter γ. The value of γ should be between 0 to 1. Experimentally, in our case, it is 0.5.

3.4 Moving object detection

Moving objects for the i^{th} frame are detected with the following steps:

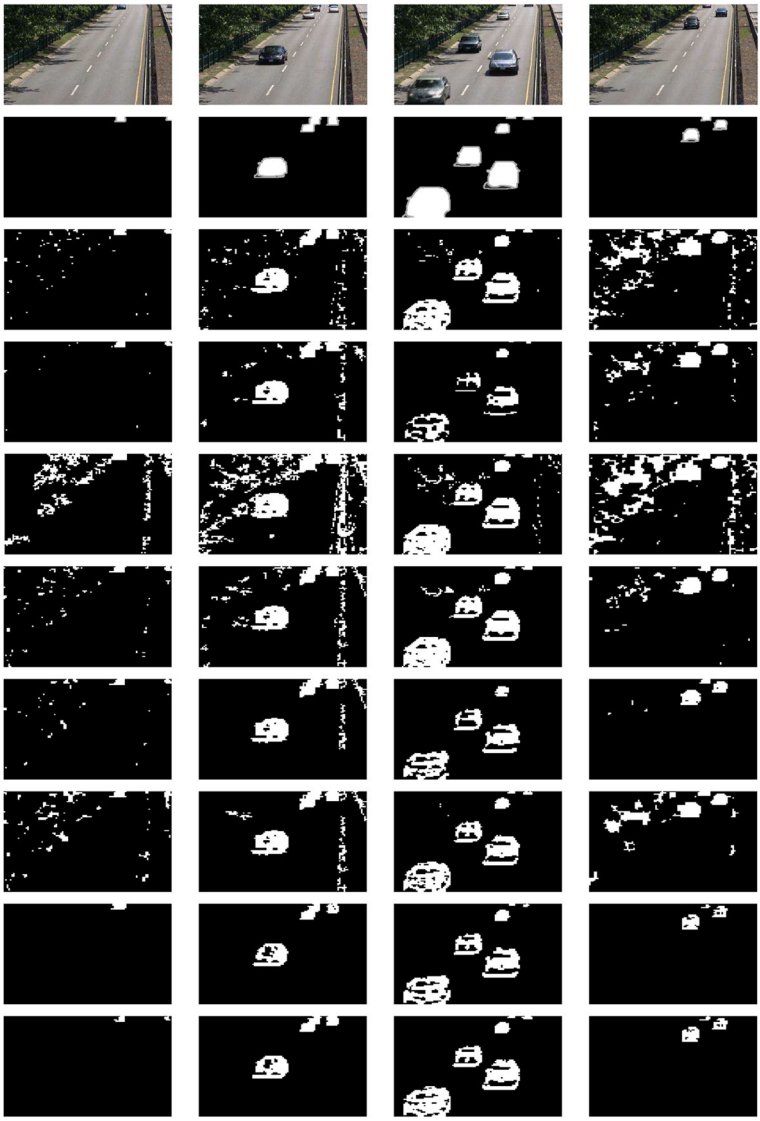

Fig. 3 Results with Highway video for the frames number **a** 505^{th} **b** 666^{th} **c** 890^{th} **d** 1277^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

1. To accurately detect the moving target, combine the output of the foreground object of each sub-bands (computed in Section 3.3).

$$FG(x, y) = FG_{LH}(x, y) + FG_{HL}(x, y) + FG_{HH}(x, y) \tag{23}$$

2. To make our approach simple and to produce the output as the size of the original frame, we super sample the combined output of previous step:

$$FG(x, y) = resize(FG(x, y), 2^l) \tag{24}$$

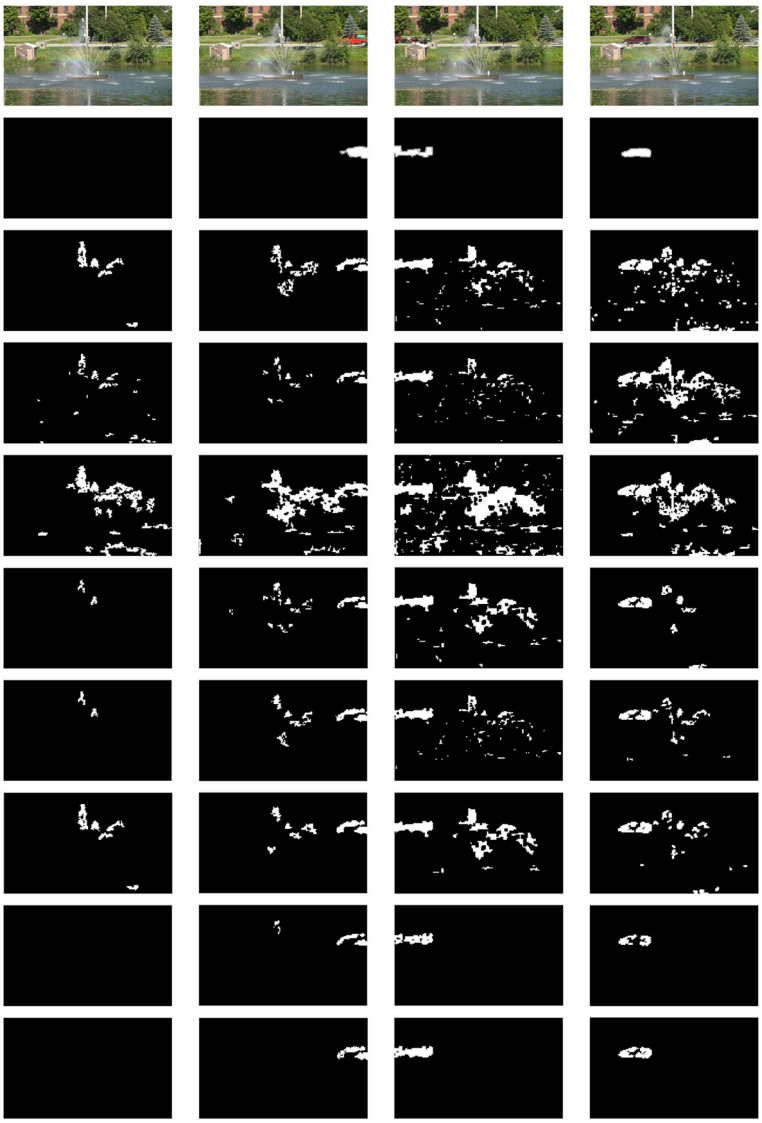

Fig. 4 Results with Fountain2 video for the frames number **a** 138^{th} **b** 671^{th} **c** 760^{th} **d** 1272^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

Where l represents the number of decomposition level by the DTCWT method.

3.5 Post-processing

We carried the post-processing steps employing morphological operation, connected component analysis, and flood-fill algorithm to suppress the noise, detect as well as label the connected components, and generate the silhouette of foreground objects from the binary

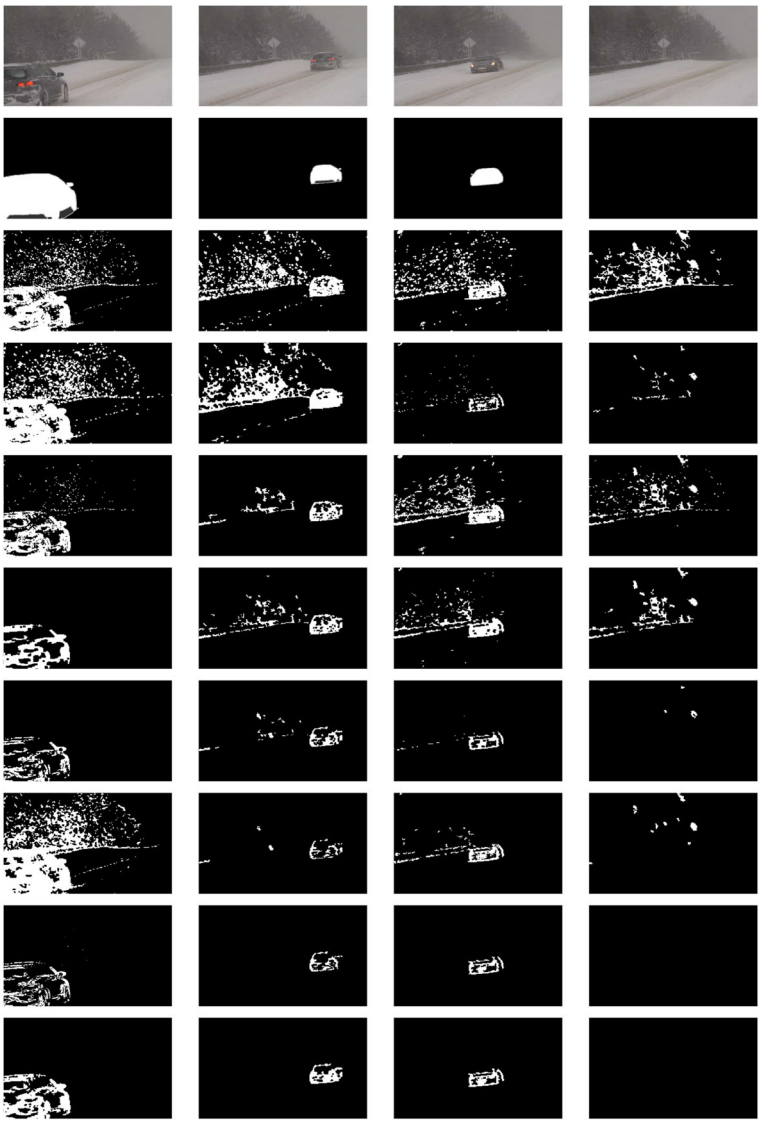

Fig. 5 Results with Snowfall video for the frames number **a** 795^{th} **b** 906^{th} **c** 1390^{th} **d** 3143^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

frame $FG(x, y)$. Morphological closing operator (as shown in (25)) with diamond structuring element (SE) of size 3×3 is applied to eliminate the noise. Here closing is the erosion of dilated version of a binary frame using the same structuring element [12].

$$FG = (FG \oplus SE) \ominus SE \tag{25}$$

Connected component labeling with thresholding operation is employed to remove the isolated small-sized noisy blobs. The binary foreground objects so obtained contain holes

within it. Therefore, we employ the flood-fill algorithm to fill up these holes and finally to obtain the silhouette of the target objects.

3.6 Background update

In illumination variations, jitter, and camera sake environments, the background frame needs to be updated regularly to make sure that the background appropriately corresponds to the current frame. Here motion-based background updating method is done to generate the adaptive background frame. The sub-bands of the current frame and the previous frame are employed to construct the sub-band of the background of the current frame using the following sets of equations.

$$B^{C}_{LH}(x,y) = \begin{cases} \alpha.B^{C-1}_{LH}(x,y) + (1-\alpha)F^{C}_{LH}(x,y) & if(x,y)\ is\ stationary \\ B^{C-1}_{LH}(x,y) & \text{otherwise} \end{cases} \tag{26}$$

$$B^{C}_{HL}(x,y) = \begin{cases} \alpha.B^{C-1}_{HL}(x,y) + (1-\alpha)F^{C}_{HL}(x,y) & if(x,y)\ is\ stationary \\ B^{C-1}_{HL}(x,y) & \text{otherwise} \end{cases} \tag{27}$$

$$B^{C}_{HH}(x,y) = \begin{cases} \alpha.B^{C-1}_{HH}(x,y) + (1-\alpha)F^{C}_{HH}(x,y) & if(x,y)\ is\ stationary \\ B^{C-1}_{HH}(x,y) & \text{otherwise} \end{cases} \tag{28}$$

According to the above equations, the coefficient of the sub-bands of the previous background (B^{C-1}_{LH}, B^{C-1}_{HL}, and B^{C-1}_{HH}) are used to update the current background (B^{C}_{LH}, B^{C}_{HL}, and B^{C}_{HH}) for the foreground coefficient. Alternatively the sub-bands of the previous frame's background and current frame's sub-bands (F^{C}_{LH}, F^{C}_{HL}, and F^{C}_{HH}) are used to update the background. The weighting parameter α is employed to show the importance or influence of previous background in the background updating procedure. The value of parameter α is 0.4 for the proposed method.

4 Experimental results and analysis

To prove the efficacy, our compressed domain-based moving object detection technique has been applied on standard benchmark datasets namely CDnet 2014,[1] Hall,[2] Walk,[3] Meet,[4] and Traffic.[5] The detailed description of these tested video datasets is provided as follows:

- *CDnet 2014* [49]*:* It is a large scale data-set contains 11 categories, where each category consists of 4 to 6 video sequences. Total number of frames are 600 to 7999 in each video sequence with spatial resolutions varying from 320 × 240 to 720 × 576. We evaluate our method on 30 videos from 6 challenging categories including bad weather, baseline, dynamic background, intermittent object motion, shadow, and low frame rate.
- *Hall:* In this indoor surveillance color video dataset, two persons have boxes in their hands and walk from opposite sides. Color of the background and the foreground are much similar and there are illumination variations. This video contains total of 299

[1] www.changedetection.net
[2] http://www.cipr.rpi.edu/resource/sequences/sif.html
[3] https://vid.me/videodata
[4] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
[5] http://clickdamage.com/sourcecode/cv_datasets.php

frames with 240 pixels × 352 pixels size (i.e. row size × column size). The bit rate and frame rate of this dataset is 24 Kbps and 30 frames/sec respectively.

– *Walk:* In this outdoor color dataset, there is a slowly moving person (foreground) having brighter color than the background. This dataset has 132 total number of frames with size 240 pixels × 368 pixels (i.e. row size × column size). The bit rate and frame rate of this dataset are 24 Kbps and 25 frames/sec respectively.
– *Meet:* In this indoor surveillance color video, two persons make entry from opposite sides and do the handshake and go together. This video has varied illumination with dark and bright areas in the background and the size of foreground objects is small. There are total number of 716 frames with size 288 pixels × 384 pixels (i.e. row size × column size). The used bit rate and frame rate of this video are 24 Kbps and 25 frames/sec respectively.
– *Traffic:* In this outdoor traffic surveillance gray scale video data, there are several fast-moving vehicles with varying numbers and size. It also has a person, moving slowly on the footpath. The background of this dataset has small illumination variation, dark or bright regions, trees, and other natural objects. Total number of 49 frames are there in this video with size 512 pixels × 512 pixels (i.e. row size × column size). The bit rate and frame rate of this dataset is 24 Kbps and 30 frames/sec respectively.

We have implemented all the tested techniques and employed the same parameters as recommended by the authors of corresponding work. In the following sections, we will compare our method with other existing approaches based on both qualitative and quantitative analysis.

4.1 Qualitative analysis

The results for our compressed domain moving object detection technique and other existing methods are displayed in Figs. 3, 4, 5, 6, 7, 8, 9 for some of the representative frames. The original video frames, corresponding ground truth, and detected object results are displayed in the aforementioned figures. The results for Highway (baseline category of CDnet 2014 dataset) (the frame number 505, 666, 890, 1277), Fountain2 (dynamic background category of CDnet 2014 dataset) (the frame number 138, 671, 760, 1272), Snowfall (bad weather category of CDnet 2014 dataset) (the frame number 795, 906, 1390, 3143), Hall (the frame number 72, 100, 191, 265), Walk (the frame number 7, 81, 105, 130), Meet (the frame number 311, 418, 610, 702), and Traffic (the frame number 10, 21, 38, 43) are shown for the proposed as well as seven existing techniques. In these Figs. 3–9, first two rows display the original video frame and ground truth respectively. Next seven rows (from top to bottom) display the results of Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] respectively.

The last row of the aforementioned figures show the result of the proposed method. As we can see in resultant figures, our method accurately detects the foreground objects and has considerably high similarity with the ground truth in comparison to other tested techniques for all the used video datasets. Other existing techniques (Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11]) cannot segment the moving objects correctly and misclassified most of the foreground pixels as background and vice-versa. Thus, segmented objects cannot be efficiently discriminated from the background region in these approaches.

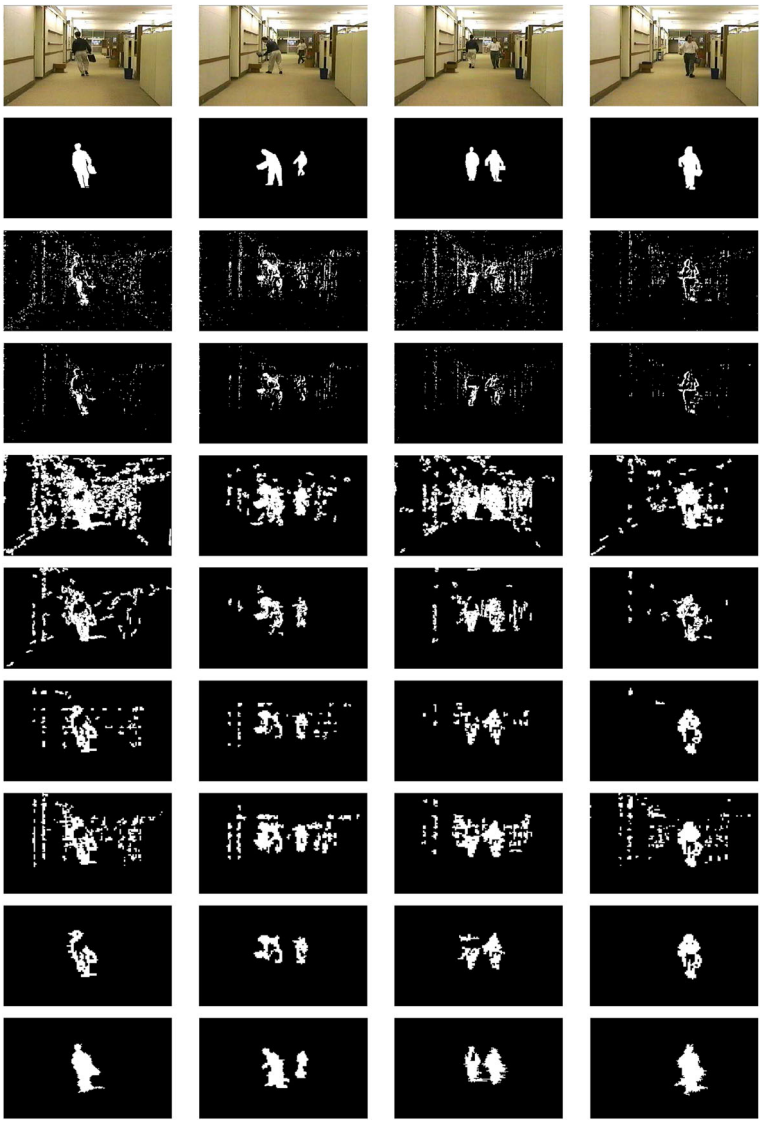

Fig. 6 Results with Hall video for the frames number **a** 72^{th} **b** 100^{th} **c** 191^{th} **d** 265^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

4.2 Quantitative analysis

Qualitative comparison of the proposed technique with other existing methods is shown in the previous section and it has been proved that our method performs reasonably better than some of the existing schemes. It has been observed from the qualitative analysis that perfect segmentation of moving objects is a very challenging task with all the techniques. Thus, it will be very difficult to judge the performance of the proposed and existing methods

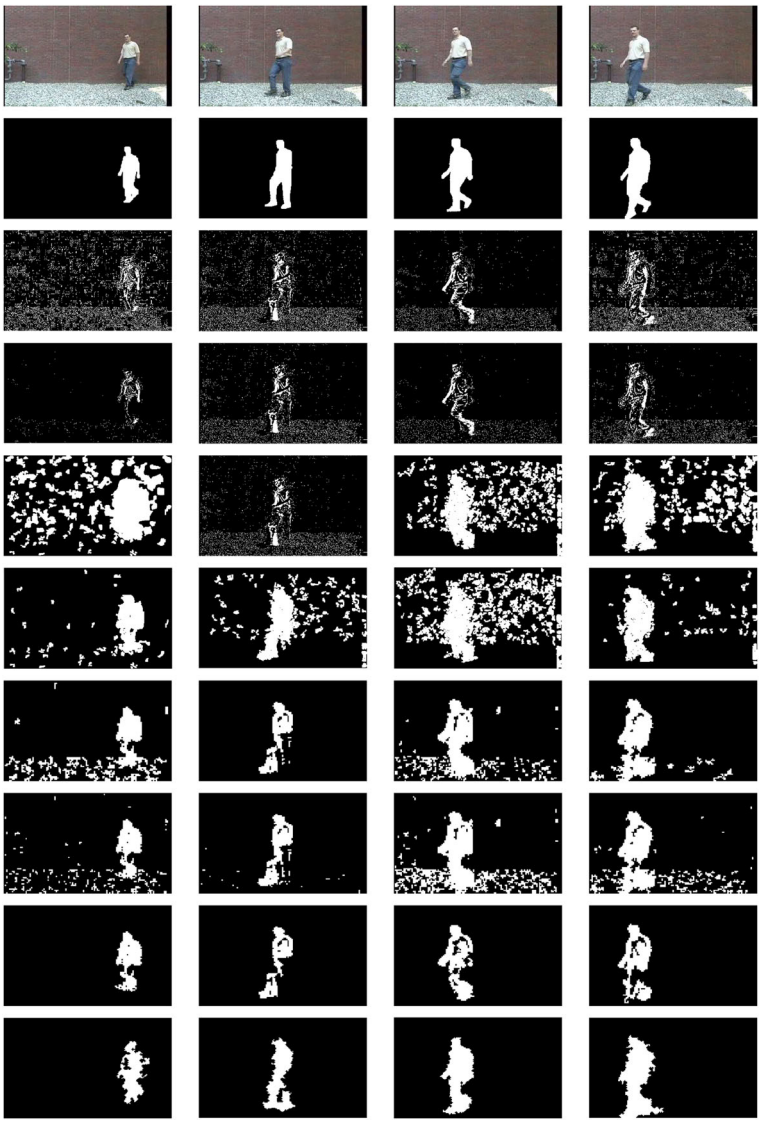

Fig. 7 Results with Walk video for the frames number **a** 7^{th} **b** 81^{th} **c** 105^{th} **d** 130^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

using human visual system only. Furthermore, quantitative analysis together with qualitative evaluation are more suitable for accurate performance measure. In order to measure the performance of the proposed and tested techniques, we use five metrics [49] based on the numbers of false positive FP (background pixels detected as foreground), false negative FN (foreground pixels detected as background), true positive TP (correctly detected foreground pixels), and true negative TN (correctly detected background pixels). These performance

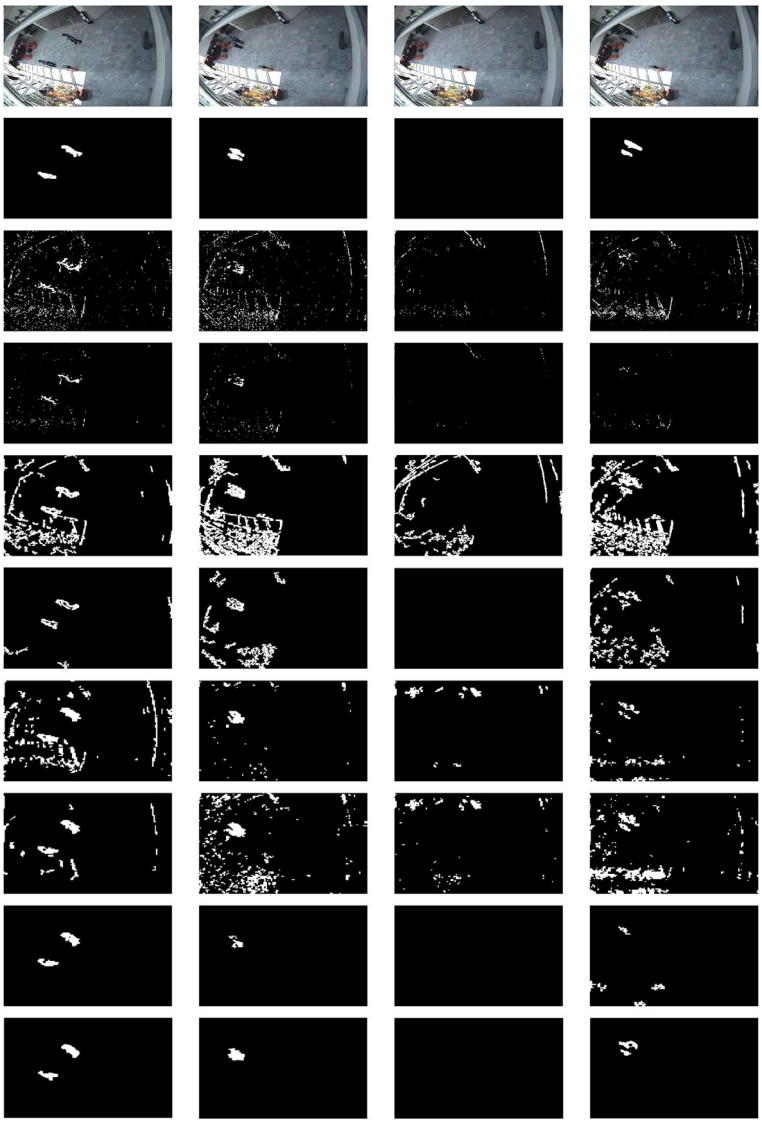

Fig. 8 Results with Meet video for the frames number **a** 311^{th} **b** 418^{th} **c** 610^{th} **d**702^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

metrics are expressed as follows:

$$\begin{array}{ll} Recall = TP/(TP + FN) & Precision = TP/(TP + FP) \\ FPR = FP/(FP + TN) & FNR = FN/(TN + FP) \\ & Specificity = TN/(TN + FP) \end{array} \tag{29}$$

To correctly detect the foreground objects, the values of recall, precision and specificity should be high as well as FPR and FNR should be low. Tables 1–2 display the recall, precision, FPR, FNR, and specificity for all the tested datasets. Table 1 displays the average

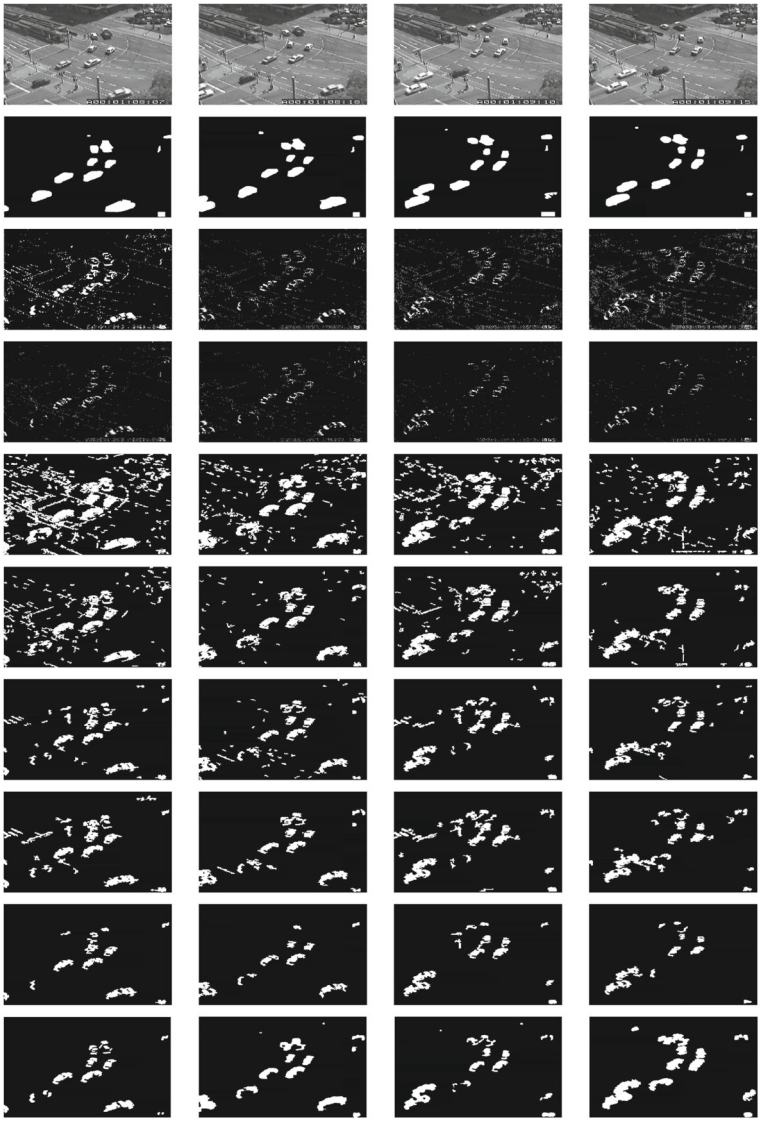

Fig. 9 Results with Traffic video for the frames number **a** 10^{th} **b** 21^{th} **c** 38^{th} **d** 43^{th}; row wise, top to bottom: original frame, ground truth, Huang et al. [22], Gangal et al. [16], Khare et al. [24], Srivastava et al. [25], Yue et al. [18], Tao et al. [17], Dou et al. [11] and the proposed method

performance on bad weather, baseline, dynamic background, intermittent object motion, shadow, and low frame rate categories of CDnet 2014 dataset (total 30 videos); whereas Table 2 shows the average performance on Hall, Walk, Meet and Traffic video sequences. As shown in the aforementioned tables, our technique has significantly large recall, precision, and specificity value and considerably small FNR and FPR value than almost all the tested techniques. Furthermore, Dou et al. [11] performs better than our proposed method

Table 1 Performance of the different methods on CDnet 2014 dataset

Category	Method	Recall	Precision	FPR	FNR	specificity
Bad	Proposed	.8017	.9568	.0009	.1983	.9991
weather	Dou	.8189	.9415	.0006	.1811	.9994
	Tao	.7151	.9015	.0916	.2849	.9084
	Yue	.7918	.8950	.0195	.2082	.9805
	Srivastava	.7816	.9110	.0081	.2184	.9919
	Khare	.7659	.8716	.0413	.2341	.9587
	Gangel	.6103	.8503	.1803	.3897	.8197
	Huang	.5501	.8010	.2001	.4399	.7999
Baseline	Proposed	.9287	.9318	.0486	.0713	.9514
	Dou	.9089	.8910	.0124	.0911	.9876
	Tao	.8805	.8915	.0391	.1195	.9609
	Yue	.9081	.9179	.0384	.0919	.9616
	Srivastava	.9201	.9189	.0182	.0799	.9818
	Khare	.8985	.9001	.0412	.1015	.9588
	Gangel	.7951	.8290	.1483	.2049	.8517
	Huang	.8005	.7951	.2014	.1995	.7986
Dynamic	Proposed	.8895	.9102	.0388	.1105	.9612
Background	Dou	.8992	.9215	.0282	.1008	.9718
	Tao	.8059	.7904	.1743	.1941	.8257
	Yue	.8803	.8521	.1849	.1197	.8151
	Srivastava	.8751	.9010	.1291	.1249	.8709
	Khare	.8504	.8703	.2011	.1496	.7989
	Gangel	.8121	.8014	.2231	.1879	.7769
	Huang	.7250	.7219	.2903	.2750	.7097
Intermittent	Proposed	.6597	.8514	.0986	.3403	.9014
object motion	Dou	.6815	.8804	.0873	.3185	.9127
	Tao	.5958	.6954	.3001	.4042	.6999
	Yue	.7102	.9012	.1891	.2898	.8109
	Srivastava	.6419	.8123	.1085	.3581	.8915
	Khare	.6108	.7158	.2482	.3892	.7518
	Gangel	.6305	.7519	.2741	.3695	.7259
	Huang	.5605	.6599	.3188	.4395	.6812
Shadow	Proposed	.9297	.6413	.0981	.0703	.9019
	Dou	.9682	.6354	.0481	.0318	.9519
	Tao	.8210	.5013	.2321	.1790	.7679
	Yue	.9782	.6954	.0749	.0218	.9251
	Srivastava	.8916	.5181	.1948	.1084	.8052
	Khare	.8915	.5190	.1805	.1085	.8195
	Gangel	.7210	.4919	.3741	.2790	.6259
	Huang	.6958	.4516	.4986	.3042	.5014

Table 1 (continued)

Category	Method	Recall	Precision	FPR	FNR	specificity
Low	Proposed	.6514	.9512	.0684	.3486	.9316
frame rate	Dou	.6019	.9118	.0982	.3981	.9018
	Tao	.4915	.7251	.1301	.5085	.8699
	Yue	.5410	.8213	.0899	.4590	.9101
	Srivastava	.5803	.8907	.1081	.4197	.8919
	Khare	.5029	.8059	.1285	.4971	.8715
	Gangel	.5217	.8305	.0978	.4783	.9052
	Huang	.4718	.6983	.1747	.5282	.8253

Table 2 Performance of the different methods on other datasets

Input video	Method	Recall	Precision	FPR	FNR	specificity
Hall	Proposed	.9712	.9410	.0381	.0288	.9619
	Dou	.9656	.9259	.0211	.0344	.9789
	Tao	.8751	.8980	.1103	.1249	.8897
	Yue	.9059	.9210	.0850	.0941	.9150
	Srivastava	.9210	.9159	.0741	.0790	.9259
	Khare	.8980	.9056	.0933	.1020	.9067
	Gangel	.8212	.8314	.1711	.1788	.8289
	Huang	.8131	.8125	.2085	.1869	.7915
Walk	Proposed	.9512	.9410	.0390	.0488	.9610
	Dou	.9401	.9315	.0482	.0599	.9518
	Tao	.8734	.8654	.1316	.1266	.8684
	Yue	.9010	.9179	.0747	.0990	.9253
	Srivastava	.9215	.9315	.0449	.0785	.9551
	Khare	.8434	.8951	.1049	.1566	.8951
	Gangel	.8130	.8057	.1699	.1870	.8301
	Huang	.8014	.7913	.1888	.1986	.8112
Meet	Proposed	.9199	.9598	.0749	.0801	.9251
	Dou	.9358	.9315	.0311	.0642	.9689
	Tao	.8056	.7687	.1931	.1944	.8069
	Yue	.8625	.8059	.1077	.1375	.8923
	Srivastava	.8915	.8457	.0813	.1085	.9187
	Khare	.8319	.7915	.1733	.1681	.8267
	Gangel	.7910	.7814	.2185	.2090	.7815
	Huang	.7257	.6957	.2947	.2743	.7053
Traffic	Proposed	.8855	.9019	.0649	.1145	.9351
	Dou	.8912	.9212	.0897	.1088	.9103
	Tao	.7457	.7151	.2313	.2543	.7687

Table 2 (continued)

Input video	Method	Recall	Precision	FPR	FNR	specificity
	Yue	.8168	.8014	.1686	.1832	.8314
	Srivastava	.8259	.8514	.0998	.1741	.9002
	Khare	.7259	.7014	.1985	.2741	.8015
	Gangel	.6817	.6524	.3042	.3183	.6958
	Huang	.6459	.6262	.3766	.3541	.6234

for some categories of datasets. However, due to complex deep learning architecture, the processing time of Dou et al. [11] is much higher than our simple technique.

5 Conclusion

In this paper, a new approach has been developed for moving object detection in the wavelet compressed domain. This method detects the motion using only the detail subband information of dual-tree complex wavelet transform without performing the inverse wavelet transform. Shift-invariance and better edge representation properties of dual-tree complex wavelet transform build our technique more appropriately for segmentation of moving target in comparison to the approaches based on other wavelet transform. Adaptive thresholding-based background subtraction technique dependent on weighted-mean and weighted-variance based statistical parameters has been employed to detect the moving target. Connected component analysis, morphological operations, and flood-fill algorithm are used as post-processing steps to accurately segment the motion and generate the silhouette of target objects. The experimental results and analysis using both the qualitative and quantitative analysis on different standard video datasets prove that the segmentation performance of the proposed wavelet domain approach is significantly good even without performing the operations on actual pixel data. This method has several additional advantages in comparison to other compressed domain methods as it is simple, computationally efficient, and produces more accurate results.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Akula A, Khanna N, Ghosh R, Kumar S, Das A, Sardana HK (2014) Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences. Infrared Physics and Technology 63:103–109
2. Antonini M, Barlaud M, Mathieu P, Daubechies I (1992) Image coding using wavelet transform. IEEE Trans Image Process 1(2):205–220
3. Babaee M, Dinh DT, Rigoll G (2018) A deep convolutional neural network for video sequence background subtraction. Pattern Recogn 76:635–649
4. Bouwmans T, Javed S, Sultana M, Jung SK (2019) Deep neural network concepts for background subtraction: a systematic review and comparative evaluation. Neural Networks

5. Bouwmans T, Silva C, Marghes C, Zitouni MS, Bhaskar H, Frelicot C (2018) On the role and the importance of features for background modeling and foreground detection. Computer Science Review 28:26–91
6. Bouwmans T, Vaswani N, Rodriguez P, Vidal R, Lin Z (2018) Introduction to the issue on robust subspace learning and tracking: theory, algorithms, and applications. IEEE Journal of Selected Topics in Signal Processing 12(6):1127–1130
7. Bradley AP (2003) Shift-invariance in the discrete wavelet transform. 7th Digital Image computing: Techniques and Applications. Sydney
8. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell (6):679–698
9. Chen Y-M, Bajic IV, Saeedi P (2011) Moving region segmentation from compressed video using global motion estimation and Markov random fields. IEEE Transactions on Multimedia 13(3):421–431
10. Choi JW, Whangbo TK, Kim CG (2015) A contour tracking method of large motion object using optical flow and active contour model. Multimed Tools Appl 74(1):199–210
11. Dou J, Qin Q, Tu Z (2019) Background subtraction based on deep convolutional neural networks features. Multimed Tools Appl 78(11):14549–14571
12. Dougherty ER, Roberto LA (2003) Hands-on morphological image processing, vol 71. SPIE Optical Engineering Press, Washington
13. Dufaux F, Konrad J (2000) Efficient, robust, and fast global motion estimation for video coding. IEEE Trans Image Process 9(3):497–501
14. Farina A (1997) Linear and non-linear filters for clutter cancellation in radar systems. Journal of Signal Processing 59(1):101–112
15. Fırat H, Uğurhan K, İsa Ş, Anıl A (2018) A novel method for robust object tracking with k-means clustering using histogram back-projection technique. Multimed Tools Appl: 1–14
16. Gangal PP, Satpute VR, Kulat KD, Keskar AG (2014) Object detection and tracking using 2D—-DWT and variance method. In: Students conference on engineering and systems (SCES), pp 1–6
17. Gao T, Z-g Liu (2008) Moving video object segmentation based on redundant wavelet transform. In: 2008 international conference on information and automation. IEEE, pp 156–160.
18. Gao T, Liu Z-g, Yue S-h, Zhang J, Mei J-q, Gao W-c (2010) Robust background subtraction in traffic video sequence. J Cent South Univ Technol 17(1):187–195
19. Haritaoglu I, Harwood D, Davis LS (2000) W4: real-time surveillance of people and their activities. IEEE Trans Pattern Anal Mach Intell 22(8):809–830
20. Hong G-S, Kim B-G, Hwang Y-S, Kwon K-K (2016) Fast multi-feature pedestrian detection algorithm based on histogram of oriented gradient using discrete wavelet transform. Multimed Tools Appl 75(23):15229–15245
21. Hsia C-H, Guo J-M (2014) Efficient modified directional lifting-based discrete wavelet transform for moving object detection. Signal Process 96:138–152
22. Huang J-C, Hsieh W-S (2003) Wavelet-based moving object segmentation. Electron Lett 39(19):1380–1382
23. Iqbal MZ, Ghafoor A, Siddiqui AM (2013) Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means. IEEE Geoscience Remote Sens Lett 10(3):451–455
24. Khare M, Srivastava R, Khare A (2014) Single change detection-based moving object segmentation by using daubechies complex wavelet transform. IET Image Process 8(6):334–344
25. Khare M, Srivastava R, Khare A (2015) Moving object segmentation in daubechies complex wavelet domain. SIViP 9(3):635–650
26. Kushwaha AKS, Srivastava R (2014) Complex wavelet based moving object segmentation using approximate median filter based method for video surveillance. In: International advance computing conference. IEEE, pp 973–978
27. Lama RK, Choi M, Kwon G (2016) Image interpolation for high-resolution display based on the complex dual-tree wavelet transform and hidden Markov model. Multimed Tools Appl 75(23):16487–16498
28. Li SZ (2009) Markov random field modeling in image analysis. Springer Science & Business Media, Berlin
29. Li Y, Zhang L, Li B, Wei X, Yan G, Geng X, Jin Z, Xu Y, Wang H, Liu X (2015) The application study of wavelet packet transformation in the de-noising of dynamic EEG data. Bio-Medical Materials and Engineering 26(s1):S1067–S1075
30. Lina J-M (1997) Image processing with complex daubechies wavelets. J Math Imag Vis 7(3):211–223
31. Ma B, Huang L, Shen J, Shao L, Yang M-h, Porikli F (2016) Visual tracking under motion blur. IEEE Trans Image Process 25(12):5867–5876
32. Robert F, Wilcox LM (1994) Linear and non-linear filtering in stereopsis. J Vis Res 34(18):2431–2438

33. Sakkos D, Liu H, Han J, Shao L (2018) End-to-end video background subtraction with 3d convolutional neural networks. Multimed Tools Appl 77(17):23023–23041
34. Sarwas G, Skoneczny S (2015) Object localization and detection using variance filter. In: Image processing and communications challenges, vol 6. Springer, pp 195–202
35. Selesnick IW (2001) Hilbert transform pairs of wavelet bases. IEEE Signal Process Lett 8(6):170–173
36. Selesnick IW, Baraniuk RG, Kingsbury NC (2005) The dual-tree complex wavelet transform. IEEE Signal Processing Magazine 22(6):123–151
37. Sengar SS, Mukhopadhyay S (2016) Moving object area detection using normalized self adaptive optical flow. Optik-International Journal for Light and Electron Optics 127(16):6258–6267
38. Sengar SS, Mukhopadhyay S (2016) Moving object tracking using Laplacian-DCT based perceptual hash. In: International conference on wireless communications, signal processing and networking (WiSPNET). IEEE, pp 2345–2349
39. Sengar SS, Mukhopadhyay S (2016) A novel method for moving object detection based on block based frame differencing. In: 3rd international conference on recent advances in information technology. IEEE, pp 467–472
40. Sengar SS, Mukhopadhyay S (2017) Foreground detection via background subtraction and improved three-frame differencing. Arab J Sci Eng: 1–13
41. Sengar SS, Mukhopadhyay S (2017) Motion detection using block based bi-directional optical flow method. J Vis Commun Image Represent 49:89–103
42. Sengar SS, Mukhopadhyay S (2017) Moving object detection based on frame difference and w4. SIViP: 1–8
43. Serbes G, Sakar BE, Gulcur HO, Aydin N (2015) An emboli detection system based on dual tree complex wavelet transform and ensemble learning. Appl Soft Comput 37:87–94
44. Shanshan Z, Dominik KA, Christian B, Armin CB (2016) Fast moving pedestrian detection based on motion segmentation and new motion features. Multimed Tools Appl 75(11):6263–6282
45. Tarik A, Sait C, Ali AS, Talha TT (2014) Robust gesture recognition using feature pre-processing and weighted dynamic time warping. Multimed Tools Appl 72(3):3045–3062
46. Töreyin BU, Cetin AE, Aksay A, Akhan MB (2005) Moving object detection in wavelet compressed video. Signal Processing: Image Communication 20(3):255–264
47. Tulsyan A, Huang B, Gopaluni RB, Forbes JF (2014) Performance assessment, diagnosis, and optimal selection of non-linear state filters. J Process Control 24(2):460–478
48. Vaswani N, Chi Y, Bouwmans T (2018) Rethinking pca for modern data sets theory, algorithms, and applications [scanning the issue]. Proc IEEE 106(8):1274–1276
49. Wang Y, Jodoin P-M, Porikli F, Konrad J, Benezeth Y, Ishwar P (2014) CDNet 2014: an expanded change detection benchmark dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 387–394
50. Yang J, Park S-T (2003) An anti-aliasing algorithm for discrete wavelet transform. Mech Syst Signal Process 17(5):945–954
51. Yang L, Cao J, Zhu W, Tang S (2015) Accurate and efficient object tracking based on passive RFID. IEEE Trans Mob Comput 14(11):2188–2200
52. Yong T, Congzhe Z, Renshu G, Peng L, Bin Y (2017) Vehicle detection and recognition for intelligent traffic surveillance system. Multimed Tools Appl 76(4):5817–5832
53. Yu R, Ozkaramanli H (2005) Hilbert transform pairs of orthogonal wavelet bases: Necessary and sufficient conditions. IEEE Trans Signal Process 53(12):4723–4725
54. Yuhao L, Dong Y, An W, Wentao W (2018) Pedestrian tracking in surveillance video based on modified CNN. Multimed Tools Appl: 1–18
55. Zheng A, Zou T, Zhao Y, Jiang B, Tang J, Li C (2019) Background subtraction with multi-scale structured low-rank and sparse factorization. Neurocomputing 328:113–121