CrossMark

Object tracking using combination of daubechies complex wavelet transform and Zernike moment

Manish Khare[1] · **Rajneesh Kumar Srivastava**[1] ·
Ashish Khare[1]

Abstract Moving object tracking is one of the challenging problems in computer vision and it has numerous applications in surveillance system, traffic monitoring etc. The goal of object tracking algorithm is to locate a moving object in consecutive video frames. Tracking of moving object in a video becomes difficult due to random motion of objects. This paper introduces a new algorithm for moving object tracking by exploiting the properties of Daubechies complex wavelet transform and Zernike moment. The proposed method uses combination of Daubechies complex wavelet transform and Zernike moment as a feature of object. The motivation behind using combination of these two as a feature of object, because shift invariance and better edge representation property make Daubechies complex wavelet transform suitable for locating object in consecutive frames whereas rotation invariance properties of Zernike moment is also helpful for correct object identification in consecutive frames. Therefore combination of these two as feature of object gives better results. The proposed method matches Zernike moments of Daubechies complex wavelet coefficients of object in the first frame to next consecutive frames. The experimental results and performance evaluation parameters validate that the proposed method gives better performance against other state-of-the-art methods.

Keywords Object tracking · Daubechies complex wavelet transform · Zernike moment · Shift-invariance · Translation and rotation invariance

---

✉ Ashish Khare
   ashishkhare@hotmail.com; khare@allduniv.ac.in

   Manish Khare
   mkharejk@gmail.com

   Rajneesh Kumar Srivastava
   rkumarsau@gmail.com

---

[1]   Department of Electronics and Communication, University of Allahabad, Allahabad 211002, India

1 Introduction

Tracking of moving object in a video sequence is a crucial problem in image and video processing applications [32]. Main objective of object tracking is to generate the trajectory of an object over time by locating its position in every frame of the video [38]. Tracking of moving object in a video sequence is useful in various applications like video surveillance, visual navigation and monitoring, target detection and interpretation, sport video analysis, traffic monitoring, medical applications etc. [11,36,38].

Major difficulties involved in moving object tracking are due to the following reasons [38]-

–   Partial or full object occlusion.
–   Presence of noise and blur in video.
–   Varying lighting condition.
–   Changing background.
–   Real time processing requirement.
–   Loss of information caused by projection of the 3D world on a 2D image.
–   Shape and size of object may vary from frame to frame.

In the present work, we have proposed a feature based object tracking algorithm based on combination of two features. A lot of work have been proposed to solve video object tracking problem using single feature set, but very less literature is available to solve object tracking problem using combination of two feature sets. Use of single feature may not be much successful for solving object tracking problem, because single feature is not rich enough for representation of wide variety of objects. We can solve object tracking problem by combining multiple types of features that can enhance the robustness and tracking accuracy. Use of combination of two or more features is beneficial because some of the features are more informative than other in some aspect, which increases chances of correct identification of the object. This is the main motivation behind using combination of two feature for object tracking. After successful use of combination of two features in some computer vision problems such as object classification [15], and image retrieval [33], we have used combination of Daubechies complex wavelet transform and Zernike moment as a feature set for object tracking. The Daubechies complex wavelet transform having advantages of shift-invariance, rotation invariance and better edge representation. Zernike moment also have many desirable properties such as translation invariance, rotation invariance, robustness to noise etc. In the proposed algorithm, we have combined these two (Daubechies complex wavelet transform coefficients and Zernike moment) for utilizing their important properties and to get better tracking results.

Three main contribution of the proposed work, is listed below:

(i)   The proposed method uses combination of Daubechies complex wavelet transform and Zernike moment as a feature of object. The motivation behind using combination of these two as a feature of object, because shift invariance and better edge representation property make Daubechies complex wavelet transform suitable for locating object in consecutive frames whereas rotation invariance properties of Zernike moment is also helpful for correct object identification in consecutive frames.

(ii)   Our proposed method rely not only on the matching of combination of energy of Zernike moments of wavelet coefficients among different frames, but also on the prediction of displacement of object calculated from Newton's equation of motion that reduces the false tracks. Motivation behind using Newton's equation of motion for prediction of displacement of object, is that by using Newton's equation of motion will can easily calculate the predicted centroid in next frame.

(iii)   We have employed detection of search window direction mechanism in the proposed method, which was described by Khansari et al. [14]. Motivation behind using search window direction mechanism in the proposed method is that, the proper search window location ensures that the object always lies within the search area, hence, this reduces the loss of object inside the search window, if a tracked object is occluded by another object, then by use of direction of motion may reduce the occlusion problem

For performance evaluation, the proposed method is tested on both case (the proposed method with search window direction mechanism as well as the proposed method without search window direction mechanism). Also the proposed method is compared with following well established object tracking methods-

1-   Tracking method based on sparsity-based collaborative model, which is proposed by Zhong et al. [41].
2-   Tracking method based on online multiple instance learning, which is proposed by Babenko et al. [1].
3-   Tracking method based on structured Multi-task sparse learning, which is proposed by Zhang et al. [40].
4-   Tracking method based on joint color-texture histogram, which is proposed by Ning et al. [22].
5-   Tracking method based on covariance method, which is proposed by Porikli et al. [26].
6-   Tracking method based on corrected background weighted histogram, which is proposed by Ning et al. [23].
7-   Tracking method based on combination of Local sparse appearance model and K-selection, which is proposed by Liu et al. [21].
8-   Tracking method based on Daubechies complex wavelet transform coefficients, which is proposed by Khare and Tiwary [18].

Also we have done some modifications by adding search window direction mechanism, in methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], and then visually compared the proposed method with all methods.

Qualitative performance is not enough to judge the quality of any method. Therefore, we have performed quantitative performance comparison of the proposed method with other state-of-the-art methods. For quantitative measures we have used Euclidian distance, Bhattacharya distance and Mahalanobis distance.

Rest of the paper is organized as follows: Literature review is given in Section 2. Section 3 describes basics of used feature set for object tracking (Daubechies complex wavelet transform and Zernike moment). Section 4 describes properties of Daubechies complex wavelet transform and Zernike moment. Section 5 describes the proposed method in detail. Experimental results and Performance evaluation results are given in Section 6 and 7 respectively. Finally conclusions of the present work are given in Section 8.

2 Literature review

A number of tracking algorithms have been proposed to solve the above mentioned problems. A good survey on tracking algorithm is provided by Yilmaz et al. [38]. In general, the object tracking algorithms can be broadly classified into four groups: region based tracking, model based tracking, contour based tracking and feature based tracking. In region based tracking [6,20], the tracker keeps on the track of frame to frame correspondence resulting in a sequence of tracks, by using region of interest in the reference frame. Region based tracking require several parameters such as object size, color, shape, velocity etc. Due to involvement of several parameters, region based tracking require high computational cost. Contour based tracking is basically developed for tracking of non-rigid objects. This type of methods use a Bayesian approach using probability density function of texture and color features. Both of these features are fused into one as independent polling strategy. Contour based tracking [39] methods are not suitable for real time applications such as in surveillance due to its slow execution time.

In feature based methods, some heuristic is needed to select suitable features. Feature based object tracking methods use color histogram processing in spatial domain [6]. Color histogram technique is simple and robust to noise, and suitable for stationary, low noisy and partially occluded object videos.

However, these methods are complex in implementation for handling occlusions. In subsequent years, improvements in tracking algorithms have been proposed by the use of histogram and spatial information. In tracking algorithm proposed by Nummiaro et al. [24] a bootstrap particle filter was used to sample the observation model. Method proposed by Zivkovic et al. [42] uses an efficient local search scheme to find the likelihood of object region and approximated this region by using Bayesian filtering. Shen et al. [30] built a robust template model instead of single image for tracking. The main drawback of the methods discussed above [24,30,42] is to handle the scale changes. Rocha et al. [28] proposed an object tracking method based on image moment. It uses different invariant properties of moment for tracking of object. Disadvantage of this method is that, it cannot handle occlusion problem and not perform well in a video with changing background. Zhong et al. [41] proposed an object tracking method via sparsity-based collaborative model, which exploits both holistic templates and local representations. Babenko et al. [1] proposed an object tracking algorithm with online multiple instance learning. In this method Multiple Instance Learning (MIL) is used instead of traditional supervised learning, which leads to more robust tracking. Zhang et al. [40] proposed visual tracking algorithm using structured Multi-task sparse learning.

A new trend is to combine two or more features for tracking. Ning et al. [22], combined both the color and texture features using LBP to obtain better tracking results. Method proposed by Porikli et al. [26] used covariance tracking which is also very useful and robust against several illumination changes. Method proposed by Ning et al. [23] combined corrected background weighted histogram with mean shift for object tracking. Wang et al. [37] proposed object tracking method by combining Hu moment and ABC shift, and claimed that ABC shift algorithm overcome drawbacks of using color features. Liu et al. [21] proposed visual tracking algorithm using combination of Local sparse appearance model and K-selection. In this method the target appearance is modeled using a sparse coding histogram based on a learned dictionary with novel selection based dictionary learning method called K-Selection.

Another class of feature based tracking algorithms deals with the processing of frequency domain values of pixels, known as transform domain processing. By transforming images from one domain to other domain, some information which is difficult to obtain in one domain can be easily and efficiently obtained in other domain. Fourier transform [9] and discrete

cosine transform [13] based object tracking algorithms were in explored at initial stage. After that wavelet transform based object tracking algorithms came into existence, wavelets are suitable for representing local features of object. Several methods exist for object tracking using wavelet transform [3,14,16,18,27]. Cheng and Chen [3] proposed an object tracking method based on discrete wavelet transform, but discrete wavelet transform is not suitable for video processing applications, because it is shift variant in nature. Dual tree complex wavelet transform is approximate shift invariant and gives better directional selectivity and these properties has been exploited by Khare et al. [16] and Prakash and Khare [27] for object tracking. Khansari et al. [14] proposed an algorithm for tracking of user defined shapes by using Undecimated wavelet packet transform. Daubechies complex wavelet transform is also used for efficient tracking [18] due to its shift invariant and rotation invariance properties.

All the object tracking methods discussed so far, are suffering from various problems such as, some techniques do not make use of multiple image resolution levels, and therefore they are not able to handle motion of variable size objects. Some methods fails to track object in case of changing background, partial occlusion, or crowded object video. Other challenging problems in above discussed methods are robustness against noise and stability of selected features in presence of various object transformation and occlusions.

3 Features used for object tracking

For any computer vision task the notion of a feature is very important. Selection of right features plays a critical role in object tracking. Feature selection property is closely related to the object representation. In general, the most important property of a visual feature is its uniqueness so that the objects can be easily distinguished in feature space. In the present work, for object tracking, we have taken combination of two different feature sets – Daubechies complex wavelet transform and Zernike moment. A brief description of these features are given in Section 3.1 and 3.2 respectively.

3.1 Daubechies complex wavelet transform

In any computer vision application, object may present in translated and rotated form among different frames therefore we require a feature which remains invariant by translation and rotation of object. Most of the real valued wavelet transform coefficients vary by translation and rotation of the object. Use of Daubechies complex wavelet transform, due to its approximate shift-invariance and better edge representation property can avoid these shortcomings of real valued wavelet transform. In the present work we have used Daubechies complex wavelet transform coefficients as a feature set. Computation of Daubechies complex wavelet transform is described as below.

The basic equation of multiresolution theory is the scaling function

$$\phi(u) = 2\sum_i a_i \phi(2u-i) \qquad (3.1)$$

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's wavelet bases $\{\psi_{j,k}(t)\}$ in one-dimension is defined using scaling function $\phi(u)$ as defined in equation (3.1) and multiresolution analysis of $L_2(\mathfrak{R})$[7,31]. During the formulation of general solution if we relax the Daubechies condition for a_i to be real [5], it leads to complex valued scaling function.

The generating wavelet $\psi(t)$ is defined as

$$\psi(t) = 2\sum_n (-1)^n \overline{a_{1-n}} \phi(2t-n) \tag{3.2}$$

where $\phi(t)$ and $\psi(t)$ share same compact support $[-L, L+1]$.

Any function $f(t)$ can be decomposed into complex scaling function and mother wavelet as:

$$f(t) = \sum_k c_k^{j_0} \phi_{j_0,k}(t) + \sum_{j=j_0}^{j_{max}-1} d_k^j \psi_{j,k}(t) \tag{3.3}$$

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

3.2 Zernike moment

Zernike moment was firstly introduced by Teague [34] to overcome the different shortcomings of information redundancy in Geometric moments [34]. Zernike moment is a type of moment function that is used for mapping of an image onto a set of complex numbers. Zernike moment can represent the properties of an image with no redundancy or overlap of information between the moments [12]. Due to these characteristics, Zernike moments have been utilized as feature set in different computer vision applications.

Zernike moment is a set of complex polynomial which form a complete orthogonal set over the interior of the unit circle of $x^2+y^2 \leq 1$ [25]. These polynomials are of the form,

$$V_{mn}(x,y) = V_{mn}(r,\theta) = R_{mn}(r).\exp(jn\theta) \tag{3.4}$$

where m is positive integer and n is positive and negative integer subject to constraints m-$|n|$ even and $|n| \leq m$, r is the length of vector from the origin to pixel (x,y) and θ is the angle between vector r and x-axis in counter clock wise direction, $R_{mn}(r)$ is the Zernike radial polynomial in (r,θ) polar coordinates and defined as

$$R_{mn}(r) = \sum_{s=0}^{\left(\frac{m-|n|}{2}\right)} \frac{(-1)^s (m-s)! r^{m-2s}}{s! \left(\frac{m+|n|}{2}-s\right)! \left(\frac{m-|n|}{2}-s\right)!} \tag{3.5}$$

here $R_{m,-n}(r) = R_{mn}(r)$

The above mentioned polynomial in equation (3.5) is orthogonal and satisfies the Othogonality principle

Zernike moments are the projection of image function $I(x,y)$ onto these orthogonal basis functions. The othogonality condition simplifies the representation of the original image because generated moments are independent [10].

The Zernike moment of order m with repetition n for a continuous image function $I(x,y)$ that vanishes outside the unit circle is

$$Z_{mn} = \frac{m+1}{\pi} \iint_{x^2+y^2 \leq 1} I(x,y)[V_{mn}(r,\theta)] dx dy \tag{3.6}$$

In case of digital image, the integrals are replaced by summation [19], as given below

$$Z_{mn} = \frac{m+1}{\pi} \sum_x \sum_y I(x,y) V_{mn}(r,\theta), \quad x^2 + y^2 \leq 1 \tag{3.7}$$

4 Properties of daubechies complex wavelet transform and Zernike moment

In this section, we describe different properties of Daubechies complex wavelet transform and Zernike moment, which are useful in object tracking.

4.1 Properties of daubechies complex wavelet transform

Daubechies complex wavelet transform have several properties, in which reduced shift sensitivity, and better edge representation properties directly influence object tracking algorithm. Brief description of these properties is given in Subsections 4.1.1 and 4.1.2.

4.1.1 Edge detection property

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

$$\alpha(w) = -\frac{\hat{v}(w)}{\hat{l}(w)} \tag{4.1}$$

Clonda et al. [5] experimentally 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.

From the above property, equation 4.1 indicates $v(t) \approx \alpha \Delta^2 l(t)$. Here Δ^2 represents second order derivative. This gives multi-scale projection as

$$\langle f(t), x_{j,k}(t) \rangle = \langle f(t), l_{j,k}(t) \rangle + i \langle f(t), v_{j,k}(t) \rangle$$

$$\approx \langle f(t), l_{j,k}(t) \rangle + i\alpha \langle \Delta^2 f(t), l_{j,k}(t) \rangle \tag{4.2}$$

From equation 4.2, it can be concluded that the real component of complex scaling function carries averaging information and the imaginary component carries strong edge information. Daubechies complex wavelet transform acts as a local edge detector because imaginary component of complex scaling coefficients represent strong edges. This helps in preserving the edges and implementation of edge sensitive object tracking method.

4.1.2 Reduced shift sensitivity and approximate rotation invariance property

A transform is said to shift-sensitive if shift in input-signal causes an unpredictable change in transform coefficients. Real valued wavelet transform are shift-sensitive whereas Daubechies complex wavelet transform is approximate shift-invariant. Further, shift-variance results in loss of information at multilevel whereas with Daubechies

complex wavelet transform, the information is not significantly less at multilevel due to its shift-invariance property [17]. Daubechies complex wavelet transform is approximate rotation invariant as well. A transform is said to be rotation invariant, if rotation in input signal causes same rotation in transform coefficients. Figure 1 shows the approximate rotation variant and invariant nature of different wavelet transform for and image. Figure 1a [(i)-(iv)] shows an original image, image rotated by 30 degree, 60 degree and 90 degree angles in clockwise direction. . Reconstructed image by high-pass coefficients of discrete wavelet transform and Daubechies complex wavelet transform at 2 levels are shown in Fig. 1b [(v)-(vii)] and Fig. 1c [(ix-(xii)] respectively. In Fig. 1, there are two types of edge structures: rectangular edge and circular edge. As both types of edge structure rotate through space with some angle, the reconstruction using real valued discrete wavelet transform changes erratically, while Daubechies complex wavelet transform reconstructs all local shifts and orientations in approximately same manner. This indicates that Daubechies complex wavelet transform is approximate rotation-invariant.
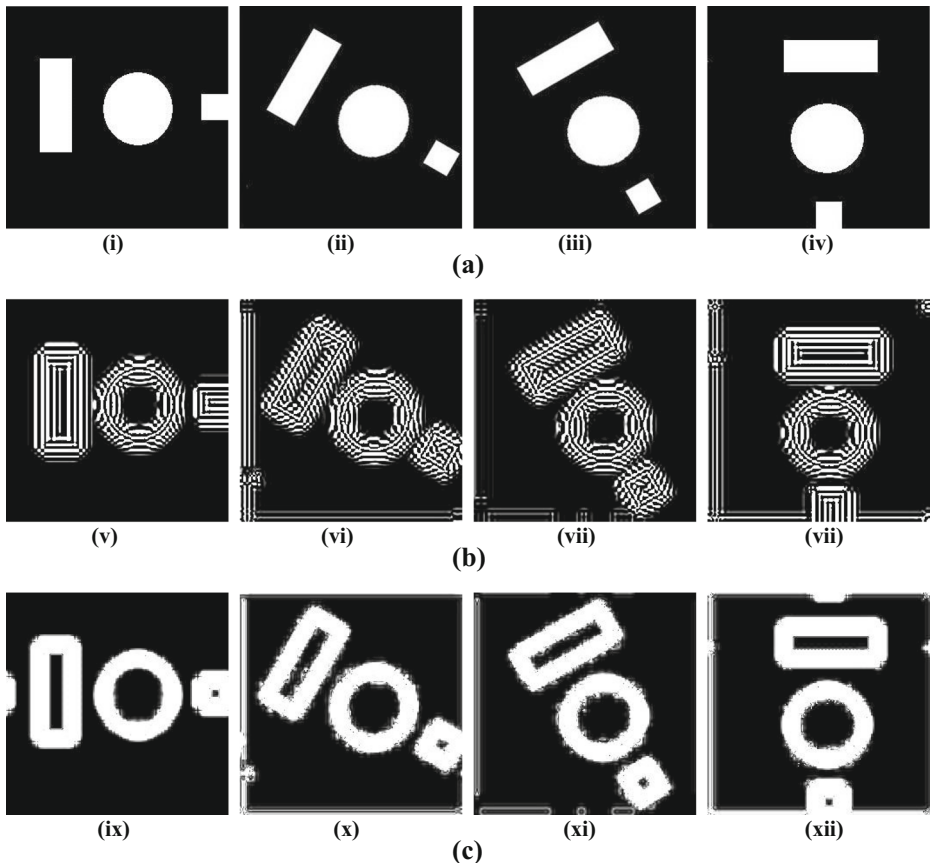

Fig. 1 a [(i)- original image, (ii)-rotation performed by 30 degree angle, (iii)-rotation performed by 60 degree angle, (iv)-rotation performed by 90 degree angle], Image reconstructed from 2 levels of wavelet coefficients using b discrete wavelet transform [(v)-(viii)], and c Daubechies complex wavelet transform [ix-xii]

4.2 Properties of Zernike moment

Zernike moment hold following important properties [4]:

(i)   Zernike moments are rotation, translation and scale invariant [2].
(ii)   Zernike moments are robust to noise and minor variations in shape.
(iii) Since the basis of Zernike moment is orthogonal, therefore they have minimum information redundancy.
(iv)   Zernike moment can characterize the global shape of pattern. Lower order moments represent the global shape pattern and higher order moment represents the detail.
(v)   An image can be better described by a small set of its Zernike moments than any other types of moments.

A big advantage of Zernike moment is that they hold simple rotation invariance property. Rotation invariance is achieved by computing the magnitudes of Zernike moments [35].

The motivation behind using combination of these two as a feature of object is that combining multiple types of features can enhance the tracking accuracy. When we use combination of two or more features, some of the features are more informative than others for an object in a particular frame, therefore chances of correct tracking will be high. In the proposed method, for object tracking we have taken combination of these two features as first we compute Daubechies complex wavelet transform coefficients and then compute Zernike moment of these wavelet coefficients.

5 The proposed method

A video contains a sequence of consecutive frames. Each frame can be considered as an image. If the algorithm can track moving object between two consecutive frames then it will be able to track object in video sequence. The algorithm is semi-automatic in the sense that the user specifies the area along the boundary of the object in the reference frame, then calculate feature value (Daubechies complex wavelet transform coefficients followed by Zernike moment of wavelet coefficients) in the rectangular area. The tracking algorithm uses calculated feature value to find the new location of pixel in the adaptive search window. To speed up the performance of the tracker, the object is searched in its neighborhood in all possible directions rather than in entire frame.

The proposed method consists of following three step:

(i)   Algorithm for segmentation.
(ii)   Algorithm for tracking.
(iii) Algorithm for search window direction for next frame.

Details of all three steps are described below.

5.1 Algorithm for segmentation

The main objective of segmentation is to retrieve object of interest in first frame of video sequence. The segmentation is done in Daubechies complex wavelet domain. For

Fig. 2 Segmentation result of frame 100 of Caviar video sequence by method of Khare et al. [17]

segmentation of object in first frame of video, we have used our earlier developed method [17]. The segmentation approach, as Khare et al. [17], consist of following steps-

(i)   Wavelet decomposition of sequence of frames.
(ii)   Application of double change detection method on wavelet coefficient.
(iii)   Application of soft thresholding of remove noise.
(iv)   Application of canny edge detector to detect strong edges in wavelet domain.
(v)   Detection of strong edges after inverse wavelet transform.

We have skipped morphological processing step of Khare et al. [17], because here we do not need much accuracy in segmentation. Result of segmentation algorithm is shown in Fig. 2, for frame no. 100 of Caviar video sequence. It is clear that segmentation algorithms works well.

5.2 Algorithm for tracking

Daubechies complex wavelet transform have property of better edge representation and shift-invariance, whereas Zernike moment have properties of translation and rotation invariant. The proposed tracking algorithm exploits these properties.

The tracking algorithm searches the object in next frame according to its predicted centroid value, which is computed from the previous four frames. In the proposed algorithm it is

Fig. 3 Distance assignment in different direction
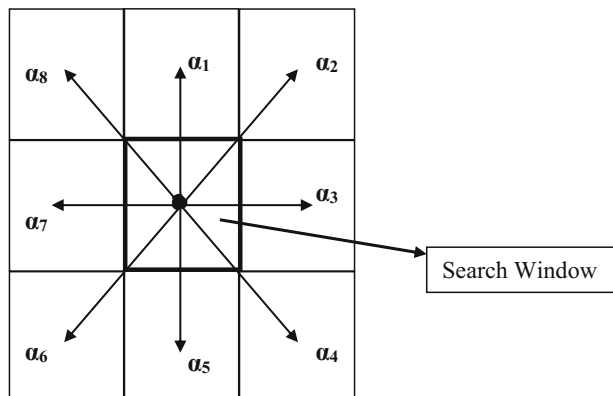
assumed that the frame rate is adequate and the size of the object should not change between adjacent frames. Complete tracking algorithm is given below-
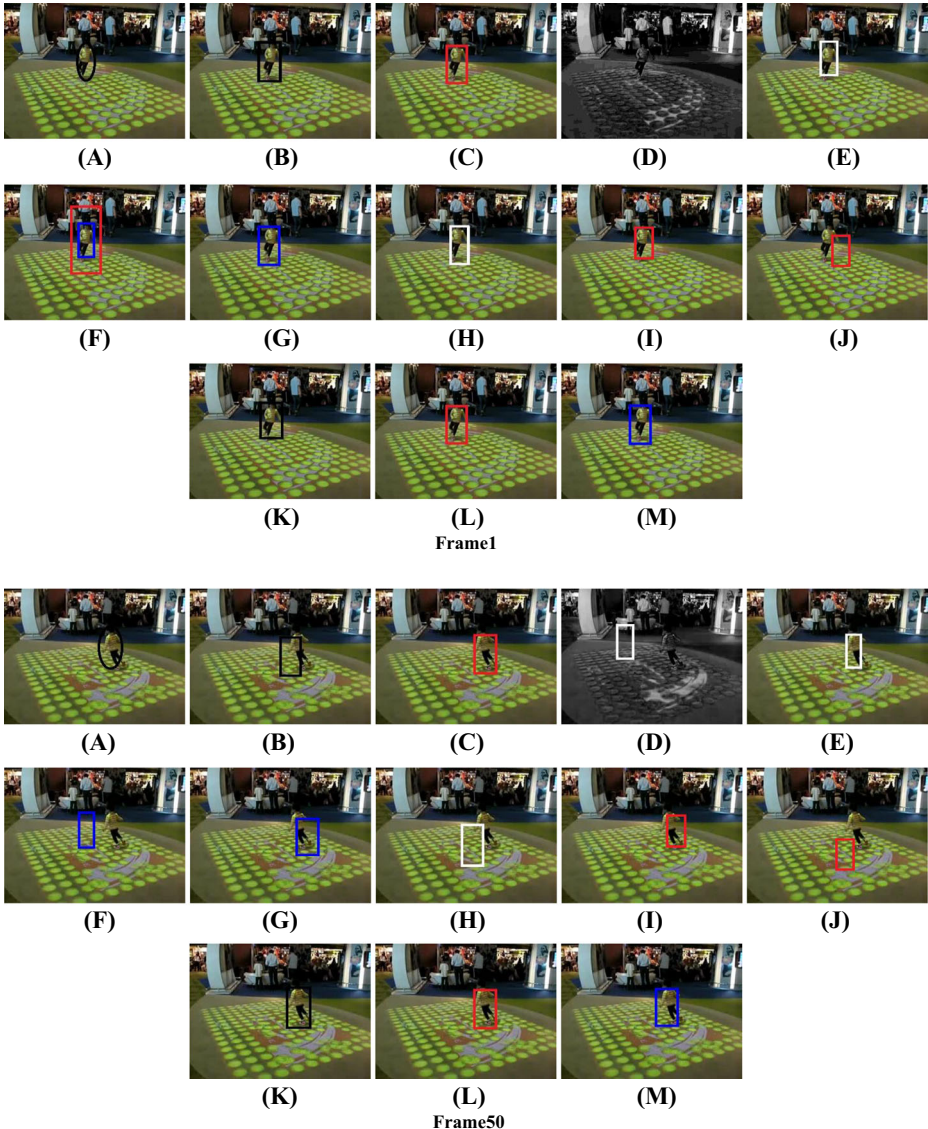
Step 1:
```
if frame_num=1
```


Fig. 4 Tracking in Child video for frame nos. 1 to 450 in steps of 50 frames [**A** tracking method proposed by Zhong et al. [41], **B** tracking method proposed by Babenko et al. [1] **C** tracking method proposed by Zhang et al. [40], **D** tracking method proposed by Ning et al. [22], **E** tracking method proposed by Porikli et al. [26] **F** tracking method proposed by Ning et al. [23], **G** tracking method proposed by Liu et al. [21] **H** tracking method proposed by Khare and Tiwary [18] **I** The proposed tracking method with search window direction mechanism, **J** The proposed tracking method without search window direction mechanism, **K-M**: modified methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], by adding search window direction mechanism]

(A) **(B)** **(C)** **(D)** **(E)**

(F) **(G)** **(H)** **(I)** **(J)**

(K) **(L)** **(M)**

Frame 100

(A) **(B)** **(C)** **(D)** **(E)**

(F) **(G)** **(H)** **(I)** **(J)**

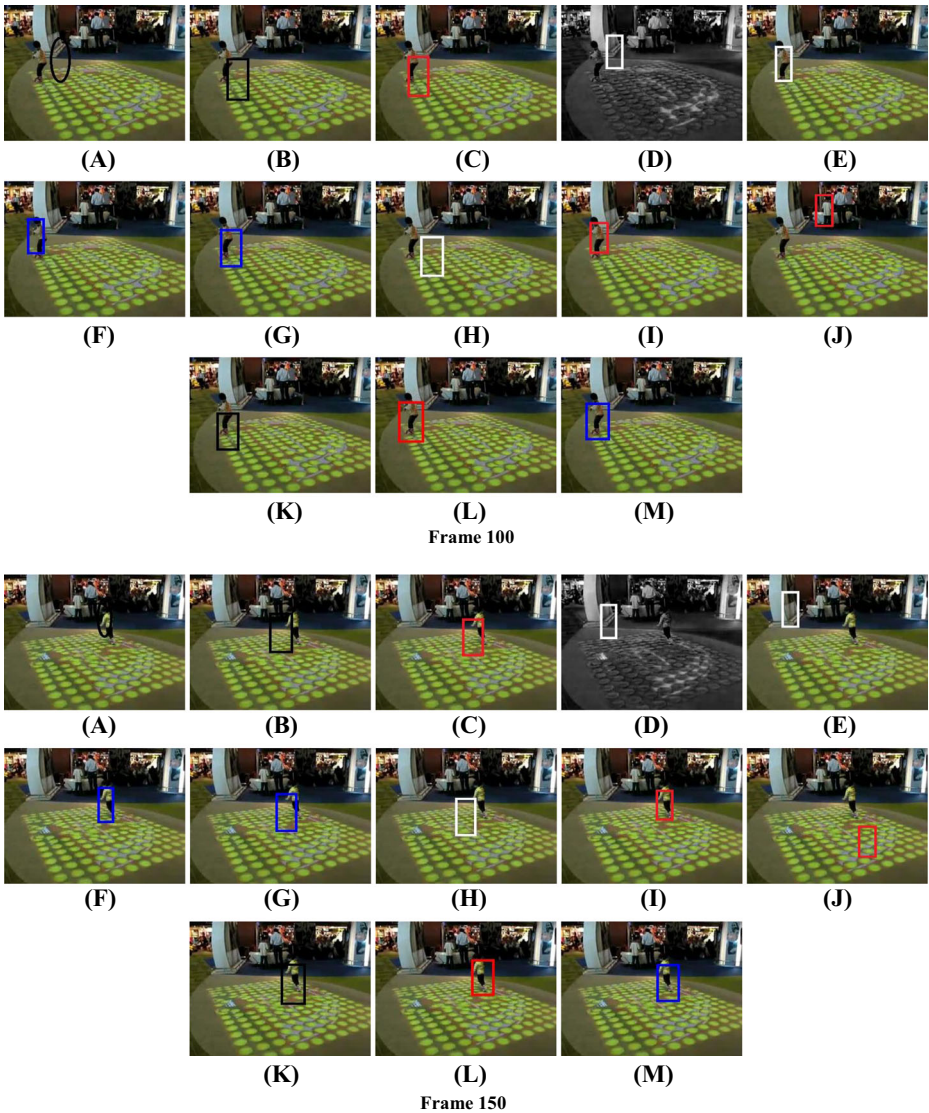**(K)** **(L)** **(M)**

Frame 150

Fig. 4 continued.

Segment the first frame, by the method described in Khare et al. [17], and re summarized in subsection 5.1 of this paper.

Make the bounding box around the object with centroid at (C_1, C_2) and compute the energy of coefficients of Zernike moment of Daubechies complex wavelet coefficients of the bounding box, say Z

$$Z = \sum_{(i,j) \in bounding\_box} \left| coeff_{i,j} \right|^2$$

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 200

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 250
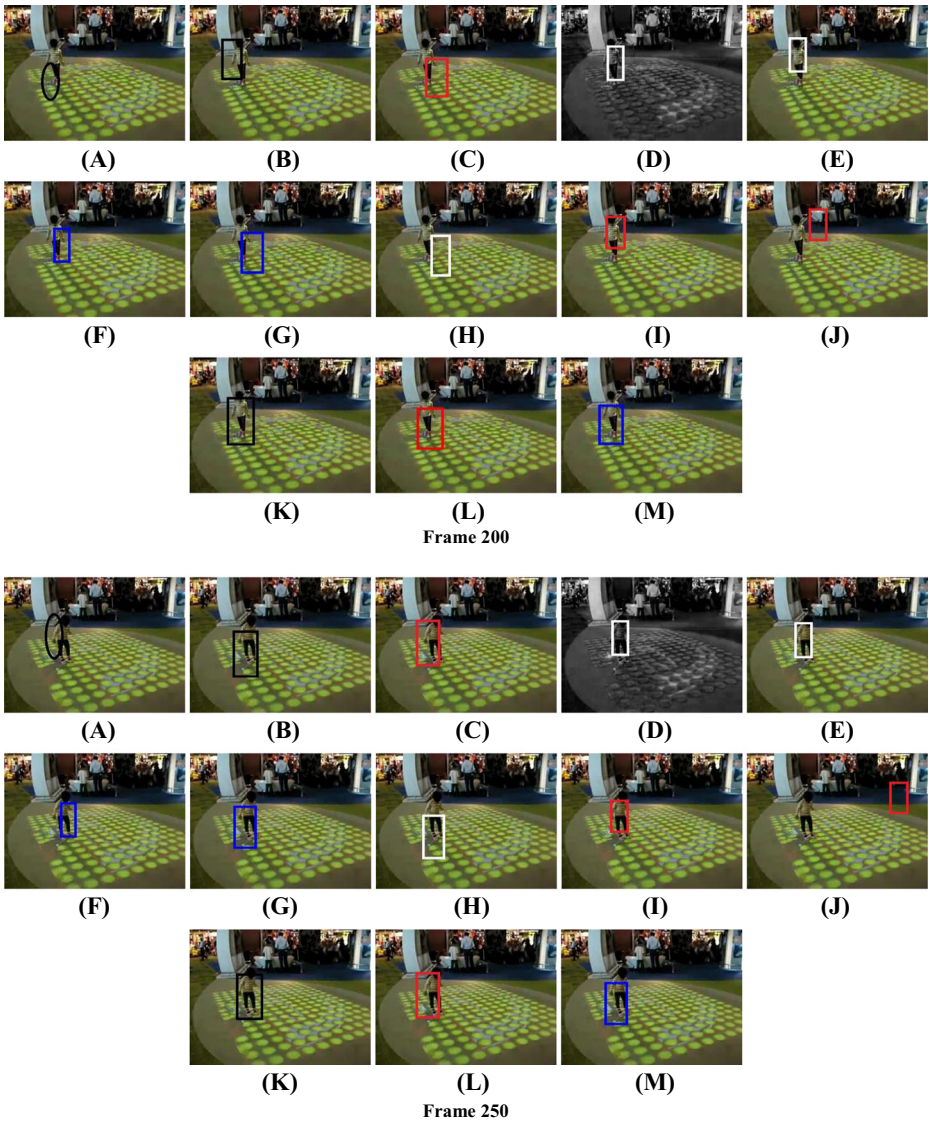
Fig. 4 continued.

where, $coeff_{i,j}$ are the Zernike moment of Daubechies complex wavelet coefficients at $(i,j)^{th}$ points.

Step 2:

for frame_num=2 to end_frame do

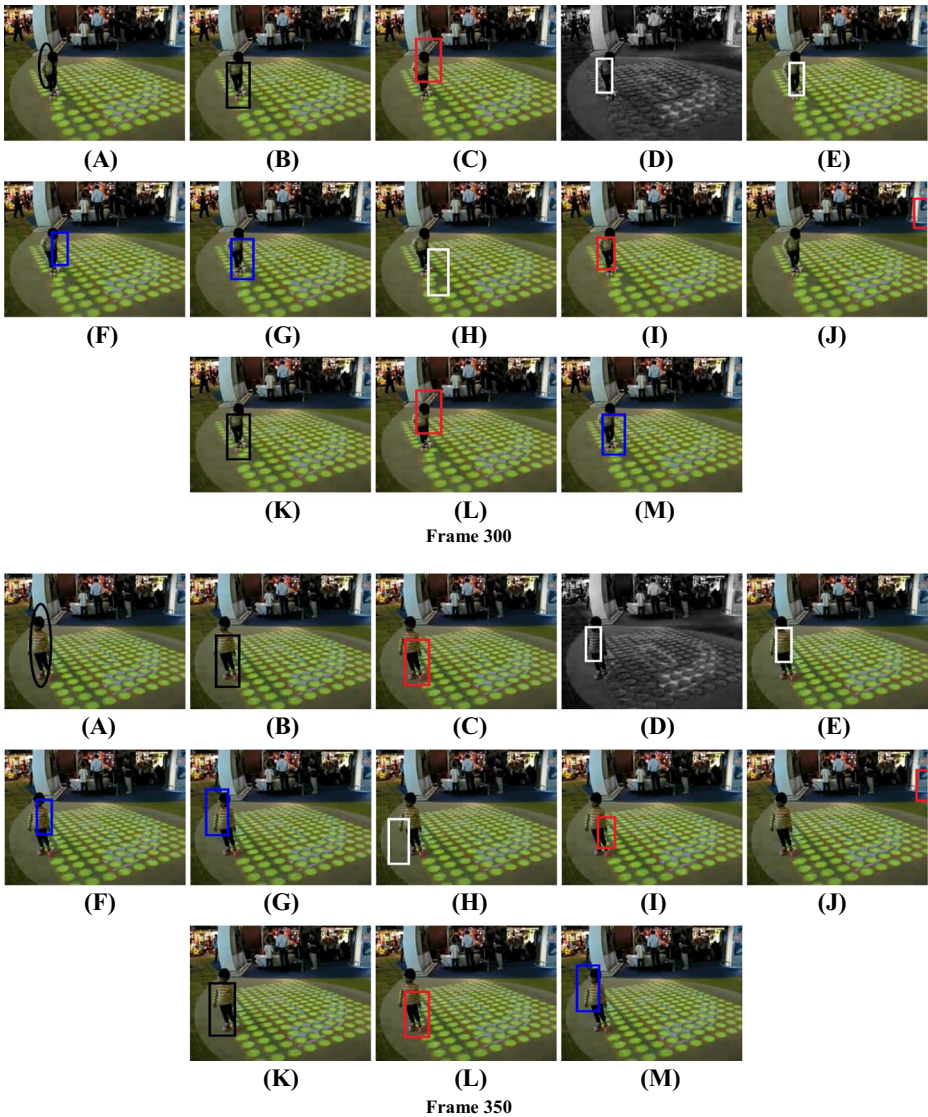Compute the Zernike moment of Daubechies complex wavelet coefficients of the frame, say $coeff_{i,j}$

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 300

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 350

Fig. 4 continued.

```
Search_region=32 (in pixels)
if frame_number>4
 Predict the centroid (C₁, C₂) of the object in current frame
```
with help of centroids of objects of previous four frames, basic
Newton's equations of motion, and search window direction mechanism
as described in section 5.3.
```
    end if
    for i=− search_region to+search_region do
```

(A) **(B)** **(C)** **(D)** **(E)**

(F) **(G)** **(H)** **(I)** **(J)**

(K) **(L)** **(M)**

Frame 400

(A) **(B)** **(C)** **(D)** **(E)**

(F) **(G)** **(H)** **(I)** **(J)**

(K) **(L)** **(M)**

Frame 450
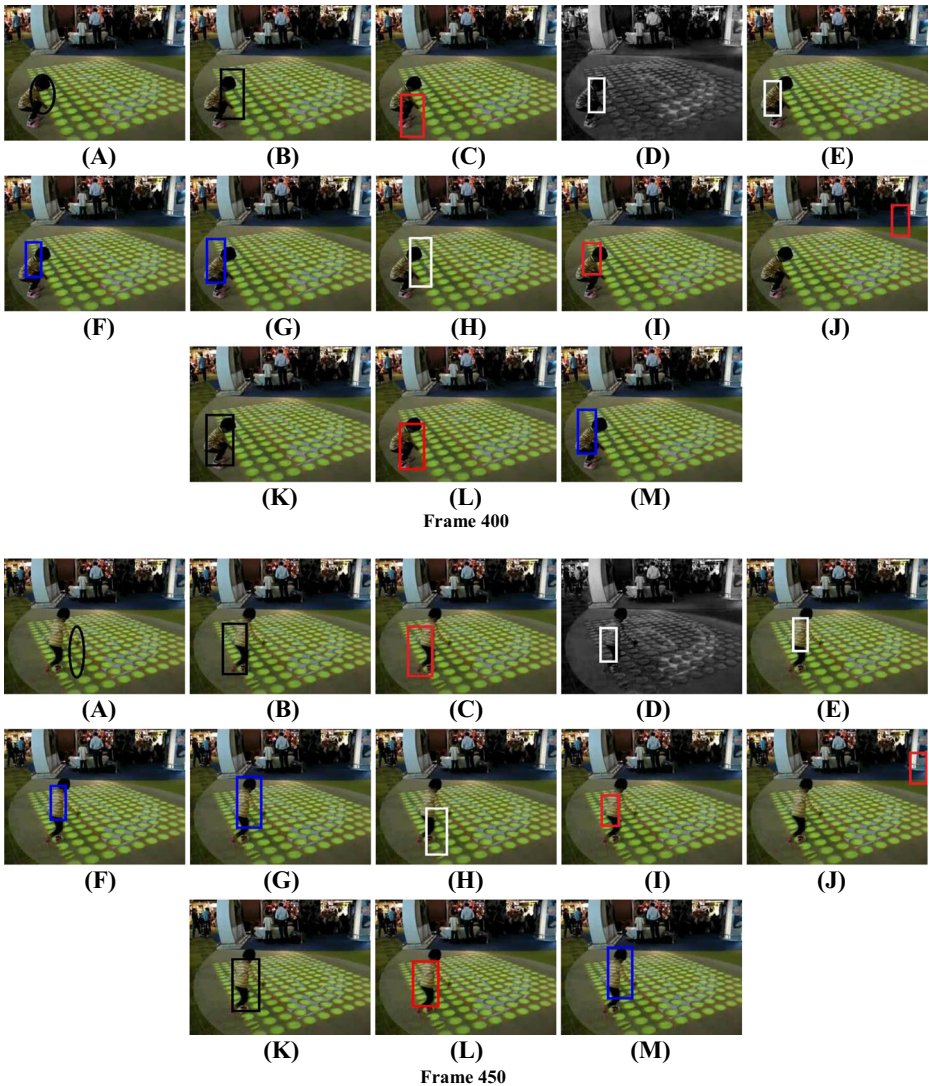
Fig. 4 continued.

```
for j=− search_region to+search_region do
C_new_1=C_1+i; C_new_2=C_2+j;
Make a bounding_box with centroid (C_new_1, C_new_2)
 Compute the difference of energy of the bounding_box, with Z,
say d_i, j
 for end
 for end
```

```
find minimum{d_i,_j} and its index, say (index_x, index_y)
C_1=C_1+index_x; C_2=C_2+index_y
```


(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 1

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 50
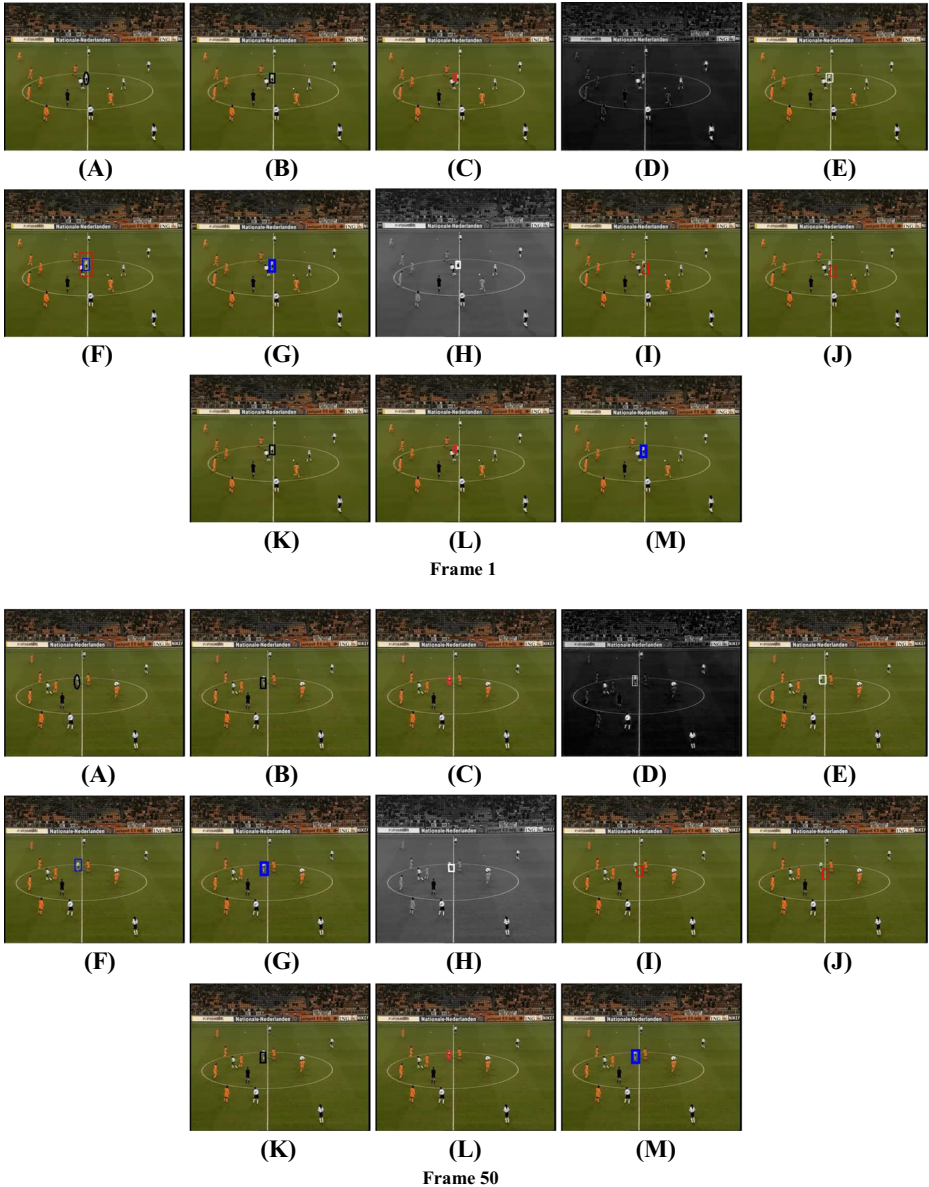
Fig. 5 Tracking in Soccer video for frame nos. 1 to 300 in steps of 50 frames [**A** tracking method proposed by Zhong et al. [41], **B** tracking method proposed by Babenko et al. [1] **C** tracking method proposed by Zhang et al. [40], **D** tracking method proposed by Ning et al. [22], **E** tracking method proposed by Porikli et al. [26] **F** tracking method proposed by Ning et al. [23], **G** tracking method proposed by Liu et al. [21] **H** tracking method proposed by Khare and Tiwary [18] **I** The proposed tracking method with search window direction mechanism, **J** The proposed tracking method without search window direction mechanism, **K-M**: modified methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], by adding search window direction mechanism]
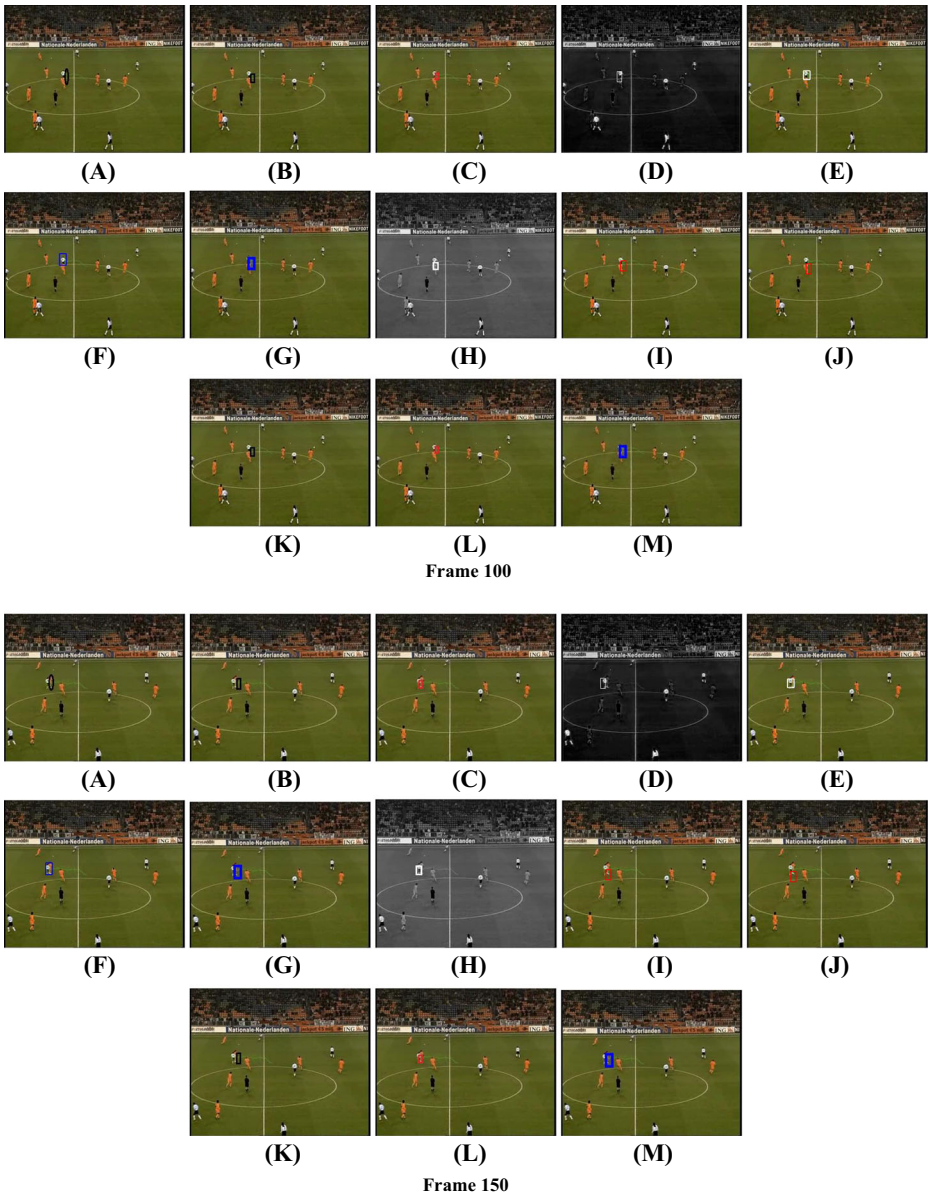
(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 100

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 150

Fig. 5 continued.

```
        Make the object in current frame with bounding_box with
centroid (C₁, C₂) and energy of bounding_box, Z as
```
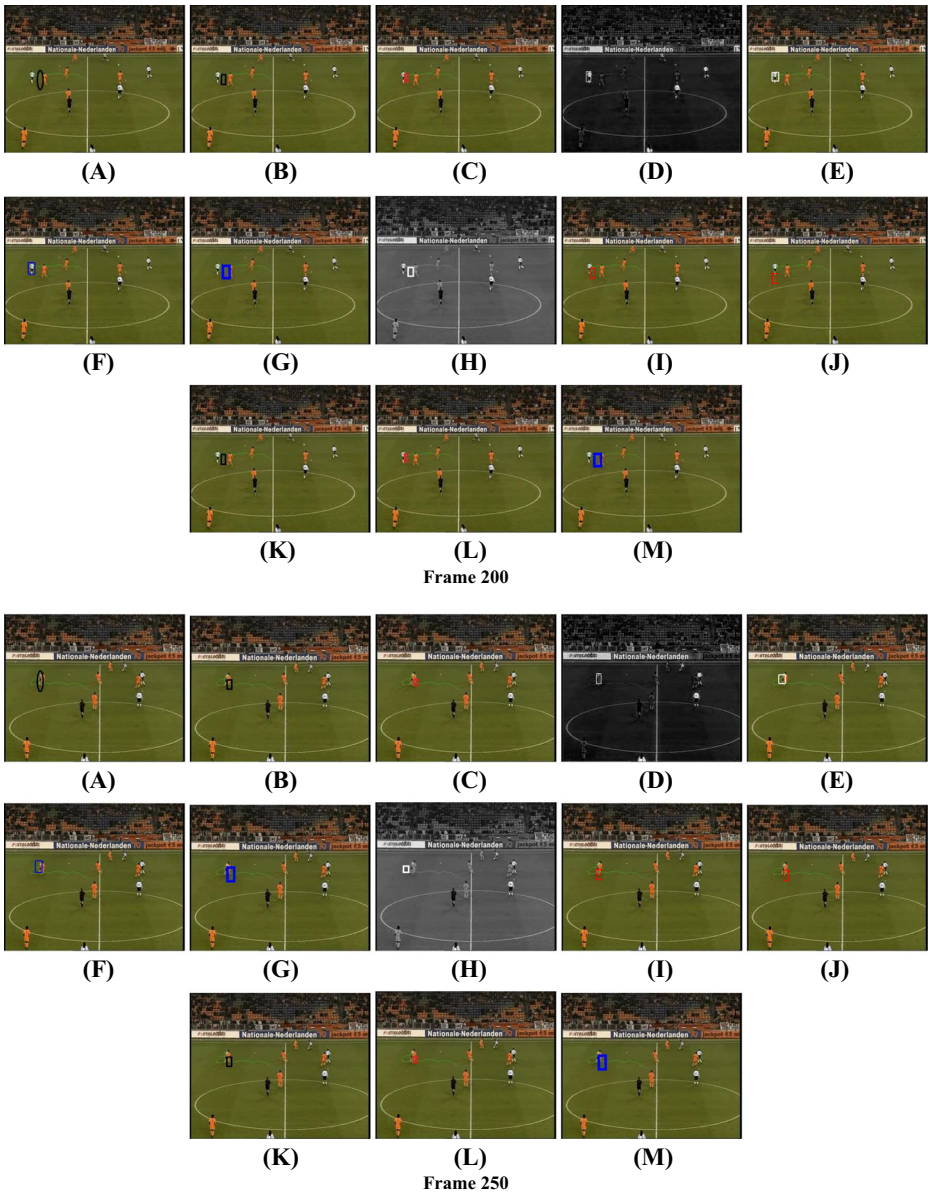
$$Z = \sum_{(i,j) \in bounding\_box} \left| coeff_{i,j} \right|^2$$

```
    end for
```

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 200

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 250

Fig. 5 continued.

5.3 Algorithm for search window direction for next frame

The change of object location requires an efficient and adaptive search window direction mechanism due to following reasons [14]:

(i)   The proper search window location ensures that the object always lies within the search area, hence, this reduces the loss of object inside the search window.

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 300
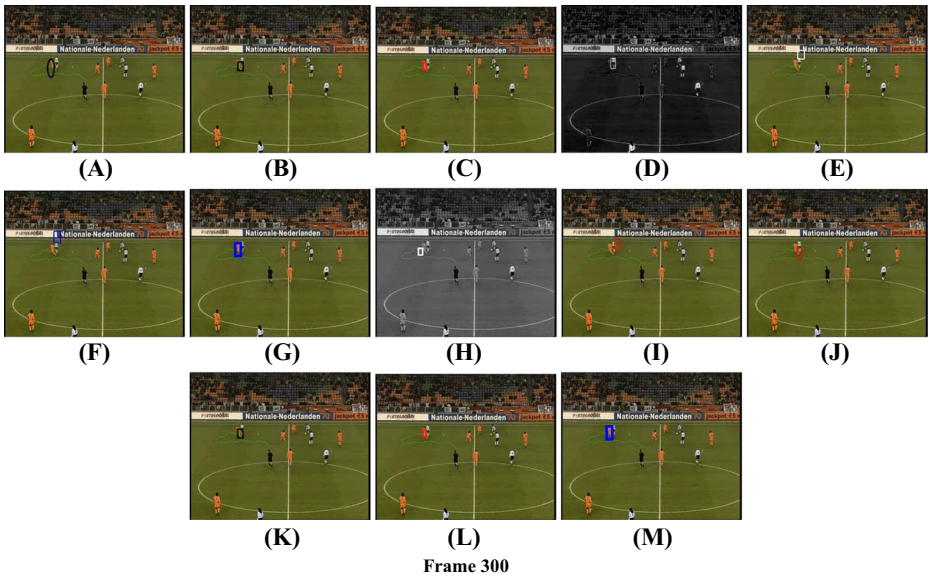
Fig. 5 continued.

(ii) If a tracked object is occluded by another object, then by use of direction of motion may reduce the occlusion problem.

In this paper, we have taken approach of Khansari et al. [14] for search window direction, which estimate the direction of motion of the object to update the location of the search window by using interframe texture analysis technique which is described by Seferidis and Ghanbari [29]. To find the direction the object motion temporal difference histogram [29] of two consecutive frame is used. Coarseness and directionality of the frame difference of two consecutive frames can be computed from temporal difference histogram. At last direction the motion is estimated from the use of temporal difference histogram of coarseness and directionality.

Temporal difference histogram – The temporal difference histogram of two consecutive frame is defined by absolute difference of gray level values of corresponding pixels at the two frames [14].

Let us consider that current search window $SA_t(x,y)$ at frame t and new search window $SA_{t+1}(x,y)$ determined by displacement value $\alpha = (\Delta x, \Delta y)$ of the current search window in the next frame. In search window, we assume that L_x and L_y are the width and height of the search window respectively. Two search windows should be of same size, then absolute temporal difference (ATD_α) of the two search windows is –

$$ATD_\alpha = \left| SA_t(x,y) - SA_{t+1}(x + \Delta x, y + \Delta y) \right| \tag{5.1}$$

After computing absolute temporal difference, we calculate the histogram of the value of ATD_α, with M bins, were M is the number of gray levels in each frame. We have used value of M is 256 for 8-bit image. Then the histogram values are normalized with respect to the number

of pixels in the search window to obtain the probability density function of each gray level value $p_\alpha(i):i=0,1,....M-1$. Number of pixels in the search window will be $(L_x \times L_y)$.

Search window direction – we have assumed that the search window is of a rectangular block, and consider eight different blocks at various direction, with distance α_i, from the center
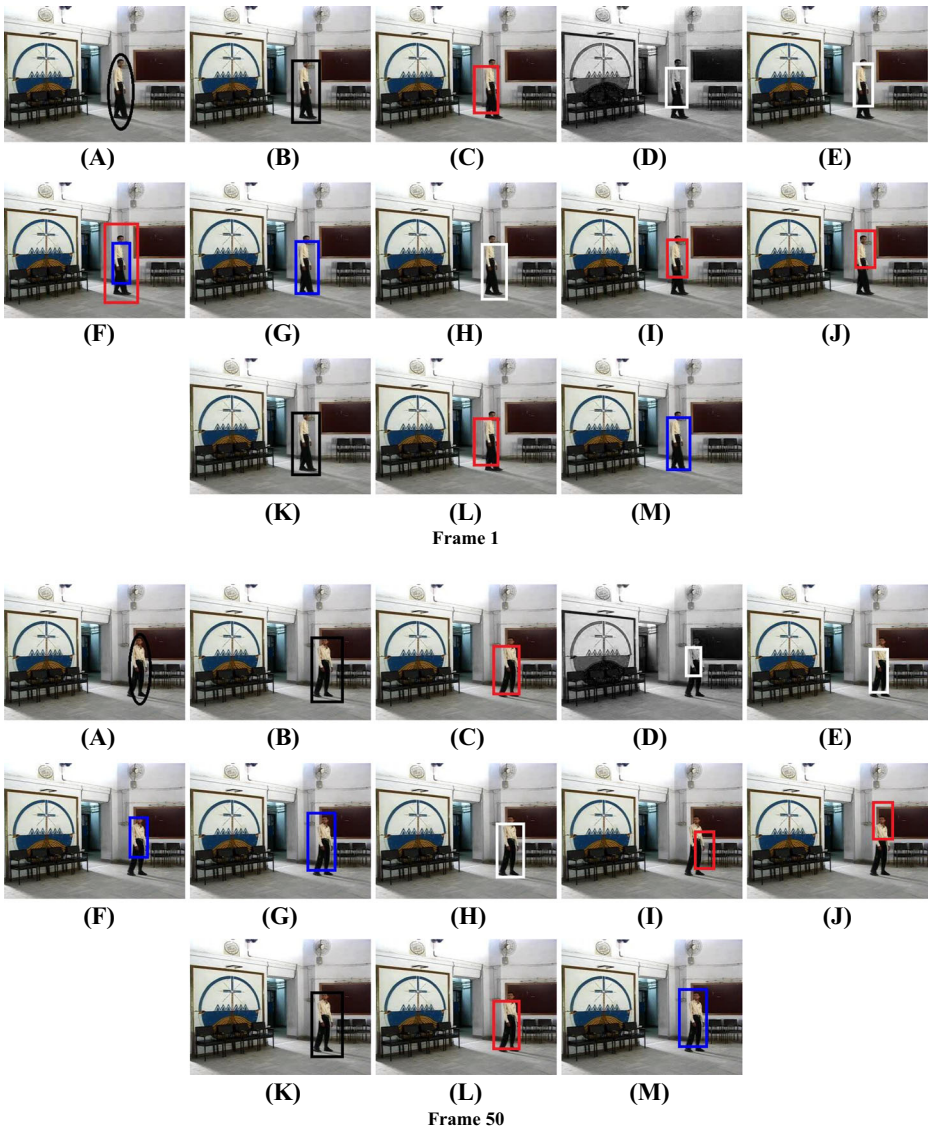

Fig. 6 Tracking in Gallery video for frame nos. 1 to 450 in steps of 50 frames [**A** tracking method proposed by Zhong et al. [41], **B** tracking method proposed by Babenko et al. [1] **C** tracking method proposed by Zhang et al. [40], **D** tracking method proposed by Ning et al. [22], **E** tracking method proposed by Porikli et al. [26] **F** tracking method proposed by Ning et al. [23], **G** tracking method proposed by Liu et al. [21] **H** tracking method proposed by Khare and Tiwary [18] **I** The proposed tracking method with search window direction mechanism, **J** The proposed tracking method without search window direction mechanism, **K-M**: modified methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], by adding search window direction mechanism]

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 100

(A) (B) (C) (D) (E)

(F) (G) (H) (I) (J)

(K) (L) (M)

Frame 150

Fig. 6 continued.

of current frame search window, as shown in Fig. 3. Calculate temporal difference histogram p_{α_i} for each block with respect to original search window block. After calculation of temporal difference histogram, we computer Inverse Difference Moment (IDM). IDM is the measure of homogeneity and it is defined as

$$IDM = \sum_{I=0}^{M-1} \frac{p_{\alpha}(i)}{i^2 + 1} \qquad (5.2)$$

Fig. 6 continued.

To derive the motion direction from texture direction, the direction that maximizes IDM should be found [27].

$$IDM_{\max} = \max\{IDM_i\}\, i = 1, 2, \ldots 8 \qquad (5.3)$$

The maximum value of *IDM*, *IDM* $_{\max}$ indicates that the frame difference is more homogenous in that direction that in the others, implying that the corresponding block in the successive frame are more correlated.
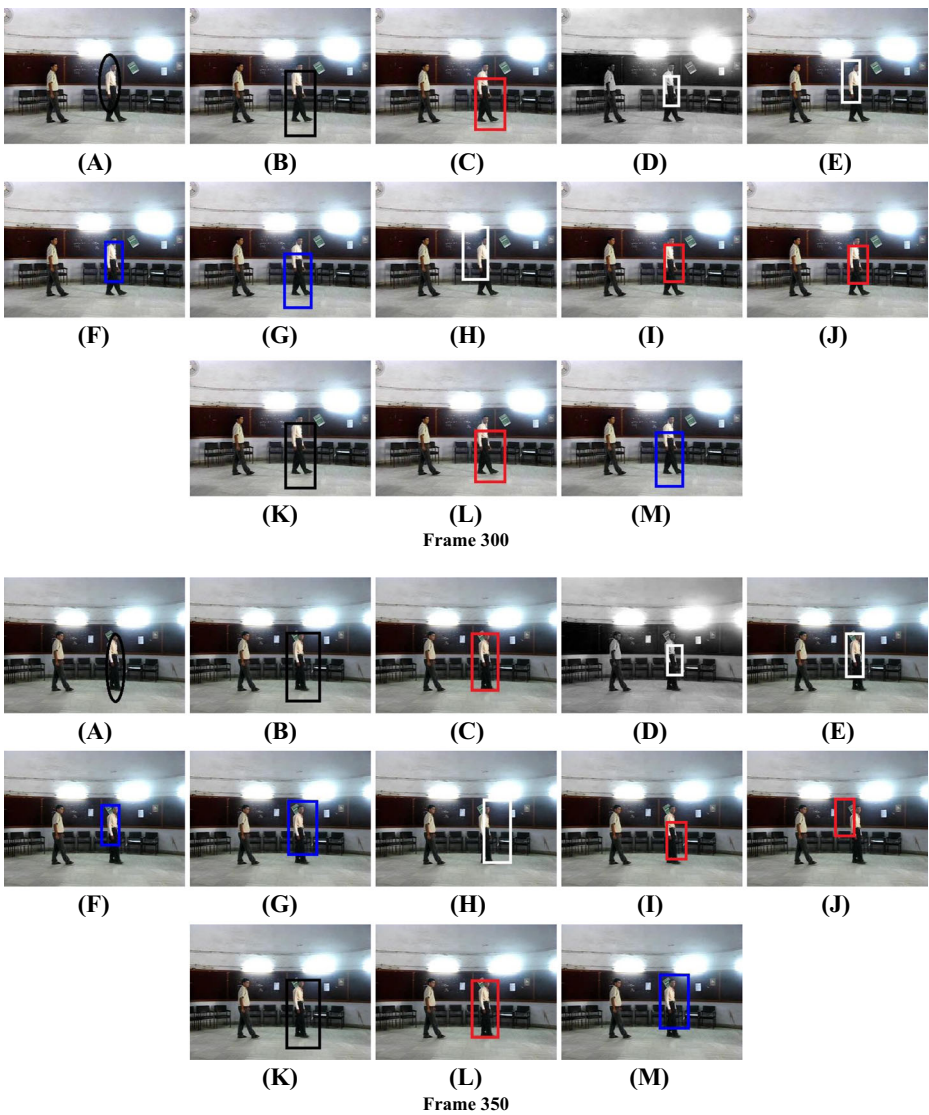
Fig. 6 continued.

6 Experimental results

The proposed method for object tracking, as described in section 5, is implemented using MATLAB and has been applied on a number of video sequences. Results are being presented here for four representative video sequences viz. Child video sequence, Soccer video sequence, Gallery video sequence and PETS video sequence.

In the proposed method, first we segment the first frame of video according to the method described in subsection 5.1. We draw an object window that fully covers the object with

(A) (B) (C) (D) (E) (F) (G) (H) (I) (J) (K) (L) (M)
Frame 400

(A) (B) (C) (D) (E) (F) (G) (H) (I) (J) (K) (L) (M)
Frame 450
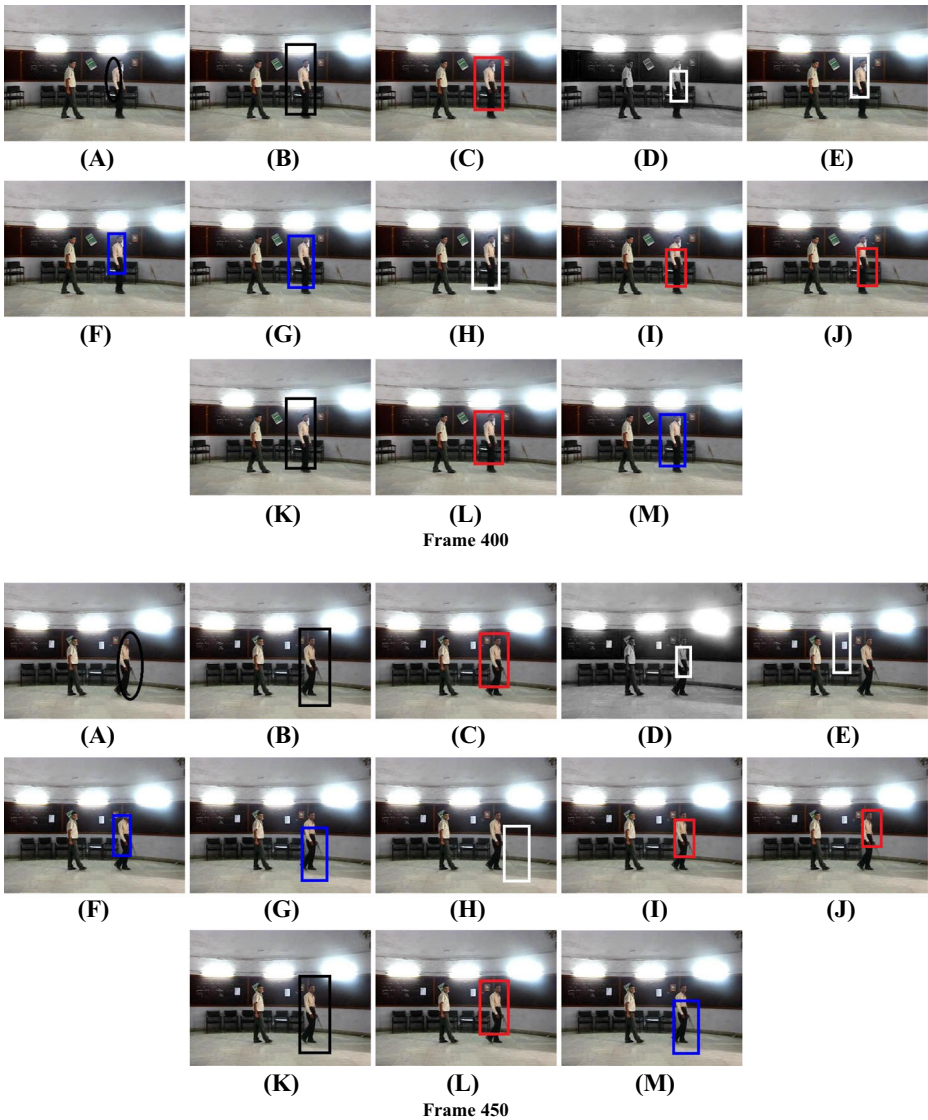
Fig. 6 continued.

centroid (C_1, C_2). Compute Daubechies complex wavelet coefficients of the object window. After computing wavelet coefficients we compute Zernike moments of complex wavelet coefficients followed by energy of Zernike moments of the object window. For the second frame and onwards, tracking of the object is performed within the search region. In our implementation we kept the search length equal to 32. The fixed search length 32 pixels of the search region allows the search window to move maximum of 32 pixels in all possible directions as per search window direction mechanism. In our proposed method, we have implemented idea of search window direction mechanism because proper search window

location ensures that the object always lies within the search area, hence this reduces the loss of object inside the search window. We have tested the proposed method in two cases (i) – the proposed method with search window direction mechanism, and (ii) – the proposed method without search window direction mechanism.
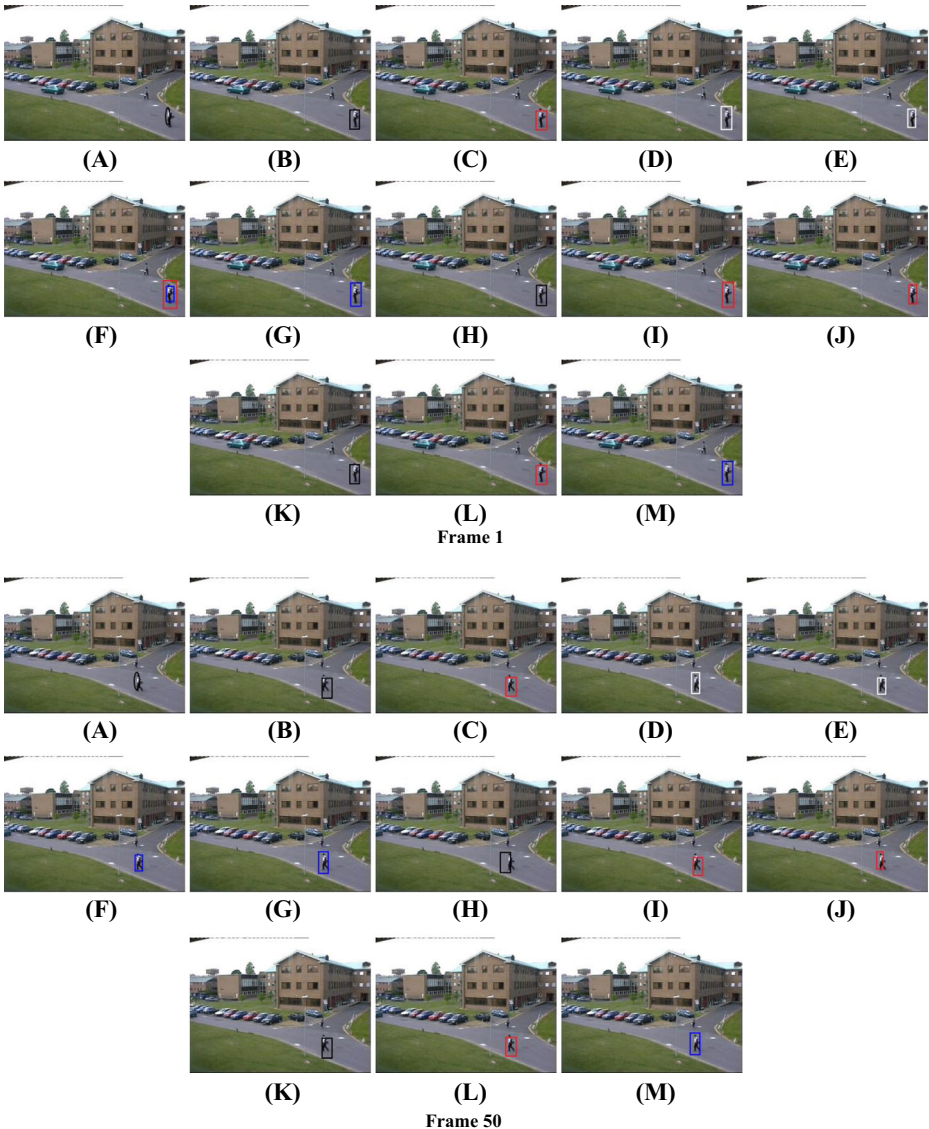

Fig. 7 Tracking in PETS video for frame nos. 1 to 400 in steps of 50 frames [**A** tracking method proposed by Zhong et al. [41], **B** tracking method proposed by Babenko et al. [1] **C** tracking method proposed by Zhang et al. [40], **D** tracking method proposed by Ning et al. [22], **E** tracking method proposed by Porikli et al. [26] **F** tracking method proposed by Ning et al. [23], **G** tracking method proposed by Liu et al. [21], **H** tracking method proposed by Khare and Tiwary [18], **I** The proposed tracking method with search window direction mechanism, **J** The proposed tracking method without search window direction mechanism, **K-M**: modified methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], by adding search window direction mechanism]

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 100

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)
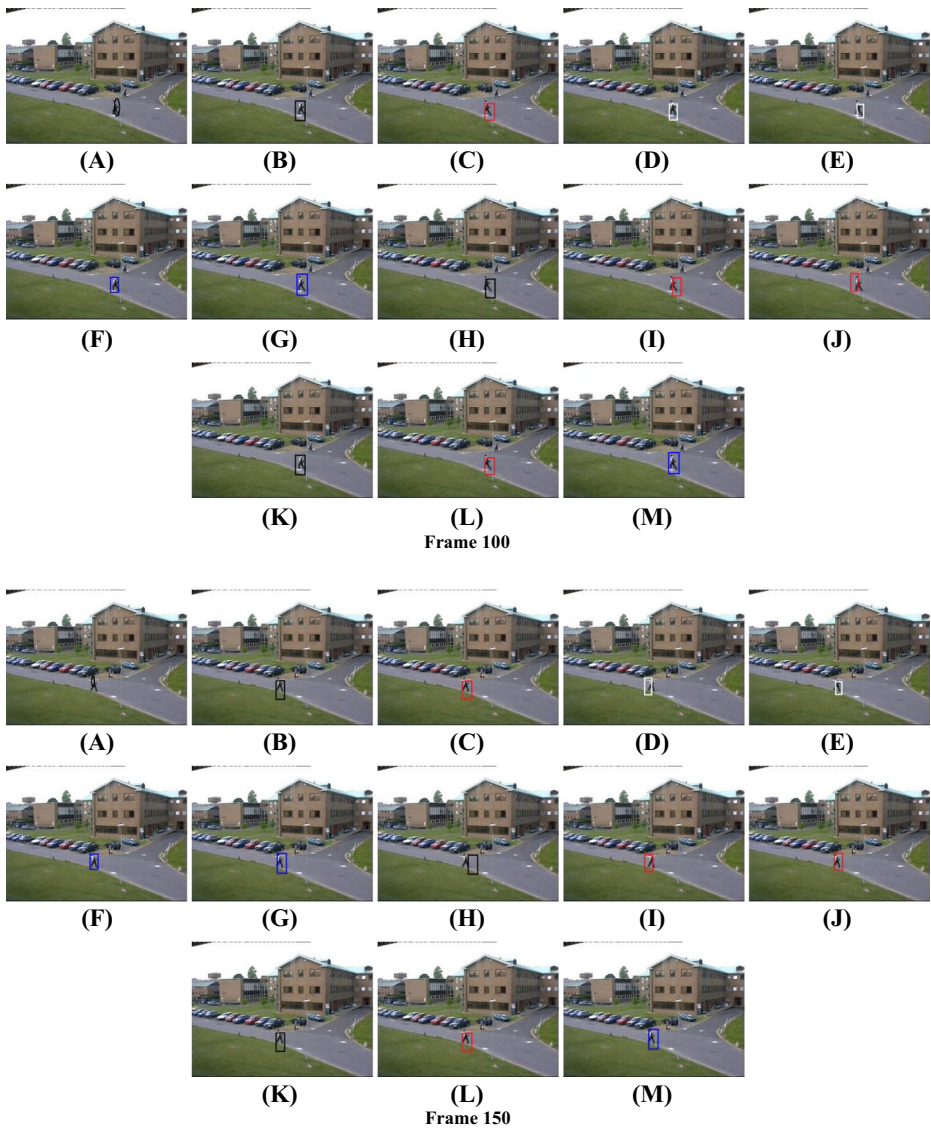
(K)   (L)   (M)

Frame 150

Fig. 7 continued.

Four experiments on Child video sequence, Soccer video sequence, Gallery video sequence and Pets video sequence are presented and analyzed one by one. In all experiments, we have presented and tested the proposed method and compared with tracking methods proposed by Zhong et al. [41], Babenko et al. [1], Zhang et al. [40], Ning et al. [22,23], Porikli et al. [26], Liu et al. [21], and Khare and Tiwary [18].

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 200

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 250
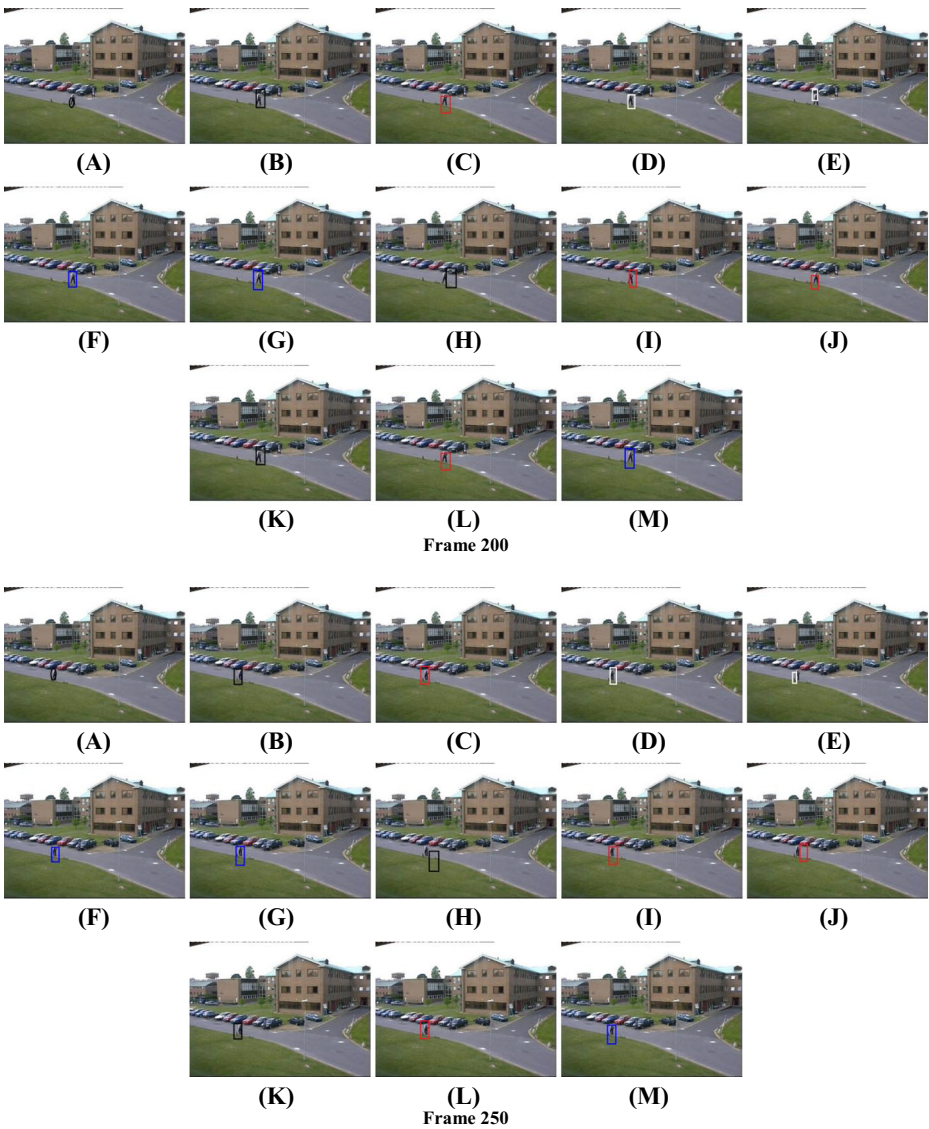
Fig. 7 continued.

6.1 Experiment 1

In experiment 1, experimental results for Child video sequence are shown. This video sequence contains 458 frames of frame size 352 x 288, but here we have shown results for frames with a difference of 50 frames, in Fig. 4.

In Child video, the child object is changing its position and motion abruptly. The abrupt motion of the child is difficult to track. In the first frame of this video, the bounding box fully covers the object. The movement and direction of movement changes in frame 2 and

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 300

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 350
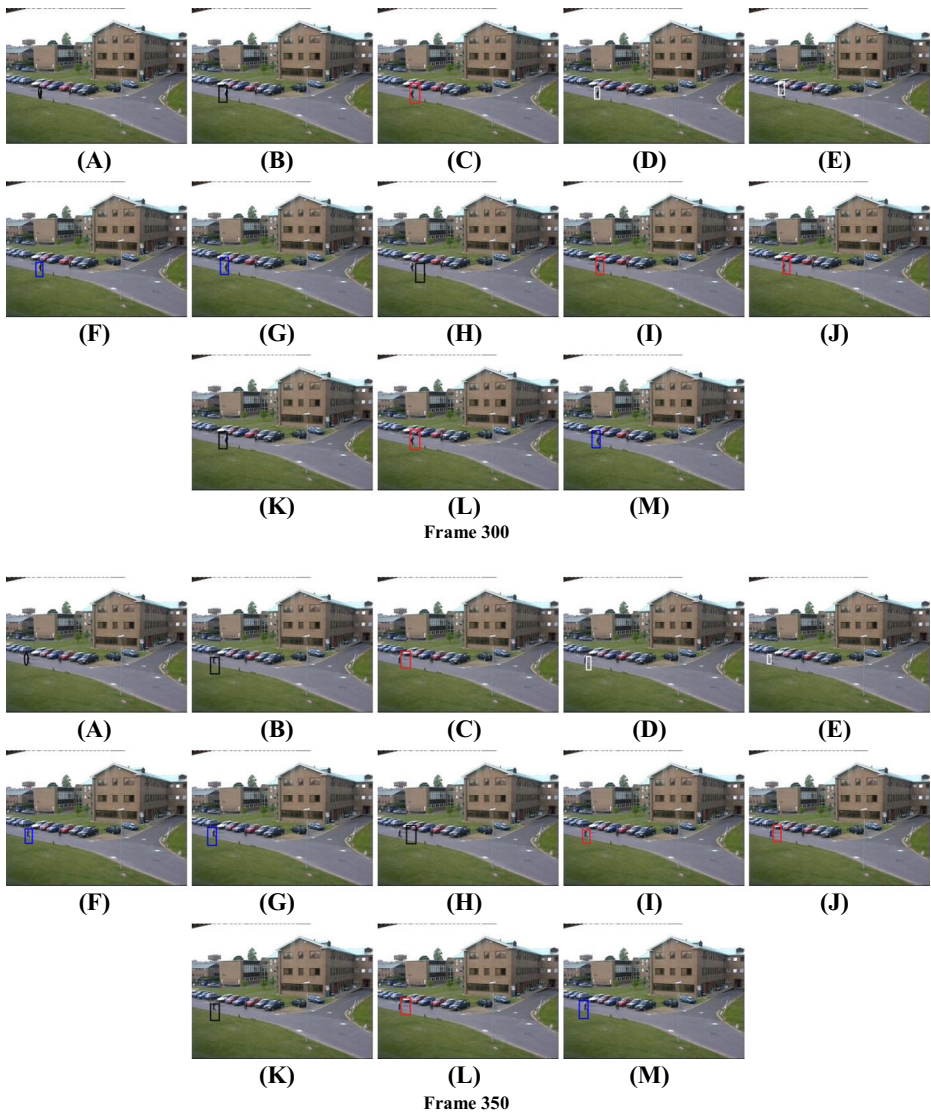
Fig. 7 continued.

continuing this the object stops in frame 100. Further we observe that from frame 150 the object reverses it direction of movement and comes at rest in frame 200. The proposed method with search window direction mechanism tracked the object in all these frames accurately. After frame 200, the child object shows some different poses such as slow motion, fast motion, walking pose, bending pose etc. and the proposed method with search window direction mechanism still keeps on tracking the object accurately.

On the other hand, in tracking method proposed by Zhong et al. [41], one can easily observed that there is clear track loss in frame 100 and frame 450, and in rest of frames

(A)   (B)   (C)   (D)   (E)

(F)   (G)   (H)   (I)   (J)

(K)   (L)   (M)

Frame 400
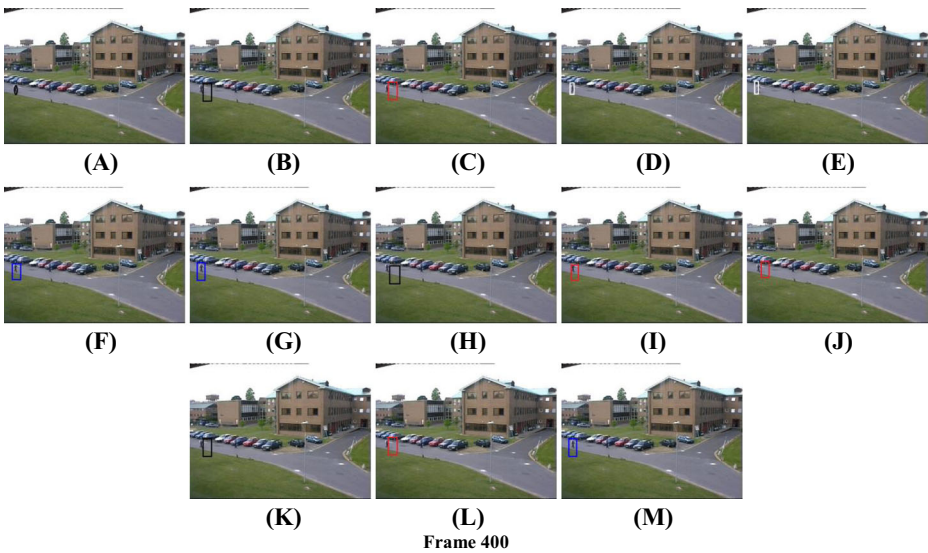
Fig. 7 continued.

bounding box is nearly exists to object. In the tracking method proposed by Ning et al. [22], the bounding box does not move correctly with child object when it changes its direction of motion, a clear miss track results can be seen in frame 50, 100 and 150. Tracking method proposed by Porikli et al. [26] again does not track the object correctly, which can be easily seen in frame 150, when object changes in direction suddenly. In tracking method proposed by Ning et al. [23], one can easily observe that there is a clear track loss in frame 50. In tracking method proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], no track loss is present in all frames, and performs better than method proposed by Zhong et al. [41] and Ning et al. [22,23]. Method proposed by Khare and Tiwary [18], which uses only Daubechies complex wavelet transform coefficients, not works well when objects changes its direction and tracking results is not good. In case of the proposed method anyone can easily see that proposed method with search window direction mechanism performs better than proposed method without search window direction mechanism. Most of tracking methods work good in simple and linear object motion, but here in this experiment the object is moving simple and linear as well as taking different abrupt directions and different poses. In this experiment the proposed method with search window direction mechanism performs better than other discussed method and no track loss is found. The proposed method with search window direction mechanism is able to track the object in different conditions such as slow/fast object motion, slight/rapid changes in object's orientation and poses. From this experimentation we have also seen that, when we use combination of two features Zernike moment and Daubechies complex wavelet transform, tracking results improves as comparison to use of single feature as shown in results of method proposed by Khare and Tiwary [18]. Modified methods, by adding search window direction mechanism, the methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], are comparable to the proposed method.

6.2 Experiment 2

In experiment 2, experimental results for soccer video sequence are shown. This video sequence contains 329 frames of size 352 × 288, but here we have shown results for frames with a difference of 50 frames, in Fig. 5

Soccer video sequence contains multiple objects. In Fig. 5 one can see that the proposed method completely covers objects in all frames, and no track loss is present there. Background condition in this video is complex, but it does not affect the prediction of object, because path of object is same throughout the video and no abruptly changes occurs. In addition to this, the size
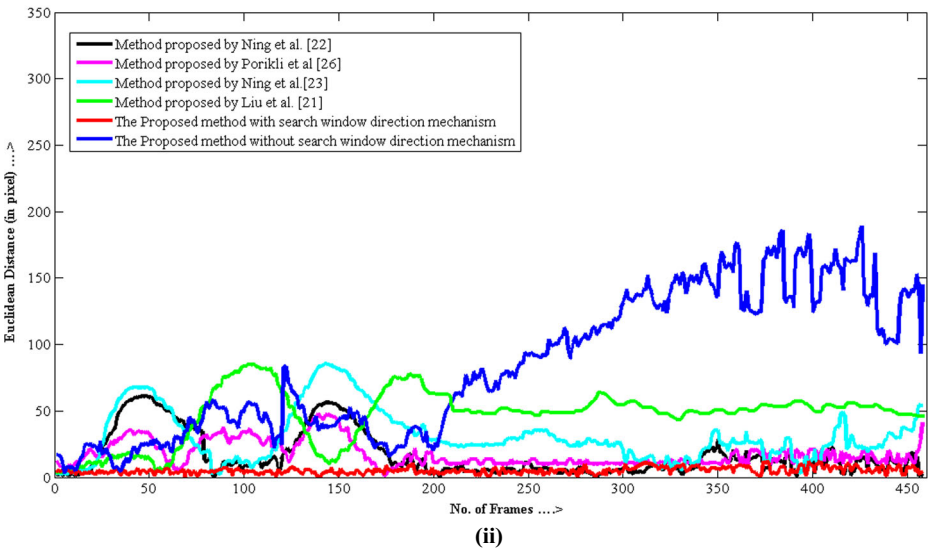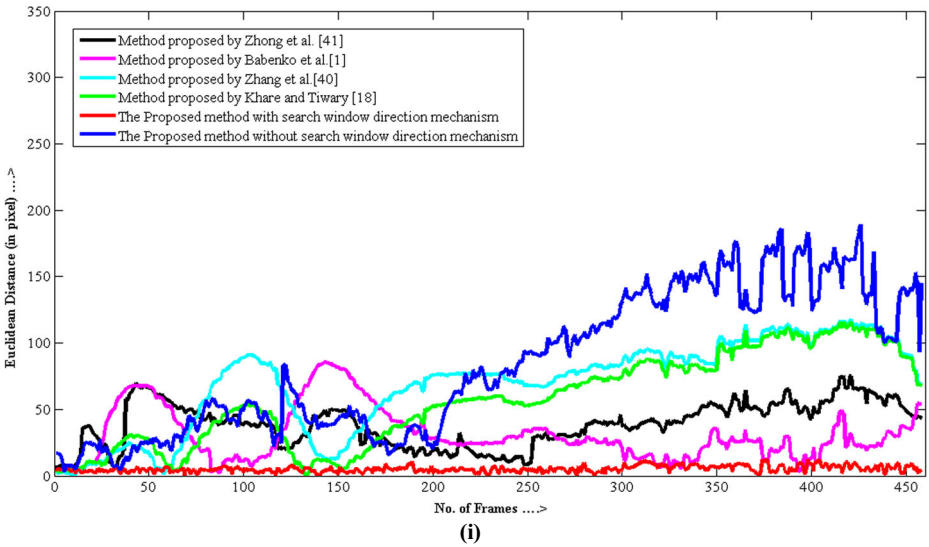
Fig. 8 Plot of Euclidean distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for child video sequence

of objects is very small in this video but the proposed method does not lose it tracks due to the size of objects. Tracking has also been tested on different objects of this video (results not given in paper), which also proves that the proposed method is suitable for any object in this video.

By applying other methods, in some frames the methods give satisfactory results whereas in others track loss is present. Tracking method proposed by Ning et al. [22] accurately tracks object upto frame 100 and start losing track slightly from frame 150. Bounding box remains still in some frames. Tracking method proposed by Ning et al. [23] track the object correctly upto frame 250 and losses the tracks when occlusion is present. Tracking method proposed by Porikli et al. [26] gives good result in comparison to method [22,23]. Method proposed by Zhong et al. [41], Babenko et al. [1], Zhang et al. [40] and Liu et al. [21] performs nearly equal and give better results in comparison to method [22,23,26]. Method proposed by Khare and Tiwary [18] not work well for small object and track loss exists from frame number 150 and onwards. The proposed method with search window direction mechanism and proposed method without search window direction mechanism performs nearly equal in this case. Modified methods, by adding search window direction mechanism, the methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], are comparable to the proposed method. From this video it is clear that, the proposed method perform well and can track the target correctly, even size of object is very small and another object occludes the target partially or fully.
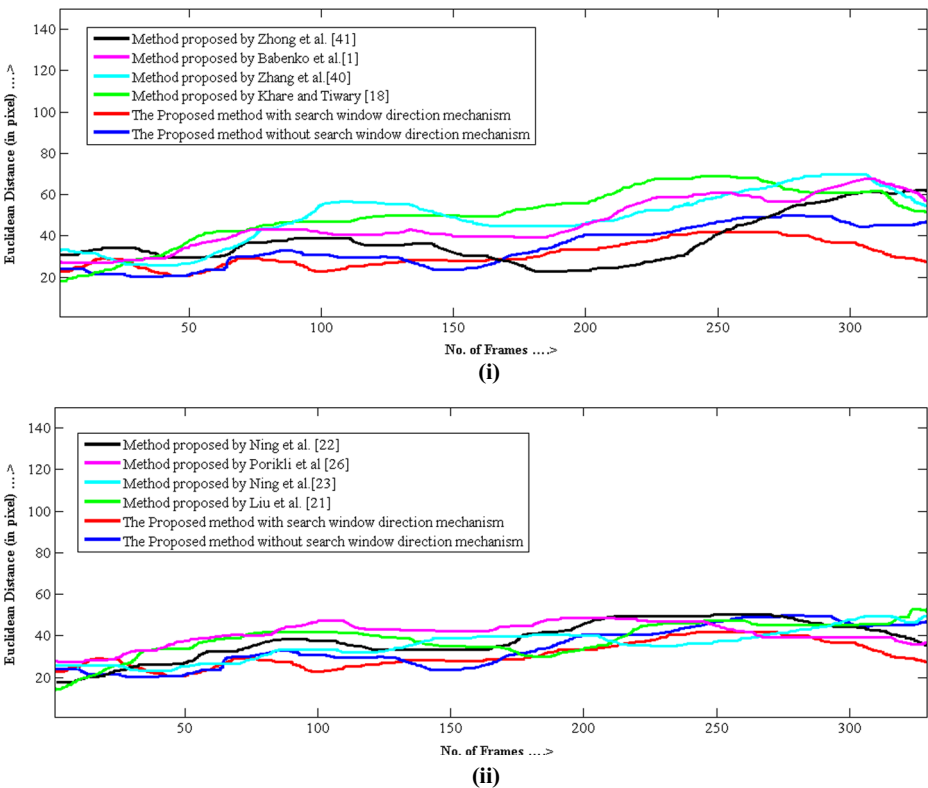

Fig. 9 Plot of Euclidean distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Soccer video sequence

6.3 Experiment 3

In experiment 3, experimental results for Gallery video sequence are shown. This video sequence is captured by author of this paper, which contains 450 frames of frame size 320 × 240, but here we have shown results for frame with a difference of 50 frame in Fig. 6.

In Gallery video, the object is slowing moving with moving background. Background is having varying lighting conditions and the camera is also moving. The proposed method tracks the object in all frames with no tracking loss. In this video background is changing and object is moving in linear manner. Due to this reason all other discussed methods [1,18,21–23,26,40,41] also performs comparatively equal to the proposed method. The proposed method with search window direction mechanism and the proposed method without search window direction mechanism perform equally in this video, because direction of object is linear. Modified methods, by adding search window direction mechanism, the methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], are comparable to the proposed method. From this video, it is clear that the proposed method tracks the object accurately even the changing background and varying lighting conditions.
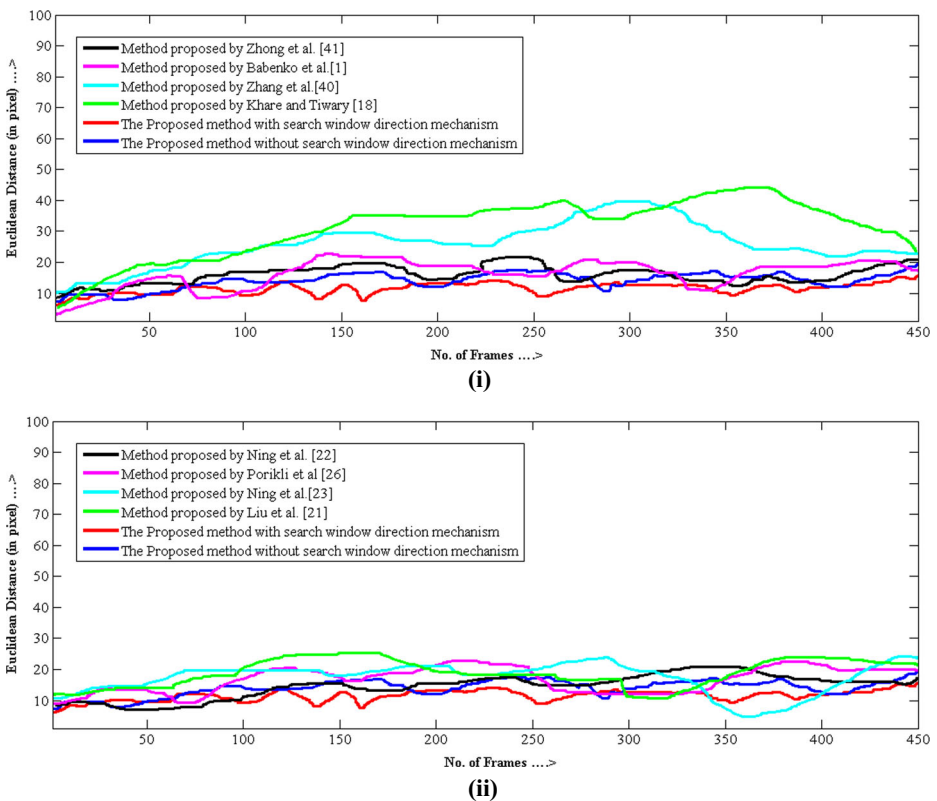

Fig. 10 Plot of Euclidean distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Gallery video sequence

6.4 Experiment 4

In experiment 4, experimental results for PETS video sequence are shown. This video sequence contains 432 frames of frame size 352 × 288, but here we have shown results for frame with a difference of 50 frame in Fig. 7.

In PETS video, the object is slowing moving in cluttered background. Background is having varying lighting conditions and the camera is also moving. The proposed method tracks the object in all frames with no tracking loss. In this video object is moving in linear manner. Due to this reason all other discussed methods [1,18,21–23,26,40,41] also performs comparatively equal to the proposed method. Some track loss exists in method proposed by Zhang et al. [40], Ning et al. [23], and Khare and Tiwary [18] due to clutter background and object color is nearly equal to background. The proposed method with search window direction mechanism and the proposed method without search window direction mechanism perform equally in this video, because direction of object is linear. Modified methods, by adding search window direction mechanism, the methods proposed by Babenko et al. [1], Zhang et al. [40] and Liu et al. [21], are comparable to the proposed method.
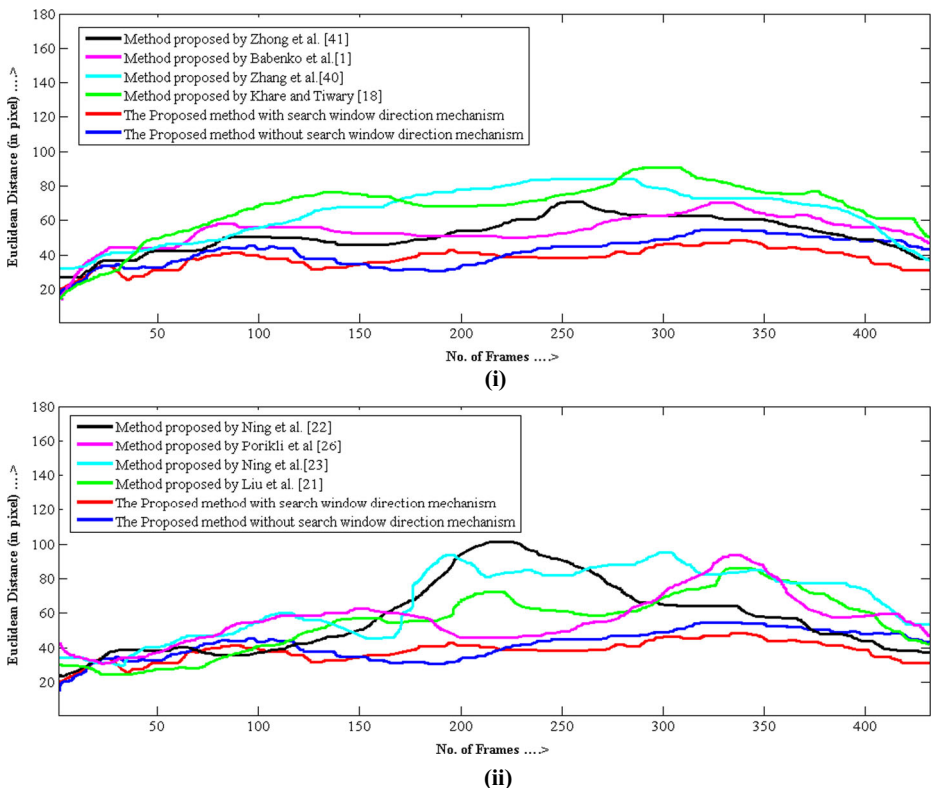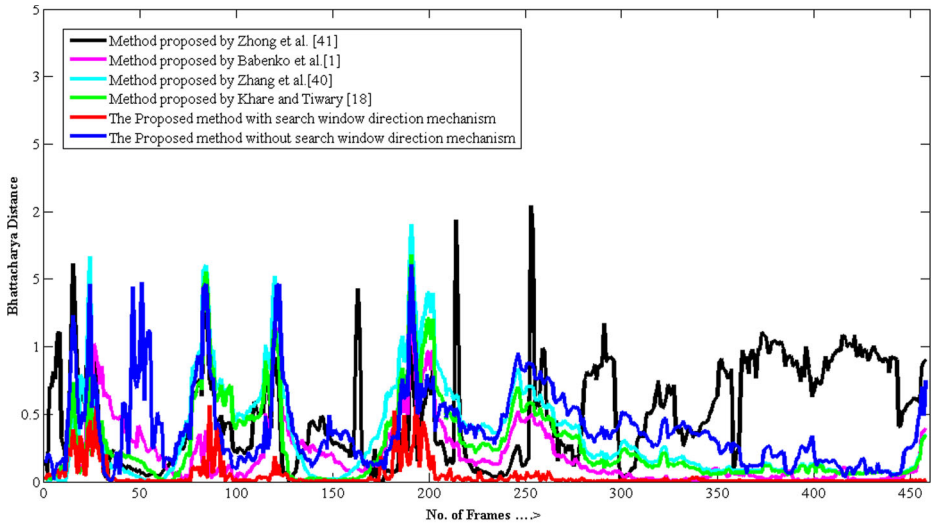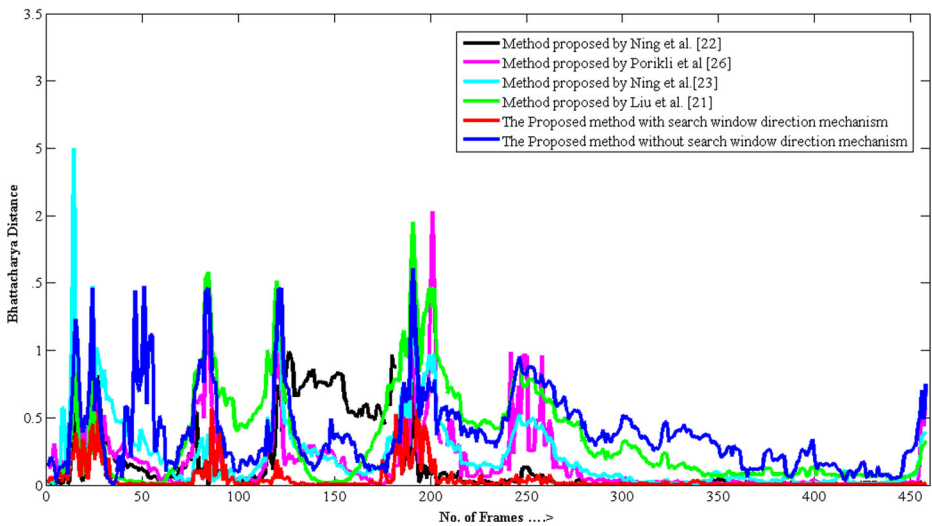

Fig. 11 Plot of Euclidean distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for PETS video sequence

7 Performance evaluation

Qualitative performance is not enough to judge the quality of any method, 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 to analyze performance of different methods. For quantitative comparison we have considered three different performance metrics: Euclidean distance, Bhattacharya distance, and Mahalanobis distance. All the simulation and

(i)

(ii)

Fig. 12 Plot of Bhattacharya distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for child video sequence

comparison of the proposed method were done on a machine with Intel 1.73 GHz Dual core processor with 2GB RAM using MATLAB 2012a software. We have performed quantitative performance comparison of the proposed method with other state-of-the-art methods proposed by Zhong et al. [41], Babenko et al. [1], Zhang et al. [40], Ning et al. [22,23], Porikli et al. [26], Liu et al. [21], and Khare and Tiwary [18]. We have shown Quantitative performance metrics results for Child video sequence, Soccer video sequence, Gallery video sequence and PETS video sequence.

7.1 Euclidean distance

The Euclidean distance between the computed centroid (X_c, Y_c) of tracked object and ground truth centroid (X_g, Y_g) is defined as

$$ED = \sqrt{\left(X_g - X_c\right)^2 + \left(Y_g - Y_c\right)^2} \tag{7.1}$$

Figures 8, 9, 10 and 11 shows plots of Euclidean distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] in case of Child video sequence, Soccer video sequence, Gallery video sequence and PETS video sequence respectively.
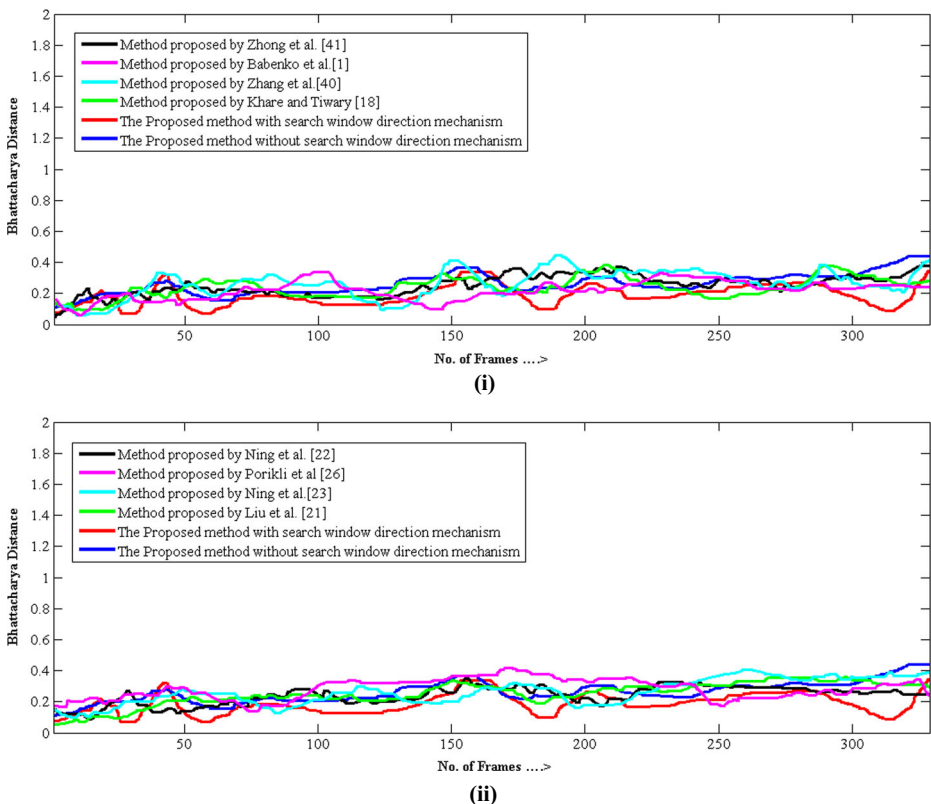

Fig. 13 Plot of Bhattacharya distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Soccer video sequence

7.2 Bhattacharya distance

The Bhattacharya distance is distance transform which is used as separability measure of tracked object region and ground truth object region. For two classes of ground truth and computed values, Bhattacharya distance (BD) is defined as

$$BD = \frac{1}{8} \left(mean_c - mean_g\right)^T \left[\frac{\text{cov}_g + \text{cov}_c}{2}\right]^{-1} \left(mean_c - mean_g\right)$$

$$+ \frac{1}{2} \ln \frac{\left|\left(\text{cov}_g + \text{cov}_c\right)\big/2\right|}{\left|\text{cov}_g\right|^{1/2} \left|\text{cov}_c\right|^{1/2}} \tag{7.2}$$

where, $mean_g$ and $mean_c$ are mean vector for ground truth object region and computed object region respectively, cov_g and cov_c are covariance matrix for ground truth object region and computed object region respectively.

Figures 12, 13, 14 and 15 shows plots of Bhattacharya distance for the proposed method and other state-of-the art methods [1,18,21–23,26,40,41] in case of Child video sequence, Soccer video sequence, Gallery video sequence and PETS video sequence respectively.
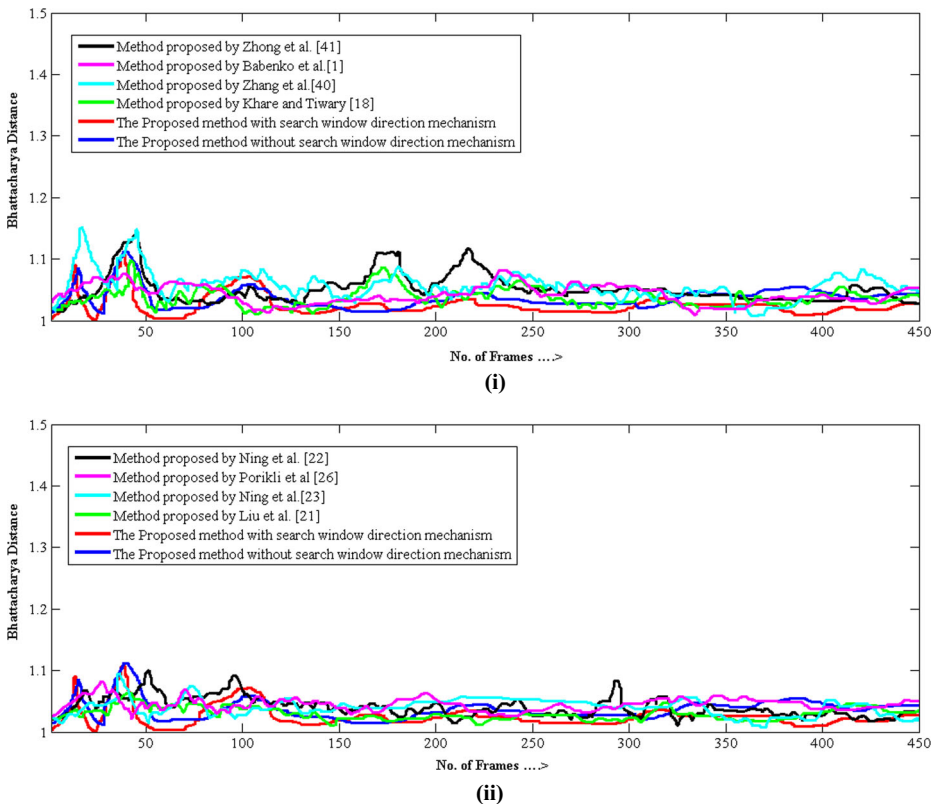

Fig. 14 Plot of Bhattacharya distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Gallery video sequence

7.3 Mahalanobis distance

Mahalanobis distance is based on correlations between variables by which different patterns can be identified and analyzed. It differs from Euclidean distance in the sense that it encounters the correlations of the data points. It is defined as a dissimilarity measure between two points $X=(X_g, Y_g)$ and $Y=(X_c, Y_c)$ with the covariance matrix C

$$MD = \sqrt{(X-Y)^T C^{-1} (X-Y)} \tag{7.3}$$

where, X and Y are centroid points of ground truth object and computed object respectively.

The covariance of two features (e.g. ground truth centroid and computed centroid) measures their tendency to vary together i.e. to co-vary. Consider two features, feature i (ground truth centroid points) and feature j (computed centroid points). Let $\{x(1,i), x(2,i), \ldots x(n,i)\}$ be a set of n examples of feature i, and let $\{x(1,j), x(2,j), \ldots \ldots x(n,j)\}$ be a set of n examples of feature j (i.e. $x(k,i)$ and $x(k,j)$ are features of same pattern, pattern k). let $m(i)$ be the mean of feature i, and $m(j)$ be the mean of feature j. Then the covariance of feature i and feature j is computed as

$$c(i,j) = \frac{\left\{ \left[x(1,i)-m(i)\right]\left[x(1,j)-m(j)\right] + \ldots\ldots\ldots + \left[x(n,i)-m(i)\right]\left[x(n,j)-m(j)\right] \right\}}{n-1} \tag{7.4}$$
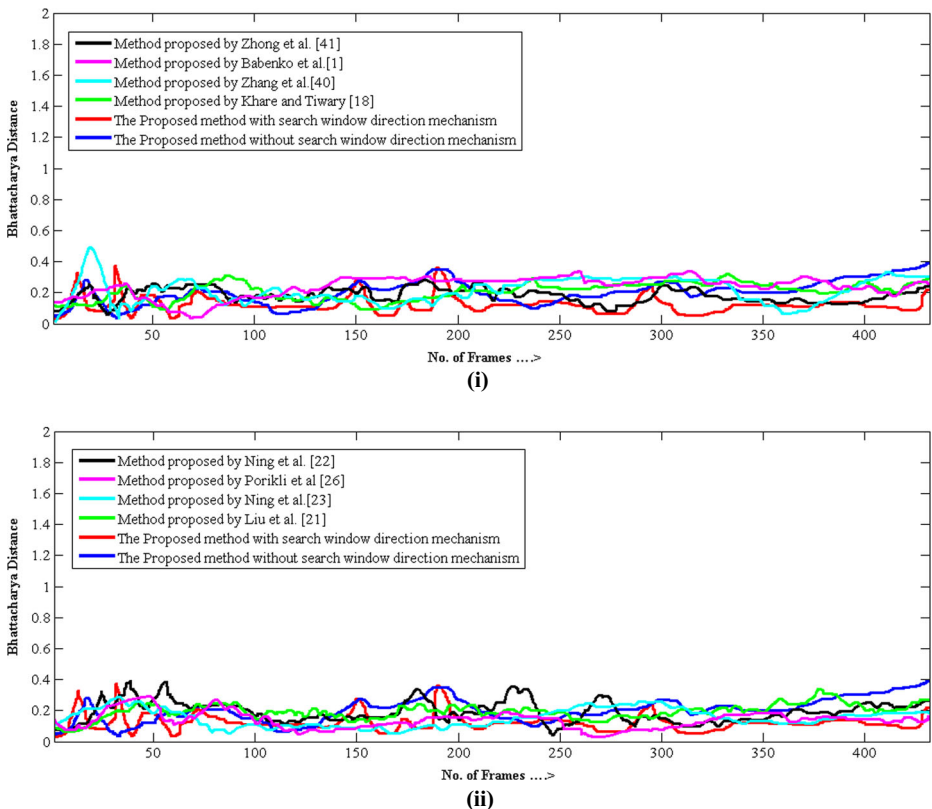

Fig. 15 Plot of Bhattacharya distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for PETS video sequence

After computation of $c(i,j)$, now covariance matrix C [8], is defined as

$$C = \begin{bmatrix} c(1,1) & c(1,2) & . & . & . & c(1,n) \\ c(2,1) & c(2,2) & . & . & . & c(2,n) \\ . & & . & . & & . \\ . & & . & . & & . \\ . & & . & . & . & . \\ c(n,1) & c(n,2) & . & . & . & c(n,n) \end{bmatrix} \tag{7.5}$$
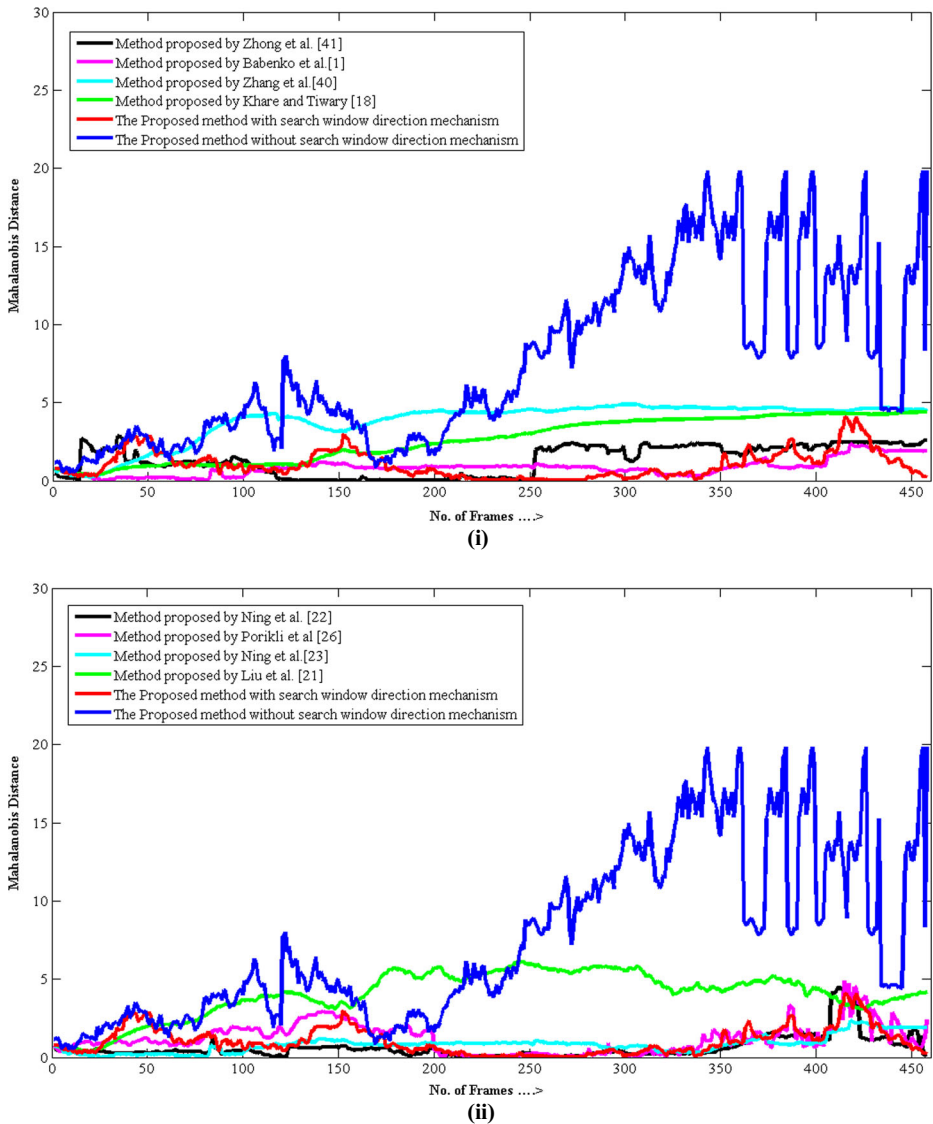

Fig. 16 Plot of Mahalanobis distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for child video sequence

This covariance matrix provides us with a way to measure distance that is invariant to linear transformation of the data.

Figures 16, 17, 18 and 19 shows plots of Mahalanobis distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] in case of Child video sequence, Soccer video sequence, Gallery video sequence and PETS video sequence respectively.

From Figs. 8, 9, 10 and 11 it is clear that the proposed method has the least Euclidean distance between centroid of tracked bounding box and ground truth centroid in comparison to other methods. From Figs. 12, 13, 14 and 15, it can be seen that the proposed method shows least deviation from actual object region in comparison to other state-of-the-art methods. From Figs. 16, 17, 18 and 19 it is clear that the values of dissimilarity measure are small in case of proposed method, i.e. the ground truth centroids and computed centroids of the proposed method are almost same. We can also see that instead of results shown in Figs 8, 12 and 16, all other results (i.e. Figs. 9, 10, 11, 13, 14, 15, 17, 18 and 19) giving plot in some linear manner this is due to fact that in Figs. 8, 12 and 16, child object is moving abruptly in frames whereas in all other video sequences (results are in Figs 9, 10, 11, 13, 14, 15, 17, 18 and 19), object is moving in linear manner and little bit predictable manner.
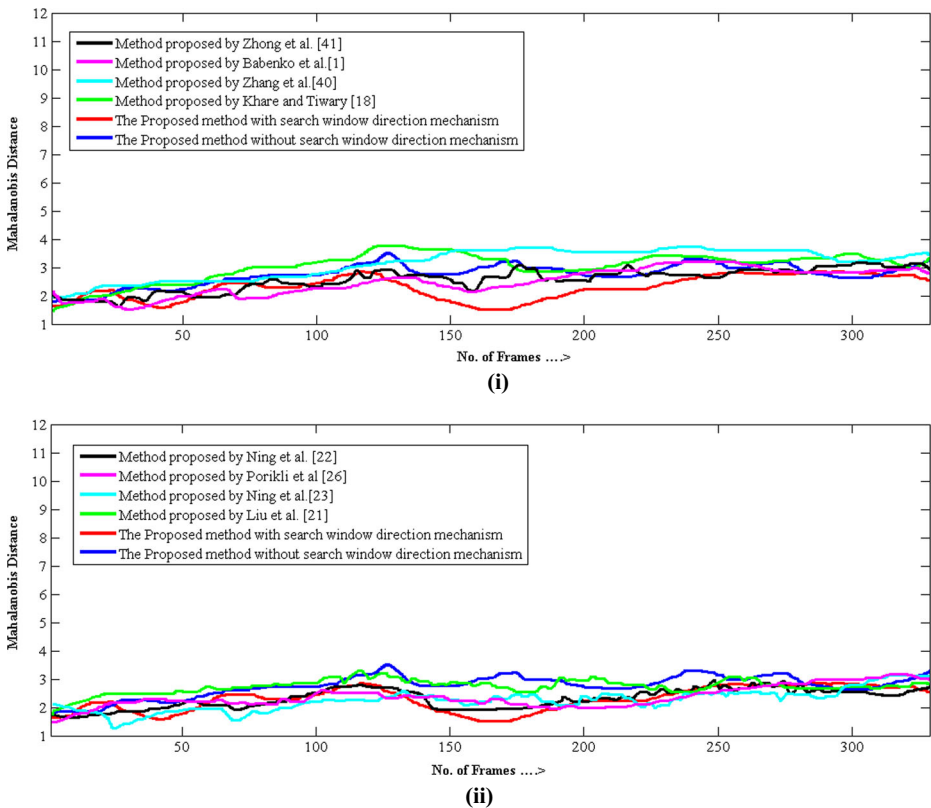

Fig. 17 Plot of Mahalanobis distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Soccer video sequence
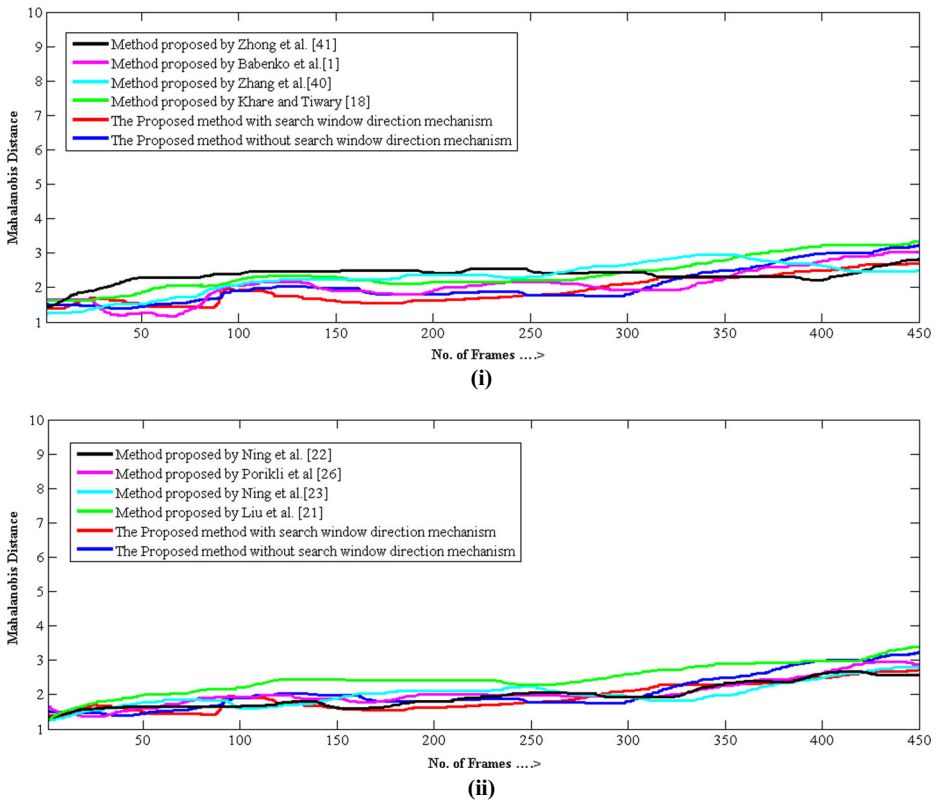
Fig. 18 Plot of Mahalanobis distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for Gallery video sequence

8 Conclusions

In the present work, we have proposed a new method for tracking of object in video using combination of Daubechies complex wavelet transform and Zernike moment as a features of object. Main motivation behind using combination of two feature set for feature based video object tracking is that some of the features are more informative than others in some aspect, therefore chances of correct tracking will increase. Reduced shift sensitivity, approximate rotation invariance and better edge detection properties of Daubechies complex wavelet transform along with translation and rotation invariance properties of Zernike moment makes the proposed method more suitable for tracking of object in video. The proposed algorithm is simple to implement since it does not need any parameter except the Zernike moment of Daubechies complex wavelet coefficients. In the proposed method, the prediction of displacement of object is calculated from Newton's equation of motion that reduces of false tracks as well as the proposed method uses search window direction mechanism, which is very useful in case of slow motion or stable motion. The proposed algorithm allows user to easily and quickly track an object in video.

We have tested the proposed method in two cases (i) the proposed method with search window direction mechanism and (ii) the proposed method without search window direction mechanism. Experimental results indicate that the proposed method performs well as compared to other discussed tracking methods [1,18,21–23,26,40,41] qualitatively and quantitatively. For
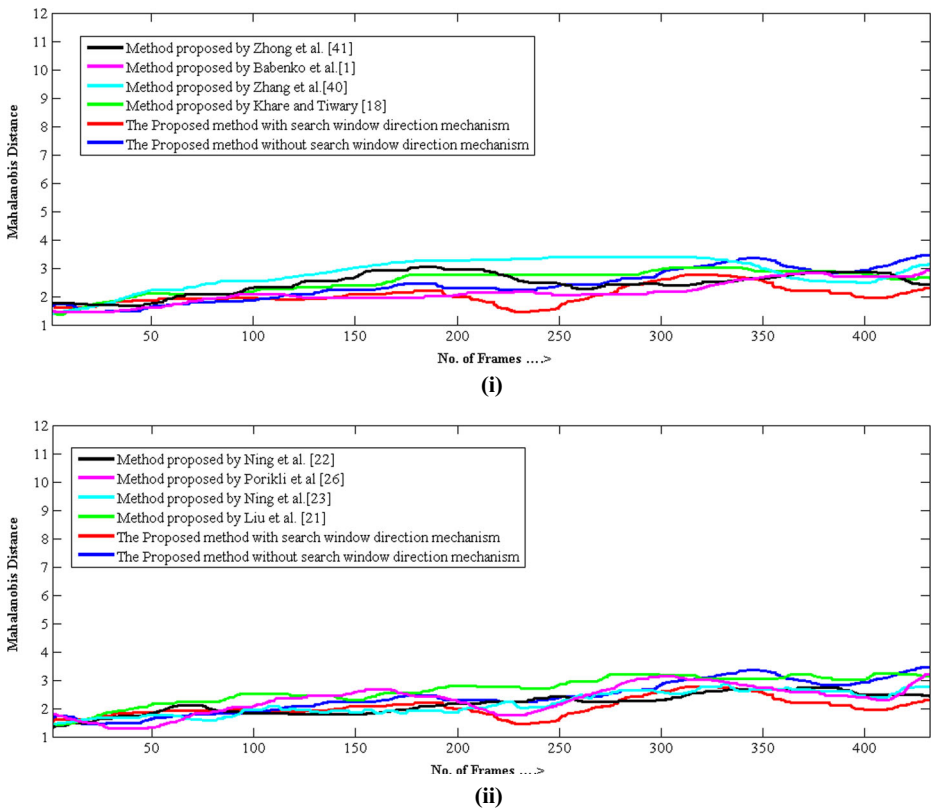
Fig. 19 Plot of Mahalanobis distance for the proposed method and other state-of-the-art methods [1,18,21–23,26,40,41] for PETS video sequence

quantitative measures we have used Euclidean distance, Bhattacharya distance and Mahalanobis distance. Unlike the other methods, the proposed method does not rely upon many properties of object such as size, shape, color etc.

The main advantages of the proposed method is given below-

(i)   The proposed method is able to track the object in the video having cluttered background, changing background, as well as changing lighting conditions.
(ii) The proposed method is able to handle occlusion problem efficiently.
(iii) The proposed method is able to track object efficiently where objects are moving at fast speed and whose direction of movement changes abruptly.

The contribution of the proposed work can be summarized as follows-

(i)   A new method for tracking of object in video, which is based on combination of Daubechies complex wavelet transform and Zernike moment as a feature of object, is developed and presented.
(ii) The proposed algorithm rely not only on the matching of combination of energy of Zernike moments of wavelet coefficients among different frames, but also on the

prediction of displacement of object calculated from Newton's equation of motion that reduces the false tracks.

(iii)   The method has been tested on several video sequences and is found to have better qualitative and quantitative performance as compared to representative state-of-the-art methods.

References

1. Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
2. Bin Y, Xiong PJ (2002) Invariance analysis of improved Zernike moments. J Opt A Pure Appl Opt 4(6):606–614
3. Cheng FH, Chen YL (2006) Real time multiple objects tracking and identification based on discrete wavelet transform. Pattern Recogn 39(6):1126–1139
4. Chong CW, Raveendran P, Mukundan R (2003) Translation invariance of Zernike moments. Pattern Recogn 36(8):1765–1773
5. Clonda D, Lina JM, Goulard B (2004) Complex daubechies wavelets: properties and statistical image modelling. Signal Process 84(1):1–23
6. Comaniciu D, Ramesh V, Meer P (2003) Kernel based object tracking. IEEE Trans Pattern Anal Mach Intell 25(5):564–577
7. Daubechies I (1992) Ten Lecture on Wavelets. SIAM
8. Duda RO, Hart PE, Stork DG (2006) Pattern classification. Wiley, UK,
    Second Edition
9. Elgamel SA, Soraghen J (2011) Enhanced monopulse tracking radar using optimum fractional fourier transform. IET Radar Sonar Navig 5(1):74–82
10. Gurwicz Y, Yehezkel R, Lachover RB (2011) Multiclass object classification for real time video surveillance systems. Pattern Recogn Lett 32(6):805–815
11. Hu W, Tan T (2006) A survey on visual surveillance of object motion and behavior. IEEE Trans Syst Man Cybern 34(3):334–352
12. Hwang SK, Kim WY (2006) A novel approach to the fast computation of Zernike moment. Pattern Recogn 39(11):2065–2076
13. Islam MM, Alam MS (2007) Human motion tracking using mean shift clustering and discrete cosine transform, Proc SPIE conference 6566, Automatic Target Recognition XVII, 656616, 10.1117/12.717921
14. Khansari M, Rabiee HR, Asadi M, Ghanbari M (2008) Object tracking in crowded video scenes based on Undecimated wavelet features and texture analysis. EURASIP Journal on Advances in Signal Processing, Article ID 243534, 10.1155/2008/243534
15. Khare M, Binh NT, Srivastava RK (2014) Human object classification using dual tree complex wavelet transform and Zernike moment. Trans Large Scale Data Knowl Cent Syst XVI, LNCS 8960:87–101
16. Khare M, Patnaik T, Khare A (2010) Dual tree complex wavelet transform based video object tracking. Proc Int Conf ICT 101(2):281–286
17. Khare M, Srivastava RK, Khare A (2015) Moving object segmentation in daubechies complex wavelet transform. SIViP 9(3):635–650
18. Khare A, Tiwary US (2007) Daubechies complex wavelet transform based moving object tracking, Proc IEEE Symposium on Computational Intelligence in Image and Signal Processing, pp. 36–40
19. Khotanzed A, Hong YH (1990) Invariant image recognition by Zernike moment. IEEE Trans Pattern Anal Mach Intell 12(5):489–497
20. Liu TL, Chen HT (2004) Real time tracking using trust region methods. IEEE Trans Pattern Anal Mach Intell 26(3):397–401
21. Liu B, Huang J, Kulikowski YL (2013) Robust visual tracking using local sparse appearance model and K-selection. IEEE Trans Patt Recog Mach Intell 35(12):2968–2981
22. Ning J, Zhang L, Zhang D, Wu C (2009) Robust object tracking using joint color texture histogram. Int J Pattern Recognit Artif Intell 23(7):1245–1263

23. Ning J, Zhang L, Zhang D, Wu C (2012) Robust mean shift tracking with corrected background. IET Comput Vis 6(1):62–69
24. Nummiaro K, Meier EK, Gool IJV (2003) An adaptive color-based particle filter. Image Vis Comput 21(1):99–110
25. Papakostas GA, Boutalis YS, Karras DA, Mertzios BG (2007) A new class of Zernike moments for computer vision applications. Inf Sci 177(13):2802–2819
26. Porikli F, Tuzelq O, Meer P (2006) Covariance tracking using model update based on lie algebra. Proc IEEE Conf Comput Vis Pattern Recognit 1:728–735
27. Prakash O, Khare A (2011) Tracking of non-rigid object in complex wavelet domain. J Signal Inf Proc 2:105–111
28. Rocha L, Velho L, Cezar P, Carvalho P (2002) Image moment based structuring and tracking of objects, Proc 15th Brazilian Symposium on Computer Graphics and Image Processing, pp. 99–105
29. Seferidis VE, Ghanbari M (1994) Adaptive motion estimation based on texture analysis. IEEE Trans Commun 42(2–4):1277–1287
30. Shen C, Kim J, Wang H (2010) Generalized kernel based visual tracking. IEEE Trans Circuits Syst Video Technol 20(1):119–130
31. Soman KP, Ramachandran KI, Resmi NG (2010) Insight into wavelets - from theory to practice. Prentice Hall India, New Delhi, **Third Edition**
32. Sonka M, Hlavac V, Boyle R (2001) Image processing: analysis and machine vision. Thomson Asia Pvt. Ltd, Singapore, Second Edition
33. Srivastava P, Binh NT, Khare A (2014) Content-based image retrieval using moments of local ternary pattern. Mobile Net Appl. doi:10.1007/s11036-014-0526-7
34. Teague M (1980) Image analysis via the general theory of moments. J Opt Soc Am 70(8):920–930
35. Vorobyov M (2011) Shape Classification Using Zernike Moments. Technical Report, iCamp at University of California, Irvine
36. Wang L, Hu W, Tan T (2003) Recent developments in human motion analysis. Pattern Recogn 36(3):585–601
37. Wang X, Wang Z, Hong K (2013) An object tracking approach based on Hu moments and ABC shifts. In: Wong W, Tinghuai M (ed) Emerging technologies for information systems, computing and Management, pp. 453–460
38. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38(4):1–45
39. Yilmaz A, Li X, Shah M (2004) Contour based object tracking with occlusion handling in video acquired using mobile camera. IEEE Trans Pattern Anal Mach Intell 26(11):1531–1536
40. Zhang T, Ghanem B, Liu S, Ahuja N (2013) Robust visual tracking via structured multi-task sparse learning. Int J Comput Vis 101:367–383
41. Zhong W, Lu H, Yang MH (2012) Robust Object Tracking via Sparsity-based Collaborative Model, Proc IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838–1845
42. Zivkovic Z, Cemgil AT, Krose B (2009) Approximate Bayesian methods for kernel based object tracking. Comput Vis Image Underst 113(6):743–749

Manish Khare received his D.Phil. degree in Computer Science from University of Allahabad, Allahabad, India in 2015. He received M. Tech. degree in Computer Science and Engineering from School of IT, Centre for Development of Advanced Computing, Noida, India, in year 2009, and M. Sc. in Computer Science from University of Allahabad in year 2006. He is a student member of IEEE. His Research interests include Image and Video Processing, Computer Vision, and applications of wavelet Transforms. He has published more than 25 papers in International Journals, book chapters, International and National Conferences.

Rajneesh Kumar Srivastava received his D.Phil. degree from University of Allahabad, Allahabad, India, in 2008. He received his M.Tech. degree from IT-BHU, Varanasi. Presently he is working as an Associate Professor in Department of Electronics and Communication, University of Allahabad, India. Earlier he worked as research staff at CRL, BEL, Ghaziabad, India. His research interests include Photo Electrets, Mobile Communication, and Image Processing. He has served as Principal Investigator of research project sponsored by UGC. He has supervised a number of Ph.D. students and has published one book, one book chapter and more than 75 research papers.

Ashish Khare received M.Sc. (Computer Science) and D. Phil. degree both from Department of Electronics and Communication, University of Allahabad, Allahabad, India, in 1999 and 2007, respectively. He did his post-doctoral research from Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2007–2008. Presently, he is working as an Associate Professor in Department of Electronics and Communication, University of Allahabad. His research interests include applications of wavelet transforms, computer vision, cyber security, and human behavior understanding. He has worked as Principal Investigator of research projects funded by UGC and DST. He has supervised a number of Ph.D. students and has published one book, one book chapter and around 100 research papers.