ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2020.0320

TRACKING OF MULTIPLE HUMAN OBJECTS USING COMBINATION OF DAUBECHIES COMPLEX WAVELET TRANSFORM AND ZERNIKE MOMENT

Manish Khare[1] and Ashish Khare[2]
[1]*Department of Information Communication and Technology, Dhirubhai Ambani Institute of Information and Communication Technology, India*
[2]*Department of Electronics and Communication, University of Allahabad, India*

Abstract
The goal of multi object tracking is to find location of the target objects in number of consecutive frames of a video. Tracking of multiple human objects in a scene is one of the challenging problems in computer vision applications due to illumination variation, object occlusion, abrupt motion etc. This paper introduces a new method for multiple human 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 objects. The motivation behind using combination of these two as a feature of object, because shift invariance and better edge representation properties make Daubechies complex wavelet transform suitable for locating object in consecutive frames whereas translation invariant property of Zernike moment is also helpful for correct object identification in consecutive frames. The proposed method is capable to handle full occlusion, partial occlusion, split and object re-enter problems. The experimental results validate the effectiveness and robustness of the proposed method.

Keywords:
Multiple Object Tracking, Daubechies Complex Wavelet Transform, Zernike Moment, Shift-Invariance, Translation Invariance, Occlusion

1. INTRODUCTION

Multi object tracking in a video sequence is a crucial problem in computer vision applications, like surveillance system and security, target detection and interpretation, traffic monitoring, sport video analysis [1]-[3] etc. The main attributes of a good tracking algorithms should be as below [1]:

- *Occlusion*: tracking algorithm should have ability to deal with presence of occlusion.
- *Multiple Object*: tracking algorithm should have ability to deal with more than one object in a scene.
- *Robustness*: tracking algorithm should have ability to deal with noisy data.
- *Non-Rigid Object*: tracking algorithm should have ability to deal with non-rigid object.

Tracking task can be performed in two ways: single object tracking and multiple object tracking. Multiple object tracking is much more complicated task than single object tracking, due to varying lighting condition, shape and size of object may vary from frame to frame, loss of information caused by projection of the 3-D world on 2-D image, partial or full occlusion, object merging and object splitting. Occlusion occurs due to one another object is hiding object or object is being hide by some background component. Split occurs due to splitting of merged object in a scene [4].

Many literatures are available on single object tracking [5]-[8], but only a few and far works have been reported on multiple object tracking [4]. Moving object segmentation is primary and mandatory step for multi-object tracking. Once the object is separated by its background, it can be characterized by some features, which are used to locate objects for tracking. Inaccurate moving object segmentation may result in failure to track the object or false object being tracked. Three different types of moving object segmentation methods are popular in multi object tracking, and they are background subtraction method, temporal differencing method and optical flow-based method. Algorithmic complexity and execution time are major drawbacks of optical flow and temporal difference based methods. This is the reason why background subtraction based methods are broadly used in multi-object tracking. In recent years, large number of solutions have been proposed for background subtraction [9]. In recent researches, wavelets are used for moving object segmentation [10]-[13], as wavelet domain methods are more beneficial than spatial domain methods because they provide more object related information. Since shadow creates problem in accuracy of moving object segmentation, so if shadow of moving object gets removed during segmentation then accuracy of segmentation increases. Guan [14] and Khare et al. [15]-[17] proposed a method for foreground segmentation with shadow removal from moving object using HSV color space in wavelet domain. Guan [14] used multiscale dyadic wavelet transform whereas Khare et al. [15]-[17] used complex wavelet transform.

Multi object tracking algorithm should be able to establish unique correspondences between objects in each frame of a video [4]. Mittal and Davis [18] proposed Kalman filtering based approach for tracking of multiple objects, but Kalman filtering approach creates problem and it will become difficult to correct, if the number of object increases. Smith et al. [19] proposed an algorithm for tracking of multiple objects by particle filtering approach which reduces the problem present in Kalman filtering. Nillius et al. [20] proposed a multi object tracking method by using Bayesian network inference, this method is suitable for sport video analysis to track multiple objects. Han et al. [21] proposed a probabilistic framework for multiple object tracking which is based on Hidden Markov Model. These methods [19-21] are capable to handle partial short duration occlusions.

Rad and Jamzad [22] proposed three criterions to predict and detect occlusion of vehicle in a highway. These criterions are based on examining the trajectory of each vehicle and size of foreground region. In Rad and Jamzad [22] approach, occlusion can be predicted, if the centroids of foreground region are too close to each other. Amer [23] proposed a two-stage non-linear feature for tracking of multiple objects in presence of heavy occlusion, but this method fails to track the objects when object suddenly disappears, or object re-enter in scene or object changes its direction. Cheng and Chen [24] proposed a method for real time multiple object tracking by using discrete wavelet transform, but it has been proven that real valued wavelet transform is less suitable for computer vision applications. Jalal and Singh [4] proposed complex wavelet transform based multi-object tracking, which exploits the properties of Daubechies complex wavelet transform, Jalal and Singh [4] combined wavelet coefficient with appearance based occlusion handling, which was proposed by Senior et al. [25], and then claimed that their framework is suitable for tracking of multiple objects with occlusion handling.

A new trend is to combine two or more features for tracking. Combination of two or more features in one is very useful and gives good results in case of single object tracking, so in the present work, we have used combination of two features in case of multiple object tracking. In the present work we have proposed a multiple human object tracking algorithm based on combination of Daubechies complex wavelet transform and Zernike moment. The Daubechies complex wavelet transform having advantages of shift invariance and better edge representation. Zernike moment also have many desirable properties such as translation invariance, robustness to noise, etc. In the proposed work our desirable task is to track multiple human objects. The experimental results validate that the proposed method for multi object tracking performs well for both indoor and outdoor video sequences, and also solve the problems of partial/full occlusion, split and merge object, as well as object re-enter in scene.

Rest of the paper is organized as follows: section 2 describes basics of used feature set for multi-object tracking (Daubechies complex wavelet transform and Zernike moment). Section 3 describes properties of Daubechies complex wavelet transform and Zernike moment. Section 4 describes the proposed framework in detail. Experimental results and discussions are given in section 5. Finally conclusions of the work is given in section 6.

2. FEATURE USED IN MULTI-OBJECT TRACKING

Selection of right features play a critical role in multi 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 multiple objects can be easily distinguishable in feature space. Use of single feature may not be much successful for solving multi object tracking problem, because single feature is not rich enough for representation of wide variety of objects. When we use combination of two or more features, some of the features will be more informative than others in certain aspects. Therefore combining multiple type of features can enhance the robustness and tracking accuracy. In the present work, for multi object tracking, we have taken combination of two different feature sets – Daubechies complex wavelet transform and complex Zernike moment. A brief description of these features are given in subsections 2.1 and 2.2 respectively.

2.1 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 multi-resolution theory is the scaling function

$$\phi(u) = 2\sum_i a_i \phi(2u - i) \tag{11}$$

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 basis $\{\psi_{j,k}(t)\}$ in one-dimension is defined using scaling function $\phi(u)$ as defined in Eq.(2) and multiresolution analysis of $L_2(\Re)$ [26][27].

During the formulation of general solution if we relax the Daubechies condition for a_i to be real [28], it leads to complex valued scaling function and further Daubechies complex wavelet transform.

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

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

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:

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

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

2.2 ZERNIKE MOMENT

Zernike moment was firstly introduced by Teague [29] to overcome different shortcomings of information redundancy in Geometric moments [29]. 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 [30]. Due to these important characteristics, Zernike moments have been used 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 i.e. $x^2 + y^2 \leq 1$ [31]. These polynomials are of the form,

$$V_{mn}(x, y) = V_{mn}(r, \theta) = R_{mn}(r) \cdot \exp(jn\theta) \tag{4}$$

where m is positive integer and n is positive and negative integer subject to constraints m-$|n|$ is 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:

$$R_{mn}(r)=\sum_{s=0}^{\left(\frac{m-|n|}{2}\right)}\frac{(-1)^s(m-s)!\,r^{m-2s}}{s!\left(\frac{m+|n|}{2}-s\right)!\left(\frac{m-|n|}{2}-s\right)!} \tag{5}$$

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

The above mentioned polynomial in Eq.(5) is orthogonal and satisfies the othogonality principle

$$\iint_{x^2+y^2\le 1} V_{mn}(x,y).V_{pq}(x,y)dxdy=\frac{\pi}{n+1}\delta_{mp}\delta_{nq} \tag{6}$$

where $\delta_{ab}=1$ for $(a=b)$ and $\delta_{ab}=0$ otherwise, is the Kronecker symbol.

Zernike moments are the projection of image function $I(x,y)$ onto these orthogonal basis functions. The orthogonality condition simplifies the representation of the original image because generated moments are independent [32].

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

$$Z_{mn}=\frac{m+1}{\pi}\iint_{x^2+y^2\le 1} I(x,y)\left[V_{mn}(r,\theta)\right]dxdy \tag{7}$$

In case of digital image, the integrals are replaced by summation [33], as given below

$$Z_{mn}=\frac{m+1}{\pi}\sum_x\sum_y I(x,y)V_{mn}(r,\theta),\quad x^2+y^2\le 1 \tag{8}$$

Zernike moments are the coefficients of the image expansion into orthogonal Zernike polynomials as can be seen in the following reconstruction equation as in Eq.(9):

$$f(x,y)=\sum_{m=0}^{m_{\max}}\sum_n Z_{mn}V_{mn}(r,\theta) \tag{9}$$

3. 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.

3.1 PROPERTIES OF DAUBECHIES COMPLEX WAVELET TRANSFORM

Daubechies complex wavelet transform have several properties, in which better edge representation, and reduced shift sensitivity properties directly influence the multi object tracking algorithm. Brief description of these properties are given in subsections 3.1.1 and 3.1.2.

3.1.1 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

$$\alpha(w)=-\frac{\hat{v}(w)}{\hat{l}(w)} \tag{10}$$

Clonda et al. [28] 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, Eq.(10) indicates $v(t)\approx\alpha\Delta^2 l(t)$. Here Δ^2 represents second order derivative. This gives multi-scale projection as

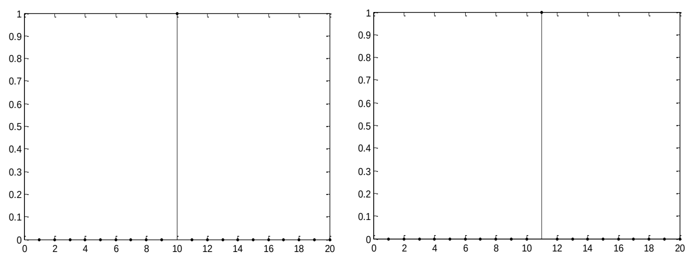
$$\left\langle f(t),x_{j,k}(t)\right\rangle=\left\langle f(t),l_{j,k}(t)\right\rangle+i\left\langle f(t),v_{j,k}(t)\right\rangle$$
$$\approx\left\langle f(t),l_{j,k}(t)\right\rangle+i\alpha\left\langle \Delta^2 f(t),l_{j,k}(t)\right\rangle \tag{11}$$

From Eq.(11), 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 methods.

3.1.2 Reduced Shift Sensitivity:

A transform is said to be shift-sensitive if shift in input-signal causes an unpredictable change in transform coefficients. Real valued wavelet transforms 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 loss at multilevel due to its shift-invariance property.

The Fig.1 shows the reduced shift-sensitivity of Daubechies complex wavelet transform. The Fig.1(a) is an input signal, whereas Fig.1(b) is the shifted form of the input signal by one sample. The Fig.1(c) and Fig.1(d) shows high-pass wavelet coefficients of the original and the shifted signal using real valued discrete wavelet transform. The Fig.1(e) and Fig.1(f) shows high-pass wavelet coefficients of the original and the shifted signal using Daubechies complex wavelet transform. From Fig.1, it is clear that, nature of magnitude of complex wavelet coefficients remain approximately same by shifting the input signal. Thus we can say that discrete wavelet transform is highly shift sensitive, whereas the Daubechies complex wavelet transform is approximate shift-invariant.



(a) (b)

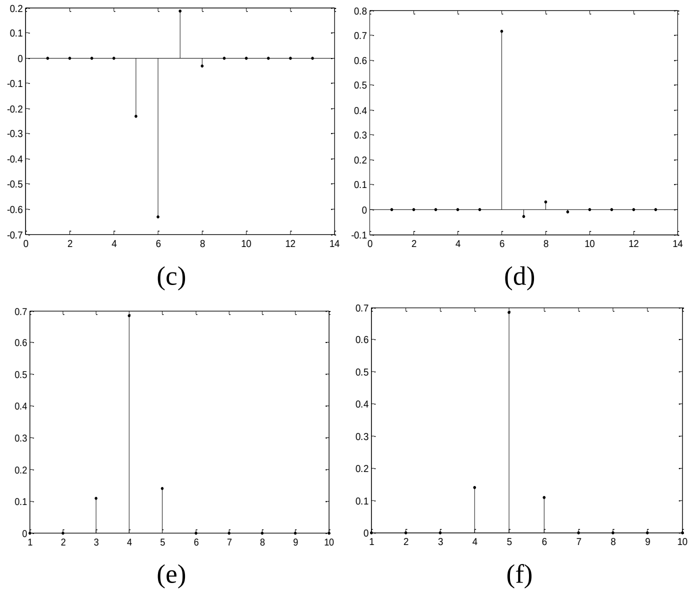


Fig.1. (a) Original Signal, (b) signal shifted by one sample, (c)-(d) high-pass wavelet coefficient of original signal and shifted signal using real db4 wavelet, (e)-(f) Magnitude of complex wavelet coefficients of original signal and shifted signal using SDW6 wavelet

3.2 PROPERTIES OF ZERNIKE MOMENT

Zernike moment hold following important properties [34]:

- Zernike moments are rotation, translation and scale invariant [35].
- Zernike moments are robust to noise and minor variations in shape.
- Since the basis of Zernike moment is orthogonal, therefore they have minimum information redundancy.
- Zernike moment can characterize the global shape of pattern. Lower order moments represent the global shape pattern and higher order moment represents the detail.
- An image can be better described by a small set of its Zernike moments than any other types of moment.

3.2.1 Rotation Invariance Property:

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. The rotation of an image is expressed in polar coordinates by simple change of angle [36].

Thus, relationship between the rotated image and the original image is defined as

$$f^{/}(r,\theta) = f(r,\theta-\alpha) \tag{12}$$

Since, Zernike polynomials are defined only on the unit circle, it is easy to compute Zernike moment in polar coordinate.

$$Z_{mn} = \frac{m+1}{\pi}\int_0^{2\pi}\int_0^1 f(r,\theta)\cdot R_{mn}(r)\cdot\exp(-jn\theta)\,rdrd\theta \tag{13}$$

If, we apply the rotated image into Eq.(13), we get

$$Z_{mn}^{/} = \frac{m+1}{\pi}\int_0^{2\pi}\int_0^1 f(r,\theta-\alpha)\cdot R_{mn}(r)\cdot\exp(-jn\theta)\,rdrd\theta \tag{14}$$

Changing of variable $\theta_1 = \theta-\alpha$, Eq.(14) is again expressed as

$$Z_{mn}^{/} = \frac{m+1}{\pi}\int_0^{2\pi}\int_0^1 f(r,\theta_1)\cdot R_{mn}(r)\cdot\exp\left(-jn(\theta_1+\alpha)\right)rdrd\theta$$

$$= \left[\frac{m+1}{\pi}\int_0^{2\pi}\int_0^1 f(r,\theta_1)\cdot R_{mn}(r)\cdot\exp(-jn\theta_1)\,rdrd\theta\right]\cdot\exp\left(-jn(\alpha)\right)$$

$$A_{mn}^{/} = A_{mn}\exp(-jn(\alpha)) \tag{15}$$

After taking magnitude of the rotated image, we get

$$\left|A_{mn}^{/}\right| = \left|A_{mn}\exp(-jn(\alpha))\right| = \left|A_{mn}(\cos(n\alpha) - j\sin(n\alpha))\right| = \left|A_{mn}\right|$$

So,

$$\left|A_{mn}^{/}\right| = \left|A_{mn}\right| \tag{16}$$

Hence, we find that magnitudes of Zernike moments possess rotation invariance.

Rotation invariance property help in development of accurate object tracking methods. In a video, object may be present in rotated form, hence value of Zernike moments of original object and rotated object remains same. The motivation behind using combination of these two (Daubechies complex wavelet transform and Zernike moment) as a feature of object is that combining multiple types of features can enhance the tracking accuracy, as Daubechies complex wavelet transform is approximate shift invariance and Zernike moment is rotation invariant, therefore combination of these two is expected to accurately track the shifted and rotated object in a video. 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 object 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.

4. THE PROPOSED METHOD

In this section, we have described the proposed method in detail. A video contains a sequence of consecutive frames. Each frame can be considered as an image. One or more objects may be present in each frame. In our proposed framework we have taken object as human object. If the algorithm can track different moving objects between two consecutive frames then it can track multiple human objects in video sequence.

The proposed method consists of following three steps-

- Algorithm for segmentation with shadow removal.
- Algorithm for classification of human object.
- Algorithm for multi-object tracking.

The first step of the proposed method is segmentation of moving objects along with shadow removal. Here we segment moving object in each frame, and if some shadow is associated with object, then we remove it. The second step of the proposed method is classification of human object, in this step we check whether extracted object is human object or not. The third step is the multi-object tracking. This step uses a correspondence process, which is associated with each foreground object with one of the existing tracked objects. Details of all these three steps are described below.

4.1 MOVING OBJECT EXTRACTION AND SHADOW REMOVAL

In moving object segmentation, we compare current frame and reference frame. We have used HSV color model, for detection of pixels, which have been changed due to movement of moving object and its shadow, because this color model corresponds closely with human perception of color and it easily separates chromaticity and luminosity. We have used value and saturation component of each frame for processing in this step. For segmentation of object with shadow removal, we have used our earlier developed method [16] [17]. The segmentation approach, with removal of shadow, as Khare et al. [16] [17], consists of following steps

1. Represent the reference frame and current frame in HSV color model.
2. Take absolute of reference frame and current frame with respect to Hue, Saturation and Value components respectively.
3. Perform wavelet decomposition of value component and saturation component of difference image (ΔV and ΔS) using Daubechies complex wavelet transform. The wavelet coefficients are $DW_{\Delta V}$ and $DW_{\Delta S}$.
4. Compute standard deviation of wavelet coefficients of ΔV and ΔS say $(\sigma)_{DW_{\Delta V}}$ and $(\sigma)_{DW_{\Delta S}}$.
5. Detect segmented object (foreground object) without shadow using following condition

$$SR = \begin{cases} 1, & if\ DW_{\Delta V} \geq (\sigma)_{DW_{\Delta V}} \wedge DW_{\Delta S} \geq (\sigma)_{DW_{\Delta S}} \\ 0, & otherwise \end{cases} \quad (17)$$

6. Some shadow points are always misclassified in detected object. Therefore for improvement, some morphological operation is needed for better shape structure. For better shape structure we have used binary closing morphological operation.

4.2 OBJECT CLASSIFICATION

After segmentation we detect foreground pixels and form isolated regions of connected pixels, which are known as blobs. Next step of the proposed framework is to check whether extracted object is human object or not. For this task, we apply human object classification approach [37] for each extracted object in every frame.

For human object classification, first we mapped blob of object in segmented frame with object in original frame, and then calculate wavelet coefficient of object. After calculation of wavelet coefficient, we applied Zernike moment of 4th order. Values obtained after Zernike moment computation over wavelet coefficient (Daubechies complex wavelet transform coefficients) have been used as feature set for classification purpose. This feature value of object is passed from SVM classifier, which we have trained by standard INRIA person dataset [38]. After classifier computation, SVM classifier classify extracted object as human object and non-human object. This classification approach repeats for all extracted objects in each frame. Here in object classification phase, we check whether extracted object is human object or not. If extracted object is human object, then we track the object according to multi objet tracking mechanism described in following subsection, and if extracted object is not human object, then we skip to track the object in particular frame.

4.3 MULTI-OBJECT TRACKING

Multi object tracking algorithm establishes correspondence between objects in each frame. Multi object tracking can be done in two ways (i) detection based tracking, and (ii) prediction based tracking. Prediction based tracking is much difficult approach, because prediction of position of two or more object simultaneously is very difficult. In our proposed method, we have used detection based tracking approach. In detection based tracking approach we detect location of object in each frame by object extraction. When objects are extracted then we establish correspondence between extracted foreground object and degrades present is original frame, and then according to established association rules we further track the objects.

4.3.1 Object Correspondence:

Object correspondence is used for association of extracted object (foreground object or blobs) with the objects that are already being tracked. Each object is mapped to corresponding connected segmented blobs. Correspondence is used for correct object tracking by assigning unique ID to each object. For this we have used a matrix F_C, which shows association between foreground object extracted in current frame and the object successfully tracked in the previous frame. In the matrix F_C the rows correspond to existing tracks in previous frame and columns to foreground segmented objects (blobs) in the current frame.

Let T_i^{t-1} represent a i^{th} tracked object in $(t\text{-}1)^{th}$ frame and S_j^t represent the j^{th} segmented blob in t^{th} frame respectively, where $i=1,2,\ldots,K$ and $j=1,2,\ldots,L$, K represents the number of object that already being tracked in the previous frame and L represent the number of foreground segmented blobs in the current frame. If $\left(C_x^{T_i}, C_y^{T_i}\right)$ and $\left(C_x^{S_j}, C_y^{S_j}\right)$ are centroids of tracked object and segmented blobs respectively, then the distance between segmented blob S_j^t and tracked object T_i^{t-1} can be calculated as

$$d_x = | C_x^{T_i^{t-1}} - C_x^{S_j^t} |, \quad d_y = | C_y^{T_i^{t-1}} - C_y^{S_j^t} | \quad (18)$$

In the proposed method, we have created a data structure for object (DS_{obj}), to keep track of information about the tracked object. The information stored in data structure are: identity of object (*ID*), centroid of object (*C*), area of object (*Ar*), Height of object (*H*), Width of object (*W*), and status of object (*S*). The status of object represent current status of any object such as active (*A*), merge (*M*), object disappear (*E*), and object reappear (*P*).

The size of matrix F_C is $m\times n$, and its value can be defined as:

$$F_c[i,j] = \begin{cases} 1 & \text{if } d_x < \dfrac{W_{T_i^{t-1}} + W_{T_j^t}}{2} \text{ and } d_y < \dfrac{H_{T_i^{t-1}} + H_{T_j^t}}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (19)$$

where *W* and *H* represent width and height of object (*T*) and segmented blob (*S*).

From Eq.(19), we can see that F_C contain only binary value where entry '*1*' shows that there is correspondence between *T* and *S* and entry '*0*' shows that there is no correspondence between *T* and *S*. The analysis on F_C was given by Comaniciu et al. [39], with following association rule:

- ***Active Track*:** Single segmentation blob S in associated with single object *T* in the previous frame, if the *S* is isolated and not occluded. In such condition the corresponding column and row in F_C have only one non-zero element. The corresponding information in DS_{obj} is updated when *S* is declared as active track, and then create a bounding box for corresponding object of *S* in current frame and provide unique identity number to object.
- ***Appearing or Reappearing*:** If a column in F_C has all zero elements then corresponding *S* cannot be explained by any existing object. Thus *S* has to be a new region which is either by the entry of new object or re-entry of one of the existing object. If it is a case of existing object then feature of such segmented blob *S* is matched against the object having '*P*' status in DS_{obj}. If a match is found, then provide same identity number to that object as last identity number when object was disappeared with bounding box. If no match is found, then the segmented blob is treated as a new object. If segmented blob is detected as a new object then its details are added in DS_{obj}, and create a bounding box for object with new identity number to that object.
- ***Disappear*:** If a row in F_C has all zero element then it means corresponding object *T* is not supported by any of the segmented blob. Hence object *T* has disappeared from scene. In case of disappearing of the object *T*, the status of object *T* is updated as '*E*' in data structure DS_{obj}.
- ***Merging*:** If a column in F_C has more than one non-zero entries. It means that multiple objects form a single segmented blob.
- ***Splitting*:** If a row in F_C has more than one non-zero entries. It means that merged tracked object *T* is splitted into its different corresponding objects.

4.3.2 Detection and Correction of Occlusion:

Merging of two objects is very common problem in multi object tracking, because in multiple object tracking different objects cross each other. When two or more objects are close to each other in a scene, the bounding box of objects overlap in the frame and form one single bounding box. Hence after merging of two objects into a single segmented blob have area significantly larger than corresponding objects. Let two different objects T_A and T_B are occluded in current frame *t* and gives one single segmented blob S_M, then in the proposed tracking approach bounding box of merged object is detected as new object and tracked as in one of the two conditions (i). If object *A* is in front and object *B* is occluding behind of object *A*, then merged object is tracked as identity *A* and object *B* is declared as disappear, and (ii). If object *B* is in front and object *A* is occluding behind of object *B*, then merged object is tracked as identity *B* and object *A* is declared as disappear. This merged object will be tracked in subsequent frames until it splits.

4.3.3 Detection and Correction of Splitting:

Merged object can be split into several blobs during process of moving object segmentation process. In this case merged segmented blob S_M, as in previous section 4.3.2 is split and produces corresponding object T_A and T_B, and their identity is re-established as feature of objects stored in data structure. These split objects are then tracked as existing object.

5. EXPERIMENT AND RESULTS

The proposed method for multi object tracking as described in section 4, is implemented in MATLAB environment and has been applied on number of representative standard video sequences as well as video sequences captured by authors of this paper. Results are being presented here for five different video sequences viz. 'hall monitor' video sequence, 'JK_Campus' video sequence, 'Field' video sequence, 'Terrace' video sequence, and 'Road' video sequence. Hall Monitor video sequence is a standard representative video sequence and rest four video sequences are captured by authors of this paper.

Each video sequence have its own properties such as 'Hall monitor' video sequence is a type of indoor video sequence, in which no occlusion occurs between objects. 'JK_Campus' video sequence is a type of outdoor video sequence, in which no occlusion occurs between object but here some objects re-enter in scene many times. 'Field' video sequence is a type of outdoor video sequence, in which complete occlusion exists between two objects. 'Terrace' video sequence is a type of outdoor video sequence, in which occlusions as well as re-entering of objects in scene problem exists. 'Road' video sequence is type of outdoor video sequence with poor contrast, in which occlusion as well as re-entering of objects in scene problem exists. The proposed method does not depend on any particular model and no manual intervention is needed in whole process. In each experiment, the ID of the object is labelled at the top of the bounding box.

5.1 EXPERIMENT 1

In experiment 1, experimental results for hall monitor video sequence are shown. Hall monitor video sequence suffered from noise caused by variation in the illumination changes. In this video sequence color distribution of background is much similar to the trouser of the first object, shadow of object also exists in this video sequence. The Fig.2 shows multi-object tracking results for different representative frames of video sequence. In each frames one can see that, the proposed method tracks different objects efficiently when there is no-occlusion between objects.

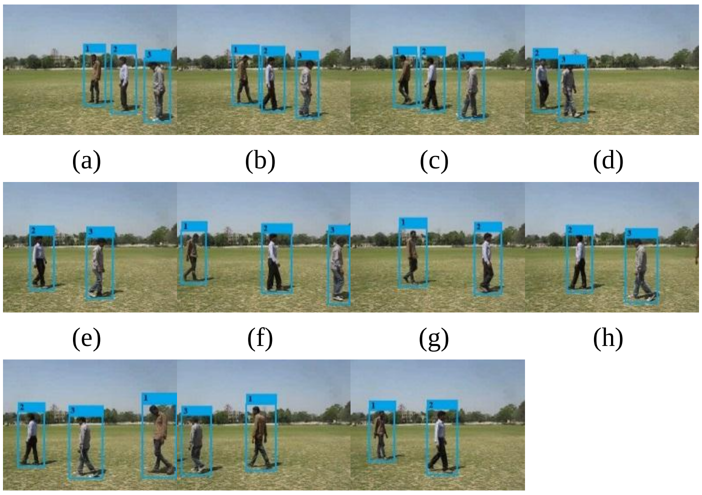
(a) (b) (c)

(d) (e)

Fig.2. Multi object tracking results for ‘hall monitor’ video sequence using the proposed method for (a) frame 50, (b) frame 100, (c) frame 150, (d) frame 200, and (e) frame 250

5.2 EXPERIMENT 2

In experiment 2, experimental results for ‘JK_Campus’ video sequence are shown. In this experiment we have tracked three different human objects. The Fig.3 shows multi-object tracking result for different representative frames of JK_Campus’ video sequence. The Fig.3(a)-Fig.3(c) shows that three objects are being tracked simultaneously. In Fig.3(d)-Fig.3(h), object #1 disappears from scene. In Fig.3(f)-Fig.3(i), object #1, again re-enter in scene with same identity number. In Fig.3(j), object #2, disappear and again re-appear in frame shown in Fig.3(k). From experiments with ‘JK_Campus’ video sequence, we can see that the proposed method works efficiently in case of object re-enter in scene.



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

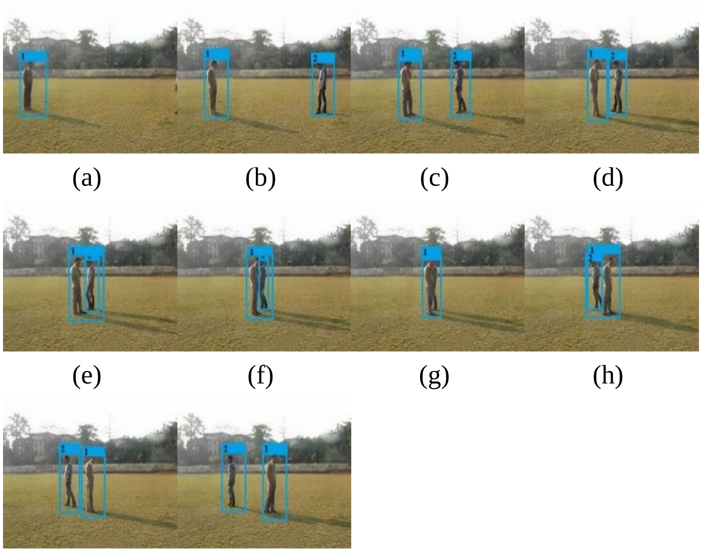
(e) (f) (g) (h)

(i) (j) (k)

Fig.3. Multi object tracking results for ‘Jk_Campus’ video sequence using the proposed method for (a) frame 50, (b) frame 100, (c) frame 150, (d) frame 200, (e) frame 250, (f) frame 300, (g) frame 350, (h) frame 400, (i) frame 450, (j) frame 500, and (k) frame 550

5.3 EXPERIMENT 3

In experiment 3, experimental results for ‘Field’ video sequence are shown. In this experiment, we have shown results which demonstrate that the proposed method have capability of occlusion handling. The Fig.4 shows multi object tracking result for different representative frames of ‘field’ video sequence. The Fig.4(a) shows only one object in frame with labelled object #1. In Fig.4(b), one new object enter in scene with labelled object #2 and in Fig.4(c), both objects #1, and #2, are being tracked efficiently. In Fig.4(d)-Fig.4(f) both objects #1, and #2, are very close to each other, but tracker tracks both object efficiently. In Fig.4(g), object #1, covers object #2 completely. Due to this reason bounding box of object #2, is lost and bounding box of object #1 is shown in Fig.4(g). In Fig.4(h), when object #2, splits from object #1 after merging, tracker again tracks object #2, with its original labelling. In Fig.4(i) and Fig.4(j), tracker tracks both object efficiently as in earlier Fig.4(b). So from this experiments we can see that the proposed method work efficiently in case of object occlusion. The proposed method also tracks object accurately at the time of splitting of object after merging situation.



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

(e) (f) (g) (h)

(i) (j)

Fig.4. Multi object tracking results for ‘field’ video sequence using the proposed method for (a) frame 50, (b) frame 100, (c) frame 150, (d) frame 200, (e) frame 250, (f) frame 300 (g) frame 350, (h) frame 400, (i) frame 450, (j) frame 500

5.4 EXPERIMENT 4

In experiment 4, experimental results for ‘Terrace’ video sequence are shown. In this experiments we have shown results for presence in partial occlusion as well as when object re-enter in scene in a video. The Fig.5 shows multi object tracking result for different representative frames of Terrace video sequence. The Fig.5(a), shows only one object in frame with labelled object #1. In Fig.5(b), one new object enters in frame with labelled object #2, and in Fig.5(c), and Fig.5(d), both object #1, and #2, are being tracked efficiently. In Fig.5(e), object #1, and object #2, partially occludes with each other, but the proposed tracking algorithm tracks both object efficiently. In Fig.5(f), one more new object enters in frame with labelled object #3, and in Fig.5(g) and Fig.5(h), both objects #1, and #3 are being tracked efficiently. In Fig.5(i), object #2, again re-enter in scene with same identity number. From this experiments, we can conclude that the proposed method efficiently handle the case of combination of object occlusion problem and object re-entering in scene problem.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

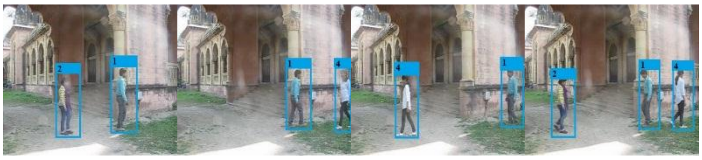
Fig.5. Multi object tracking results for ‘terrace’ video sequence using the proposed method for (a) frame 100, (b) frame 200, (c) frame 300, (d) frame 400, (e) frame 500, (f) frame 600, (g) frame 700, (h) frame 800, and (i) frame 900

5.5 EXPERIMENT 5

In experiment 5, experimental results for ‘Road’ video sequence are being presented and analyzed. ‘Road’ video sequence is a poor contrast video, and in this video we tried to prove that the proposed method works well in case of poor contrast video sequence as well. The Fig.6 shows multi-object tracking results for different representative frames of ‘Road’ video sequence. The Fig.6(a), shows only one object in frame labelled object #1. In Fig.6(b), one new object enters in scene with labelled object #2, and in Fig.6(c) and Fig.6(e), both object #1, and #2, are being tracked efficiently whereas in Fig.6(d), object #1 is partially occluded with object #2, and the proposed tracking algorithm tracks both object efficiently. In Fig.6(f), one new object enters in frame with labelled object #3, and the proposed algorithm track these two objects #1, and #3 efficiently in Fig.6(g). In Fig.6(h), object #2, again re-enters in scene with same object identity number and is being tracks efficiently with object #1. In Fig.6(i), one new object enters in frame with labelled object #4, and being tracked efficiently with object #1 in Fig.6(k). In Fig.6(l), object #2 again re-enter in scene and then object #1, #2, and #4 are being tracks simultaneously. From this experiment, we can conclude that the proposed method works efficiently in case of poor contrast video sequence.

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

(e) (f) (g) (h)



(i) (j) (k) (l)

Fig.6. Multi object tracking results for ‘Road’ video sequence using the proposed method for (a) frame 50, (b) frame 100, (c) frame 150, (d) frame 200, (e) frame 250, (f) frame 300, (g) frame 350, (h) frame 400, (i) frame 450, (j) frame 500, (k) frame 550, and (l) frame 600

After analyzing results of all five experiments, one can conclude that the proposed method works efficiently in case of different challenges of multi-object tracking such as object occlusion, object merging and splitting, re-entering of object in scene, etc.

6. CONCLUSIONS

In the proposed work, we have proposed a method for tracking of multiple objects in video using combination of Daubechies complex wavelet transform and Zernike moment as a feature of objects. Reduced shift sensitivity and better edge representation properties of Daubechies complex wavelet transform along with rotation invariance property of Zernike moment makes the proposed method suitable for tracking of multiple objects. In the proposed method our desired task is to track multiple human objects. The proposed method consists of three steps (i) Moving object segmentation with shadow removal, (ii) Moving object classification, and (iii) Multi object tracking. We have tested the performance of the proposed method in four different cases (i) no-occlusion, (ii) with occlusion, (iii) object merging and then splitting, and (iv) re-entering of object in scene. The experimental results have demonstrated the robustness of the proposed method and validates that the proposed method for multi-object tracking performs well for both indoor and outdoor video sequences, and also efficiently handles the problem of partial/full occlusion, merging and splitting objects, and re-entering of object in scene.

The main advantages of the proposed method is given as below:

- The proposed method is able to track the objects in the video having cluttered background, poor contrast as well as changing lighting conditions.
- The proposed method is able to handle different multi object tracking problems such as partial/full occlusion, re-entering of object in scene, etc.

The proposed method is able to track object efficiently where speed of moving objects is fast.

REFERENCES

[1] A. Yilmaz, O. Javed and M. Shah, “Object Tracking: A Survey”, *ACM Computing Surveys*, Vol. 38, No. 4, pp. 1-45, 2006.

[2] W. Hu and T. Tan, “A Survey on Visual Surveillance of Object Motion and Behavior”, *IEEE Transaction on System Man and Cybernetics*, Vol. 34, No. 3, pp. 334-352, 2006.

[3] L. Wang, W. Hu and T. Tan, "Recent Developments in Human Motion Analysis", *Pattern Recognition*, Vol. 36, No. 3, 585-601, 2003.

[4] A.S. Jalal and V. Singh, "A Multi-Resolution Framework for Multi-Object Tracking in Daubechies Complex Wavelet domain", *International Journal on Computational Vision and Robotics*, Vol. 3, No.1-2, pp. 52-74, 2012.

[5] M. Khare, R.K. Srivastava and A. Khare, "Object Tracking using Combination of Daubechies Complex Wavelet Transform and Zernike Moment", *Multimedia Tools and Applications*, Vol. 76, No. 1, pp. 1247-1290, 2017.

[6] A. Khare and U.S. Tiwary, "Daubechies Complex Wavelet Transform based Moving Object Tracking", *Proceedings of IEEE Symposium on Computational Intelligence in Image and Signal Processing*, pp. 36-40, 2007.

[7] O. Prakash and A. Khare, "Tracking of Non-Rigid Object in Complex Wavelet Domain", *Journal of Signal and Information Processing*, Vol. 2, No. 2, pp. 105-111, 2011.

[8] O. Prakash and A. Khare, "Tracking of Moving Object using Energy of Biorthogonal Wavelet Transform", *Chiang Mai Journal of Science*, Vol. 42, No. 3, pp. 783-795, 2015.

[9] M. Piccardi, "Background Subtraction Technique: A Review", *Proceedings of IEEE International Conference on System, Man and Cybernetics*, pp. 3099-3104, 2005.

[10] J.C. Huang and W.S. Hsieh, "Wavelet based Moving Object Segmentation", *IET Electronic Letters*, Vol. 39, No. 19, pp. 1380-1382, 2003.

[11] A. Baradarani, "Moving Object Segmentation using 9/7-10/8 Dual Tree Complex Filter Bank", *Proceedings 19th IEEE International Conference on Pattern Recognition*, pp. 1–4, 2008.

[12] M. Khare, R.K. Srivastava and A. Khare, "Moving Object Segmentation in Daubechies Complex Wavelet Domain", *Signal Image and Video Processing*, Vol. 9, No. 3, pp. 635-650, 2015.

[13] M. Khare, R.K. Srivastava and A. Khare, "Single Change Detection-Based Moving Object Segmentation by using Daubechies Complex Wavelet Transform", *IET Image Processing*, Vol. 8, No. 6, pp. 331-344, 2014.

[14] Y.P. Guan, "Spatio-Temporal Motion-Based Foreground Segmentation and Shadow Suppression", *IET Computer Vision*, Vol. 4, No. 1, pp. 50-60, 2010.

[15] M. Khare, R.K. Srivastava and A. Khare, "Daubechies Complex Wavelet based Computer Vision Applications", *Proceedings of International Conference on Recent Developments in Biometrics and Video Processing Techniques*, pp. 138-155, 2013.

[16] M. Khare, R.K. Srivastava and A. Khare, "Dual Tree Complex Wavelet Transform based Shadow Detection and Removal from Moving Objects", *Proceedings of International Conference on Electronic Imaging*, pp. 1-7, 2014.

[17] M. Khare, R.K. Srivastava and A. Khare, "Moving Shadow Detection and Removal-A Wavelet Transform based Approach", *IET Computer Vision*, Vol. 8, No. 6, pp. 701-717, 2014.

[18] A. Mittal and L. Davis, "M2tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene", *International Journal of Computer Vision*, Vol. 51, No. 3, pp. 189-203, 2003.

[19] K. Smith, D.G. Perez and J.M. Odobez, "Using Particles to Track Varying Numbers of Interacting People", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 962-969, 2005.

[20] P. Nillius, J. Sullivan and S. Carlsson, "Multi Target Tracking-Linking Identities using Bayesian Network Inference", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 2187-2194, 2006.

[21] M. Han, W. Xu, H. Tao and Y. Gong, "Multi-Object Trajectory Tracking", *Machine Vision and Applications*, Vol. 18, No. 3, pp. 221-232, 2007.

[22] R. Rad and M. Jamzad, "Real Time Classification and Tracking of Multiple Vehicles in Highways", *Pattern Recognition Letters*, Vol. 26, No. 10, pp. 1597-1607, 2005.

[23] A. Amer, "Voting based Simultaneous Tracking of Multiple Video Objects", *IEEE Transactions on Circuit Systems for Video Technology*, Vol. 15, No. 11, pp. 1448-1462, 2005.

[24] F.H. Cheng and Y.L. Chen, "Real Time Multiple Objects Tracking and Identification based on Discrete Wavelet Transform", *Pattern Recognition*, Vol. 39, No. 6, pp. 1126-1139, 2006.

[25] A. Senior, A. Hampapur, Y.L. Brown, S. Pankanti and R. Bolle, "Appearance Models for Occlusion Handling", *Journal of Image and Vision Computing*, Vol. 24, No. 11, pp. 1233-1243, 2006.

[26] K.P. Soman, K.L. Ramachandran and N.G. Resmi, "*Insight into Wavelets-From Theory to Practice*", Prentice Hall, 2010.

[27] I. Daubechies, "*Ten Lecture on Wavelets*", SIAM, 1992.

[28] D. Clonda, J.M. Lina and B. Goulard, "Complex Daubechies Wavelets: Properties and Statistical Image Modelling", *Signal Processing*, Vol. 84, No. 1, pp. 1-23, 2004.

[29] M. Teague, "Image Analysis via the General Theory of Moments", *Journal of Optical Society of America*, Vol. 70, No. 8, pp. 920-930, 1980.

[30] S.K. Hwang and Y.Y. Kim, "A Novel Approach to the Fast Computation of Zernike Moment", *Pattern Recognition*, Vol. 39, No. 11, pp. 2065-2076, 2006.

[31] G.A. Papakostas, Y.S. Boutalis, D.A. Karras and B.G. Mertzios, "A New Class of Zernike Moments for Computer Vision Applications", *Information Sciences*, Vol. 177, No. 13, pp. 2802-2819, 2007.

[32] Y. Gurwicz, R. Yehezkel and R.B. Lachover, "Multiclass Object Classification for Real Time Video Surveillance Systems", *Pattern Recognition Letters*, Vol. 32, No. 6, pp. 805-815, 2011.

[33] A. Khotanzed and Y.H. Hong, "Invariant Image Recognition by Zernike Moment", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, pp. 489-497, 1990.

[34] E.M. Celebi and Y.A. Aslandogan, "A Comparative Study of Three Moment based Shape Descriptors", *Proceedings of International Conference on Information Technology: Coding and Computing*, pp. 788-793, 2005.

[35] Y. Bin and P.J. Xiong, "Invariance Analysis of Improved Zernike Moments", *Journal of Optics A: Pure and Applied Optics*, Vol. 4, No. 6, pp. 606-614, 2002.

[36] M. Vorobyov, “Shape Classification Using Zernike Moments”, Technical Report, iCamp at University of California, 2011.

[37] M. Khare, R.K. Srivastava, A. Khare, N.T. Binh and T.A. Dien, “Human Object Classification in Daubechies Complex Wavelet Domain”, *Proceedings of 2nd International Conference on Context Aware System and Applications*, pp. 125-132, 2013

[38] INRIA Person dataset, Available at: http://pascal.inrialpes.fr/data/human, Accessed at 2017.

[39] D. Comaniciu, V. Ramesh and P. Meer, “Kernel based Object Tracking”, *IEEE Transactions of Pattern Recognition and Machine Intelligence*, Vol. 25, No. 5, pp. 564-575, 2003.