Human Tracking Based On Context Awareness In Outdoor Environment

Nguyen Thanh Binh[1], Ashish Khare[2] and Nguyen Chi Thanh[3]
[1]Faculty of Computer Science and Engineering,
Ho Chi Minh City University of Technology, VNU-HCM, Vietnam
ntbinh@hcmut.edu.vn
[2]Department of Electronics and Communication
University of Allahabad, India
ashishkhare@hotmail.com
[3]Faculty of Electronics and Computer Science Engineering
Cao Thang Technical College, Ho Chi Minh city, Vietnam
thanhngch1982@gmail.com

Received May 10, 2016; revised Decebmer 20, 2016; revised March 17, 2017; accepted March 30, 2017; published June 30, 2017

Abstract

The intelligent monitoring system has been successfully applied in many fields such as: monitoring of production lines, transportation, etc. Smart surveillance systems have been developed and proven effective in some specific areas such as monitoring of human activity, traffic, etc. Most of critical application monitoring systems involve object tracking as one of the key steps. However, task of tracking of moving object is not easy. In this paper, the authors propose a method to implement human object tracking in outdoor environment based on human features in shearlet domain. The proposed method uses shearlet transform which combines the human features with context-sensitiveness in order to improve the accuracy of human tracking. The proposed algorithm not only improves the edge accuracy, but also reduces wrong positions of the object between the frames. The authors validated the proposed method by calculating Euclidean distance and Mahalanobis distance values between centre of actual object and centre of tracked object, and it has been found that the proposed method gives better result than the other recent available methods.

Keywords: Human detection, shearlet transform, object tracking, human features, context awareness.

1. Introduction

The intelligent monitoring system has several components such as object classification, object recognition, object tracking, object measurement and have a lot of practical applications. Out of these components, the task of object tracking has an important role. In object tracking, motion of the object is analysed across different frames of a video and moving object is identified. The moving object tracking have a lot of practical applications. These applications are helpful for automatic control in surveillance applications. In some cases it give more accurate results than what humans can achieve in manual tracking and automatically handle complex situations without the need of human interventions. Some common applications of the moving object tracking are in security surveillance, traffic control, self-propelled equipment, control by gesture, etc. Moving object tracking is not an easy task as the moving object may present in different shapes and colours. Also moving object may be present in complex scenario (context) filled with turbulence. Therefore identification of moving objects in context aware situation is very difficult. The results may be affected by a number of environmental factors such as: changing lights, disturbances in monitoring devices, turbulence, obstruction etc. These factors make the moving object tracking as a more complicated task and demand results with high accuracy.

In this paper, the authors propose a new method for human tracking in outdoor environment based on human features in shearlet domain. The proposed method uses shearlet transform which exploit the transform domain features of human object with context-sensitiveness in order to improve the accuracy of human tracking in several other contexts as well. The proposed algorithm not only significantly improves the edge accuracy but also reduces wrong prediction of positions of the object between the frames. The proposed method was tested on a standard large datasets like PETS2009 dataset, SUN dataset, Caviar dataset, etc. The authors have compared the experimental results of the proposed method with other state-of-the-art methods such as Kernel Filter based method [19], Particle Filter based method [20], curvelet transform based method [21] and contourlet transform based method [22].

Main contribution of this paper is proposal of a method for human tracking based on features in uncertain outdoor environment. The rest of this paper is organized as follows: in section 2, the authors described the related works. The basics of feature selection, shearlet transform and its advantages for human object tracking are presented in section 3. Details of the proposed method for human object tracking are presented in section 4. Results of the proposed method and conclusions are given in sections 5 and 6 respectively.

2. Related work

In past time, there were many researchers who proposed methods to track moving objects. Many algorithms had been proposed with different efforts. Most of these methods were divided into four groups: contour-based [1], region-based [2], feature-based [3] and model-based [4].

Commonly used techniques of moving object tracking are based on background subtraction, statistical models, temporal differencing and optical flow [11, 12]. The algorithms based on background subtraction utilizes the current frame to compare it with the background image and detect the moving scene. Examples of some object tracking methods are median filter, Particle filter, temporal median filter, Kalman filter, sequential kernel density approximation and Eigen backgrounds [11, 13, 14, 17, 18]. The mean-shift algorithm utilizing color feature has been used to track the objects in video. This method has given

good tracking results [5, 6, 7]. However, performance of this algorithm is poor for blurred environment. The object tracking algorithms using features based on the points, shape and contour are used in many domains [8, 9, 10, 26].

A method for modelling the background to detect moving objects based on probability was proposed by Stauffer et.al [15]. Gaussian Mixture Model based tracking method is one example of such method. The key of this approach is treatment of a pixel value with a Gaussian mixture model and if a pixel does not match with the background then it is distributed to object motion. This technique is good in those cases where an object appears fixed in the background or fixed objects in the background disappear. In this method the background image will be updated after certain interval. Major disadvantage of this technique is that its performance is poor when object lighting conditions changes constantly or abruptly.

Andrzej et.al [16] proposed a tracking method using the optical flow. Optical flow method was originally suggested by Lucas-Kanade et.al [40]. The optical flow method calculates the motion vectors of the pixels between frames. Based on the motion vectors one can detect the moving object. This method is quite good because it is not sensitive to noise, but its major disadvantage is that it perform very poor in sudden changes in light conditions.

3. Shearlet transform and its advantages for human object tracking

3.1 Shearlet transform

Shearlets are similar to curvelets in the sense that both of them perform a multiscale and multidirectional analysis. There are two different types of shearlet systems: band-limited shearlet system and compactly supported shearlet system [23]. Computational complexity of the band-limited shearlet is high.

The digitization of discrete shearlet transform is performed in the frequency domain. The discrete shearlet transform is of the form [24]:

$$f \mapsto \langle f, \psi_n \rangle = \langle \hat{f}, \hat{\psi}_n \rangle$$

$$\langle \hat{f}, \hat{\psi}_n \rangle = \langle \hat{f}, 2^{-j\frac{3}{2}} \hat{\psi}(s_k^T A_{4^{-j}}) \exp^{2\Pi i <A_4^{-b}>} \rangle \quad with \quad b = jS_k m,.. \tag{1}$$

$$f \mapsto \langle \hat{f}, \hat{\psi}_n \rangle = \langle \hat{f}, 2^{-j\frac{3}{2}} \hat{\psi}(s_k^T A_{4^{-j}}) e^{2\Pi i <A_4^{-} jS_k m,..>} \rangle$$

where n = (j, k, m, i) are indexes - scale j, orientation k, position m, and cone i.

Shearlets perform a multiscale and multidirectional analysis. If f(x) is piecewise C^2, the approximation error of reconstruction with N-largest coefficients ($f_N(x)$) in the shearlet expansion is given by [25]:

$$\left\| f - f_N \right\|_2^2 \le B.N^{-2} (\log N)^3, \qquad N \to \infty \tag{2}$$

The authors have chosen shearlet transform because of its high directionality and representation of salient features (edges, curves and contours) of the image in a better way

compared with wavelet transform. Shearlet transform is useful for human detection and tracking due to its following properties [6]:

(i) Frame property: It is helpful for stable reconstruction of image.

(ii) Localization: Each of the shearlet frame elements need to be localized in both space and frequency domain.

(iii) Sparse approximation: It is useful for providing sparse approximation comparable to the band-limited shearlets.

The shearlet transform will produce a highly redundant decomposition when implemented in an undecimated form [27]. Similar to the curverlet transform, the most essential information of the image is compressed into relatively few large coefficients, which coincides with the area of major spatial activity in shearlet domain. On the other hand, noise is spread over all coefficients and at a typical noise level the important coefficients can be well recognized [21]. Thus setting the small coefficients to zero will not affect the major spatial activity of the image.

3.2 Adaboost classifier for human object tracking

For a given feature set and a training set of positive and negative images, Adaboost can be used both to select a small set of features and to train the classifier. Viola and Jones [28] firstly used binary Adaboost for face detection system. Boosting is a method to improve the performance of any learning algorithm, generally consist of sequential learning classifier [29]. Adaboost itself trains a cluster of weak-learners to form a strong classifier which performs at least as well an individual weak learner [30]. Adaboost clusters are particular features, where each feature represents an observable quantity associated with the target. The weak classifiers were basically thresholds for each data attribute. In our proposed work, we have used Adaboost algorithm which is described by Viola and Jones [28]. Complete Adaboost algorithm for classifier learning is given below.

Adaboost Algorithm [31]:

Given example images (I_1, J_1), (I_2, J_2),… (I_n, J_n) where $J_i = 0, 1$ for negative and positive example respectively.

Initialize weights $W_{1,i} = \frac{1}{2n}.\frac{1}{2p}$ for $j_i = 0,1$ respectively, where p and n are the number of positive and negative examples respectively.

For t = 1, 2 , … ,T (Number of iterations)

(i) Normalize the weights

$$W_{t,i} \leftarrow \frac{W_{t,i}}{\sum_{j=1}^{m} W_{t,j}} , \tag{3}$$

where, W_t is the probability distribution, m is the number of samples.

(ii) For each feature, *j*, train a classifier h_j, which is restricted by using a single feature. The error (E_j) is evaluated with respect to W_t

$$E_j = \sum_i W_i \left| h_j(I_i) - J_i \right| \tag{4}$$

(iii) Choose the classifier, h_t with the lowest error E_t.

Update the weights

$$W_{t+1,i} = W_{t,i}\beta_t^{1-\epsilon_i} \tag{5}$$

where ϵ_i=0, if example x_i is classified correctly and ϵ_i=1, otherwise

and

$$\beta_t = \frac{E_t}{1-E_t} \tag{6}$$

(iv) The final strong classifier is

$$h(\mathrm{I}) = \begin{cases} 1, & if \sum_{t=1}^{T} \alpha_t h_t(\mathrm{I}) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, & otherwise \end{cases} \tag{7}$$

where, $\alpha_t = \log \frac{1}{\beta_t}$

The advantages of adaboost classifier for human object tracking are as below:

- It is simple and easy to implement.
- There is no need for tuning of parameters.
- It is versatile in nature. It has been extended to learning problems well beyond binary classification.
- It is provably effective, provided that it can consistently find rough rules of thumb
- It can be combined with any learning algorithm.

In our proposed work for human tracking, we have taken the shearlet transform coefficients as a feature set and used adaboost classifier for training.

4. Human tracking based on features in context- awareness

In this section, the authors describe a method for human tracking using shearlet transform combined with human features. The overall proposed method for object tracking is shown in **Fig. 1**. The proposed method for human object tracking perform in two phases: training and testing.

In the training phase, there are three steps: First is detection of moving object. In this step, the authors used shearlet filter for computing search area and detect moving objects. In second step the authors extract the object features from the scene depending on color, shape and motion of human in context aware situation. These objects are saved as blob. Third step uses adaboost classifier for correct classification of human objects in changed context.

In the testing phase, there are four steps: the step 1 and step 2 are similar to that in training phase. The third step is matching and condition check. In this step, the authors match the results obtained at step 2 of the training phase with the result of step 2 of testing phase. Matched results imply that tracking of human object in current frame is correct and one can move to next frames (step 4). This process is repeated up to final frame. Otherwise, go to step 1 of the testing phase.

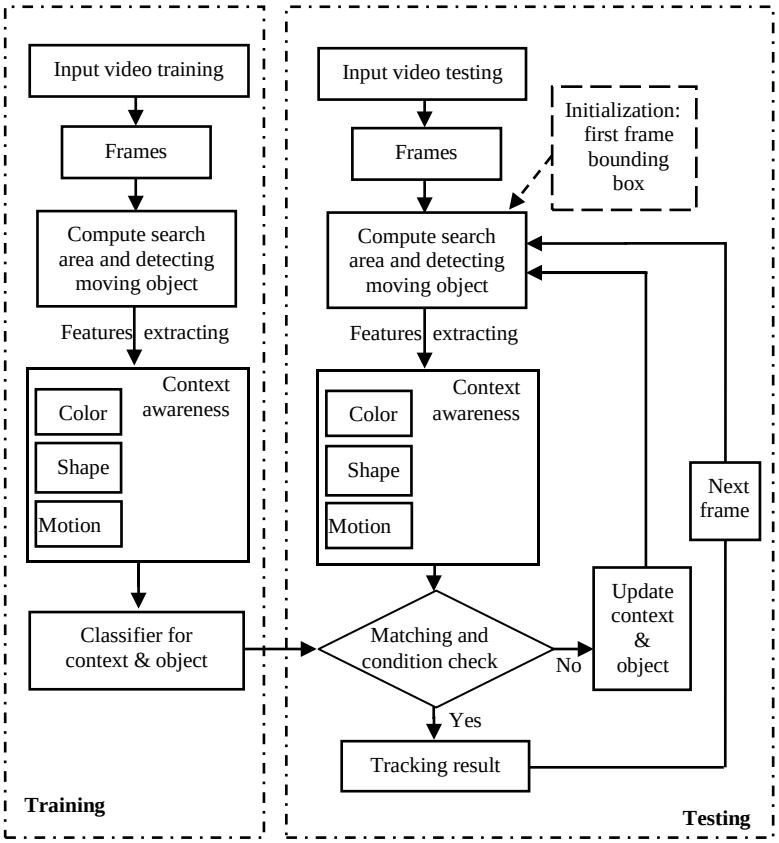


Fig. 1. The proposed method for object tracking.

Stages shown in **Fig. 1** are explained as below. -

4.1. Detection of moving objects

A video sequence contains a series of frames. Each frame can be considered as an image. The shearlet coefficients of each frame (image) are computed by using a filter bank and mean-shift algorithm [6] has been applied to detect moving objects in shearlet domain. This method reduces computation time significantly by utilizing the characteristics of high correlation of object area between adjacent frames. As mentioned in subsection 3.1, the authors have chosen shearlet transform because it has high directionality and it also represents salient features (edges, curves and contours) of the image in a better way compared to wavelet and other transforms.

The proposed method compute shearlet coefficients corresponding to the object in one frame and then by using the context aware information (as discussed in section 4.2) in form of color and shape based features, possible positions of object is identified. Then in these possible locations, shearlet coefficients have been computed and these coefficients are matched with shearlet coefficients computed in previous frame, to track the object. Thus the proposed algorithm checks the repetitive shearlet coefficients between two consecutive frames depending on the context aware information in videos, thereby reducing the frequency of calculation. The results have shown that the proposed method reduces the computation time, and it goes beyond the real-time requirements.

After detecting the objects, the method uses Grass-Fire method [13] for adding label. The adjacent pixels of the same object are assigned the same label. The centre of the object is calculated by taking the average of all point coordinates of objects. Bounding rectangle coordinates are based on the smallest and largest coordinates of the pixels present in the object. After that an MO-list is formed. MO-List is the list of moving objects. The objects which are inside MO-List are named as Mi.

4.2. Feature extraction for context awareness

After extracting the moving objects, the authors extracted the features of human objects. Here, we used three features: shape-based feature, color and motion-based features. For an independent human object, the positional distance of human object between two consecutive frames is small. Therefore we have compared the positional distance of human object in two consecutive frames for tracking purpose.

The objects in a scene are related to each other through spatial relations, or space-time relations in context domain and this relation is expressed through the relation between spatial locations of the object. There are two types of common spatial relations: metric and topology relations.

The metric spatial relation determine the distance between objects. Topological relation between the objects include the location of objects, their intersect, overlaps, matched exactly or relative positions in the direction of north-south-east-west.

Relations between the locations of objects are expressed by the relations of the object present at a time in those locations. The location itself has no temporal relations, which exist only through the presence of the object at that location. The location exists through the presence of the objects, but these objects may not be related to each other.

Based on the relationship with respect to space and time, the objects were classified as spatial objects (entities, location), temporal objects (events, phenomenon, time), spatio-temporal objects (entities, locations, times) and moving objects (entities, locations, times, trajectories).

Suppose the coordinates of objects, during tracking, in the n^{th} frame are (x, y). Then we need to search the object in $(n+1)^{th}$ frame. The authors calculated the distance coordinate (x_1, y_1) of these objects and compare them with coordinates (x, y). The distance K_c between the object at the n^{th} frame and $(n+1)^{th}$ frame is calculated using the formula of Euclidian distance:

$$K_c = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} \tag{8}$$

If T is the chosen distance threshold then if value of Kc is less than threshold T then this is treated as object tracked in the $(n+1)^{th}$ frame.

Important properties of the moving object are color, distance, area, velocity and speed [36]. Color feature is more informative and is useful for object detection. Color of the object is identified by the color of clothes, skin color, etc. The basic color model consists of three colors channels: red, green and blue (R, G, B). Here HSV color model is considered for detection of color of human object. Object and non-object pixels are classified according to the distribution of hue and saturation [17] as:

$$p(I(x, y)|\ \alpha) = p(h(x, y)|\ \alpha)\ p(S(x, y)|\ \alpha) \tag{9}$$

Object gradient information is also used to determine exact boundaries of human object.

The authors set a rectangle as:

$$X_i = [x_i, y_i, w_i, h_i], \; i = 1,\ldots,N \tag{10}$$

where locations are (x_i, y_i) and (width, height) are (w_i, h_i), are within a certain range. Detected human rectangles are tracked by comparing shearlet coefficients.

In the present work, we have calculated the minimum bounding rectangle which contain the object area. Location coordinates of the left upper and right lower positions of rectangle are stored and two shape-based features are extracted for the rectangular area as follows [34, 35]:

Aspect ratio (AR) is the ratio between the width and height of a rectangular envelope:

$$\text{Aspect ratio} = \frac{height \;\; of \;\; \text{rectangle}}{\text{width} \;\; of \;\; \text{rectangle}} \tag{11}$$

Complexity of shape (CS) is the dispersion used to measure the complexity of an object (dispersedness)

$$\text{Dispersedness} = \frac{perimeter^2}{area} \tag{12}$$

where, perimeter is the number of boundary pixels of a region containing moving objects and area is the number of pixels lying inside the moving object.

The above shape-based features are used to discriminate a human object from other objects. Because the human object has the more complex shape therefore they will have a greater dispersion value.

The objects present in different context situation may be dynamic objects. The behavior of the object must be determinated based on the features of the object. We have used variance of optical flow vectors for dynamic objects. The main idea of the method is to use optical flow vector direction of moving objects over the time. This method seeks to change the position of the pixel from frame t^{th} to the next frame $(t+1)^{th}$. This idea can be used to cluster the pixels of the human body parts to analyze the motion of an object.

Optical flow methods are used to distinguish which objects are flexible or not (rigid and non-rigid). The authors have observed that: the movement of human object is flexible whereas the movement of other objects is not flexible. The flexible human objects like humans [41] have the parts like arms, legs, etc. moving in different directions so each angle is greater than the average motion vectors. Therefore, the feature, average gradient G of human object is higher than the average gradient G of other objects. Other objects are not flexible and they will have typically low value of gradient G. The non-human objects contain nearly the same pixels therefore the motion vectors are same and the angles are smaller than the average vectors (nearly 0). Also, the characteristic gradient G, generated by the movement of human object will be cyclical. Moreover, depending on the context awareness in the frame, the authors have used this method in order to make distinction between the movement of human object from other objects such as vehicles.

4.3. Adaboost classifier for human object tracking

After feature extraction from positive and negative datasets, training is performed using the Adaboost classifier as presented in subsection 3.2. Adaboost is a supervised learning

algorithm which is used for data classification. Adaboost is very effective for training of large dimensionality data and solve over fitting problem very well [11]. The authors collected sample images for training and testing the classifier. The authors have collected images for two classes: human class and classes which do not contain any types of human, from images taken from some standard datasets like PETS2009 dataset, SUN dataset, caviar dataset, etc.

4.4. Matching and condition check for tracking

MO-List is list of moving object mask of the current frame. Consider a null TO-list, which contain list of tracked object. If TO-List is null then M_i will be added to the TO-List for matching and tracking. If the TO-List object contain the matching object then the manipulation and checking conditions will be implemented. To match up T_j and M_i, Euclidean distance is used to measure the distance between the center of each M_i and the predicted center of all the T_j. If centre of M_i and T_j have a difference less than a given threshold δ then $M_i \equiv T_j$. The δ threshold is defined as the greatest distance between two objects of the frame.

To increase the accuracy, boundary of M_i is compared with estimated boundary of T_j. If the difference is smaller than the threshold δ then $M_i \equiv T_j$. The above matching can lead to following three situations:

If M_i do not coincide with any T_j then M_i will be considered as a new object and it will be added to the TO-List. If the single M_i coincides with the T_j then M_i will be seen as a tracked object and T_j will update the information in the new location. If the existence of T_j is not updated with the information then T_j will be evaluated if it is obscured and it will be eliminated if already out of the viewport.

As M_i is added to TO-List, the index of T_j object which is the largest T_j will be increased by 1. If T_j is the first object then there will be an index of 1. Label of T_j object will be assigned based on the object index of T_j. Object center and rectangular envelope of T_j will be taken from Mi.

If there is a M_i which overlap with a single T_j then M_i will be seen as tracked object and update the information in the new location for T_j. If the central predictions of T_j are still within regions then T_j is considered as hidden object and is not rejected.

We are also updating the information estimates. After T_j has updated the information, the label of object will be kept. Object center and rectangular of T_j will be taken from M_i. Now, the authors predict the location of objects. After adding T_j, updating T_j and estimating T_j, the authors calculate the central predictions of T_j . The position and size of the object X_i are:

$$X_i = (x_i, y_i, w_i, h_i) \tag{13}$$

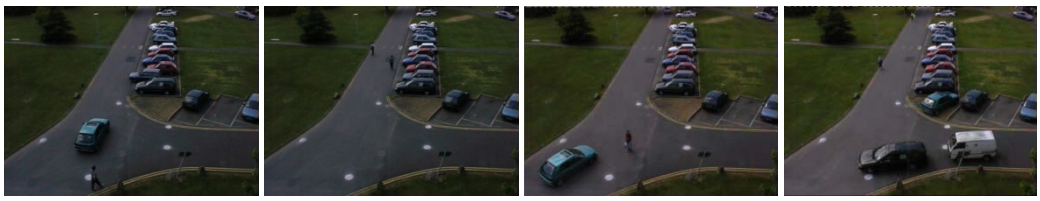
The central predictions of T_j will be used to match the center of M_i. The objects in the context are the dynamic objects. The behaviour of the object must be determined based on the features of the object. After that, the object and context will be updated for next frame.

5. Experiments and results

In this section, we present experiments of human object tracking. For tracking, we determine the object in each frame of the video. The object area is determined in the first frame and the tracking algorithm needs to track the object from frame to frame.

5.1 Materials

Experiments were conducted using Matlab 2013a and carried out on a computer of Intel i7 4700MQ 2.4 GHz CPU and 16 GB DDR3 memory. The experiments were focusing on the outdoor scenes so PETS 2009 dataset [37], CAVIAR dataset [38], and SUN dataset [39] are used. Videos used are of the resolution of 384 × 288 or 768 x 576 and the frame rate is 25 frames per second. First 2000 frames of the video are used for training. Training data comprise of 1364 human and 892 car objects. Images of some scene and sample bounding boxed images of human and car are shown in **Fig. 2**.



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

Fig. 2. Scenes in dataset. (a) and (b) are scenes in training data. (c) and (d) are scenes in testing data.

5.2 Result analysis

Here, we report the results obtained from some of video clips. Our first experiment is on person video clips with frame size 384 by 288. The proposed method processes this video clip at 25 frames/second. The authors have experimented on the video up to 1000 frames. Here, we have reported the results in difference of 50 frames. Some snapshots of results achieved are shown in **Fig. 3**.

Frame 100 Frame 250 Frame 350 Frame 500

Fig. 3. Tracking in human video clips up to 500 frames

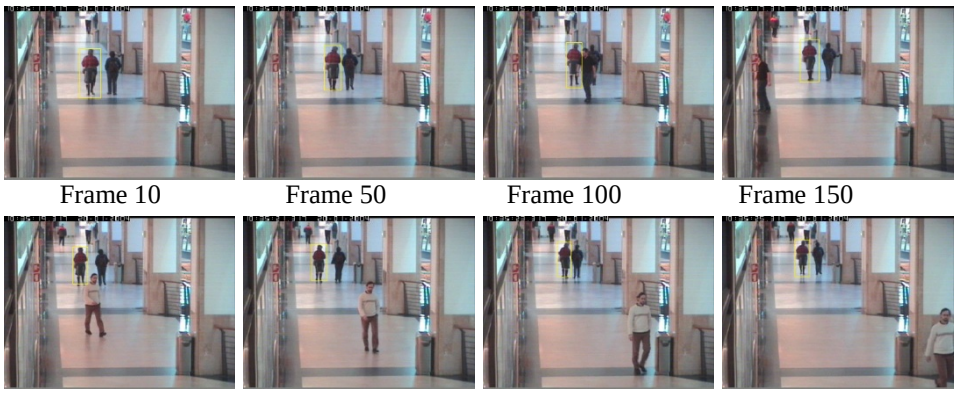
Our second experiment is on person video clips with frame size 384 by 288. The proposed method processes this video clip at 25 frames/second. The authors have experimented on the video up to 1000 frames. Here, we report the results up to 1000 frames. Some results achieved are shown in **Fig. 4**.

Frame 1 Frame 50 Frame 100 Frame 150

Frame 200 Frame 250 Frame 300 Frame 350

Fig. 4. Tracking in human video train clips up to 350 frames

Our third experiment is on person video clip, of Caviar dataset, with frame size 384 by 288. The proposed method processes this video clip at 25 frames/second. The authors have experimented on the video up to 382 frames. Here, we have reported the results up to 350 frames. Some results achieved are shown in **Fig. 5**.

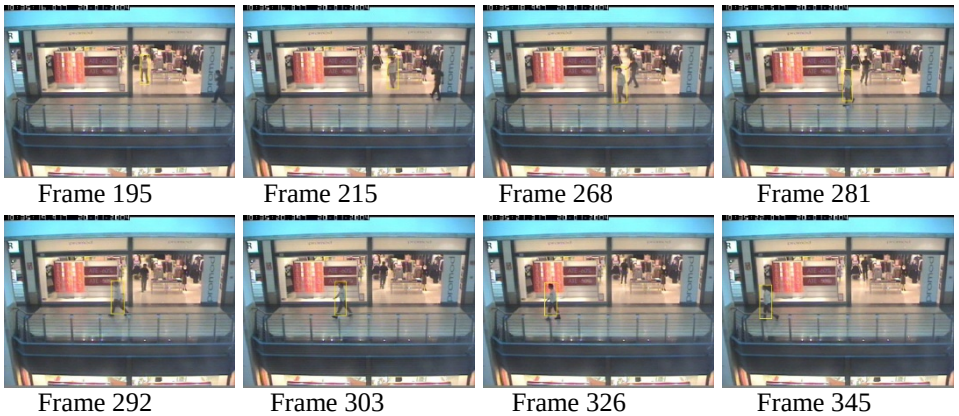


Frame 10 Frame 50 Frame 100 Frame 150

Frame 200 Frame 250 Frame 300 Frame 350

Fig. 5. Tracking in human video train clips up to 350 frames

Our fourth experiment is on person video clip, of Caviar dataset, with frame size 384 by 288. The proposed method processes this video clip at 25 frames/second. The authors have experimented on the video up to 382 frames. Here, we have reported results up to 345 frames. Some results achieved are shown in **Fig. 6**.



Frame 195 Frame 215 Frame 268 Frame 281

Frame 292 Frame 303 Frame 326 Frame 345

Fig. 5. Tracking in human video train clips up to 345 frames

From **Fig. 3** to **6**, we have observed that the proposed method performs well.

Table 1. Comparison of the object tracking error of the proposed method with other methods.

Frame Number	Kernel tracking [19]	Particle tracking [20]	Curvelet tracking [21]	Contourlet tracking[22]	Proposed Tracking
100					
150					
200	Error				
250					
300		Error	Error		
350				Error	
400	Error				
450					
500		Error			
550					
600	Error	Error	Error	Error	Error
650					
700	Error				
750		Error	Error		
800					
850	Error				
900	Error			Error	
950					
1000		Error			

In **Table 1**, the authors compared the error in human tracking of the proposed method to other methods for one video. The 'error' shown in table 1 is the error where the method could not tract the accurate object. At 600th frame, the quality of video is not good and there are many abrupt changes therefore the result of all methods are showing 'error' in this frame. From **Table 1**, we observed that the results of the proposed method are better than other methods.

Table 2. Comparison of the object tracking error of the proposed method with other methods.

Frame Number	Kernel tracking [19]	Particle tracking [20]	Curvelet tracking [21]	Contourlet tracking[22]	Proposed Tracking
100					
234	Error				
278					
315			Error	Error	
427	Error				
512		Error			
578			Error		
621					
692	Error	Error		Error	
714	Error				

748					
789	Error	Error	Error	Error	Error
800					

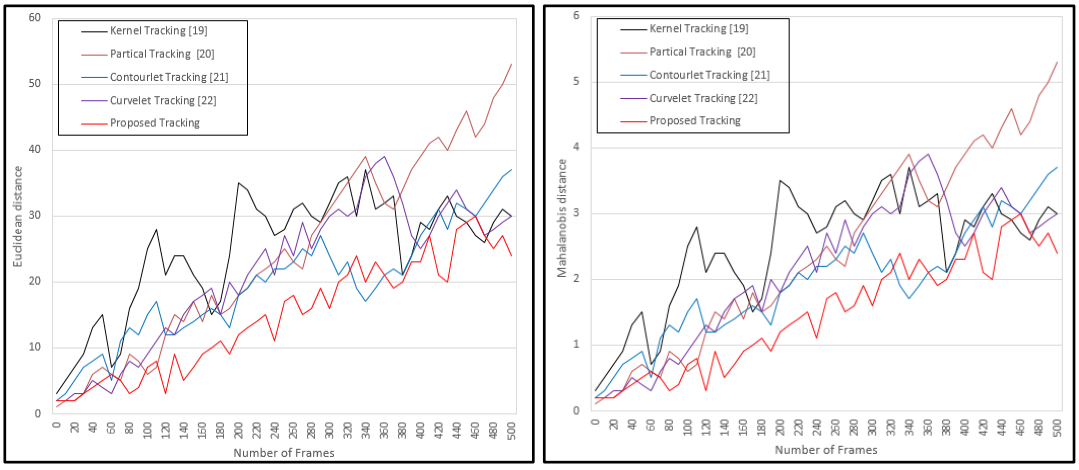
In **Table 2**, the authors compared the error in human tracking of the proposed method to other methods for another video. The 'error' shown in table 2 is the error where the method could not track the accurate object. At 789th frame, the quality of video is not good and there are many abrupt changes therefore the result of all methods are showing 'error' in this frame. From **Table 2**, we have observed that the results of the proposed method are better than other methods.

5.3 Performance evaluation

The visual results combined with the quantitative performance metrics are more appropriate to evaluate different tracking methods. For performance evaluation, we used two different performance metrics: Euclidean distance and Mahalanobis distance. The Euclidean distance between the computed centroid (x_C, y_C) of tracked human window and actual centroid (x_A, y_A) is as follows:

$$K_c = \sqrt{(x_A - x_c)^2 + (y_A - y_c)^2} \tag{14}$$

Fig. 7a shows the plot of Euclidean distance between centroid of the tracked object computed by the proposed method and other tracking methods. The video test as **Fig. 2.** From the Fig. it is clear that the proposed method has the least Euclidean distance between the centroid of tracked bounding box and the actual centroid in comparison with other methods.



(a) The Euclidean distance (b) The Mahalanobis distance

Fig. 7. The Euclidean and Mahalanobis distance of the proposed method and other tracking methods

The Mahalanobis distance is a measure of the distance between a point P (x_p, y_p) and a distribution D. The Mahalanobis distance of an observation $a = (a_1, a_2, a_3, \ldots, a_n)^T$ from a set of observations with mean $\alpha = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n)^T$ and covariance matrix S is defined as [40, 41]:

$$D_M(a) = \sqrt{(a-\alpha)^T S^{-1}(a-\alpha)^2} \tag{15}$$

It is a dissimilarity measure between two points A = (x_A, y_A) and B = (x_B, y_B) with the covariance matrix S.
Suppose, we have two features: feature i (ground truth centroid points) and feature j (computed centroid points). Let {x(1, i), x(2, i),….. x(n, i)} be a set of n examples of feature i. Let {x(1, j), x(2, j),…… x(n,j)} be a set of n examples of feature j. Let m(i) be the mean of feature i, and m(j) is mean of feature j. The covariance of feature i and j is computed as [44]:

$$s(i,j) = \frac{\{[x(1,i)-m(i)][x(1,j)-m(j)]+\ldots+[x(n,i)-m(i)][x(n,j)-m(j)]\}}{n-1} \tag{16}$$

Mahalanobis distance is calculated as procedure described in [44]:

Fig. 7b shows the plot of Mahalanobis distance between centroids of the tracked object computed by the proposed method and other tracking methods. From **Fig. 7b** it is clear that the values of dissimilarity measure are small in the proposed method.

6. Conclusions and feature works

Object tracking is a task to find a series of actions of moving objects in between video frames. This gives the information about the object such as the path of the object, the speed or direction of motion of the object. The moving objects with different shapes can be both moving and changing in shapes, color and may present in a complex context and full of turbulence. In outdoor environment, the objects are usually occluded, blurred or noisy. The shape of the object may change from scene to scene and from frame to frame. Therefore, object tracking in these cases is a challenging problem. In this paper, a new method of object tracking by combining the features of the object in form of shearlet coefficients with context-sensitive information is proposed, in order to improve the accuracy of object tracking. The authors evaluated the proposed method by calculating Euclidean distance and Mahalanobis distance between centroids of actual object and tracked object values. Experimental results have demonstrated that the proposed method perform well as compared to the other methods such as Kernel Filter based method [19], Particle Filter based method [20], curvelet transform based method [21] and contourlet transform based method [22]. The proposed method was tested on standard datasets like PETS2009, SUN dataset, and Caviar dataset. The proposed algorithm not only significantly improves the edge accuracy, but it reduces the wrong position of objects between the frames. The proposed method have a good degree of tracking accuracy and its performance is also in real time. The limitation of the proposed method is that it does not work in case of tracking of multiple objects. In future work, we will improve the accuracy of the proposed method by incorporating more human features to track multiple objects and also for real-time implementations in very complex enviorment, the authors will try to track the objects with help of hardwares like DSP kits.

Acknowledgements

We would like to thank to Reviewers of this paper for given important comments. We are also thankful to the authors of datasets which are used in this paper.

References

[1]. Erdem C. E., Tekalp A. M., Sankur B., "Video object tracking with feedback of performance measures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no 4, pp 310–324, 2003 Article (CrossRef Link)

[2]. Liu T. L., Chen H. T., "Real-time tracking using trust region methods," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no 3, pp 397–402, 2004. Article (CrossRef Link)

[3]. Wang D., "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no 5, pp 539–546, 1998. Article (CrossRef Link)

[4]. Zhou S. K., Chellappa R. , Moghaddam B., "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Transactions on Image Processing*, vol. 13, no 11, pp 1491–1506, 2004. Article (CrossRef Link)

[5]. Comaniciu D., Ramesh V. , Meer P., "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no 5, pp 564–577, 2003. Article (CrossRef Link)

[6]. Comaniciu D., Meer P., "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no 5, pp 603–619, 2002. Article (CrossRef Link)

[7]. Zivkovic Z. , Krose B., "An EM-like algorithm for color histogram- based object tracking," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, USA, pp 798–803, 2004. Article (CrossRef Link)

[8]. Yilmaz A., Javed O., Shah M., "Object Tracking: A survey," *ACM Computing Surveys*, vol. 38, no 4, 2006. Article (CrossRef Link)

[9]. Wang D., "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no 5, pp 539–546, 1998. Article (CrossRef Link)

[10]. Moeslund T.B., Granum E., "A survey of computer vision based human motion capture," *Computer Vision and Image Understanding*, vol. 81, pp 231-268, 2001. Article (CrossRef Link)

[11]. Ruolin Zhang, Jian Ding, "Object tracking and detecting based on adaptive background subtraction," in *Proc. of International workshop on information and electronics engineering*, Vol. 29, pp1351–1355, 2012. Article (CrossRef Link)

[12]. Wei Shuigen Chen Zhen, Dong Hua, "Motion detection based on temporal difference method and optical flow field," in *Proc. of Second international symposium on electronic commerce and security*, vol. 2, pp 85-88, 2009. Article (CrossRef Link)

[13]. Masafumi S., Thi T. Z., Takashi T., Shigeyoshi N, "Robust rule-based method for human activity recognition," *International journal of computer science and network security*, vol.11, no.4, pp. 37-43, 2011. Article (CrossRef Link)

[14]. Johnsen, S., Tews, A, "Real-time object tracking and classification using a static camera," in *Proc. of International Conference on Robotics and Automation, workshop on People Detection and Tracking*, 2009. Article (CrossRef Link)

[15]. Stauffer, C., Grimson, W. E. L, "Adaptive background mixture models for real-time tracking," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999. Article (CrossRef Link)

[16]. Andrzej Głowacz, Zbigniew Mikrut, Piotr Pawlik, "Video Detection Algorithm Using an Optical Flow Calculation Method," in *Proc. of 5th International Conference on Multimedia Communications, Services and Security*, pp 118-129, 2012 Article (CrossRef Link)

[17]. Peng Dai, Linmi Tao, Guangyou Xu, "Dynamic context driven human detection and tracking in meeting scenarios," in *Proc. of 2nd International Conference on Computer Vision Theory and Applications*, Volume Special Session, pp 31-38, 2007. Article (CrossRef Link)

[18]. Qi Zang, Reinhard Klette, "Object classification and tracking in video surveillance," *Computer Analysis of Images and Patterns*, vol. 2756, pp. 198-205, 2003. Article (CrossRef Link)

[19]. Comaniciu D., Ramesh V., Meer P., "Kernel based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 25, no 5, pp 564–577, 2003. Article (CrossRef Link)
[20]. Nummiaro K., Meier E. K., Gool L.J.V., "An Adaptive Color-based Particle Filter," *Image and Vision Computing*," vol. 21, pp 99-110, 2003. Article (CrossRef Link)
[21]. Nguyen Thanh Binh, A.Khare, "Object tracking of video sequences in curvelet domain," *International Journal of Image and Graphics*, vol. 11, no 1, pp 1-20, 2011. Article (CrossRef Link)
[22]. Nguyen Thanh Binh, T.A.Dien, "Object detection and tracking in contourlet domain," in *Proc. of the First International Conference on Context-Aware Systems and Applications*, Viet Nam, vol. 109, pp 192–200, 2012. Article (CrossRef Link)
[23]. Gitta Kutyniok, Demetrio Labate, "Shearlets: multiscale analysis for multivariate data," Birkhauser, ISBN 0817683151 9780817683153, 2012.
[24]. Wang-Q Lim, "The discrete shearlet transform: a new directional transform and compactly supported shearlet frames," *IEEE Transactions on image processing*, vol 19, no 5, pp 1166 – 1180, 2010. Article (CrossRef Link)
[25]. K. Guo, D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Anal.*, vol. 39, pp. 298–318, 2007. Article (CrossRef Link)
[26]. Nguyen Thanh Binh, "Vehicle tracking in outdoor environment based on curvelet domain," in *Proc. of the International Conference on Nature of Computation and Communication*, Viet Nam, vol. 144, pp 360-369, 2014. Article (CrossRef Link)
[27]. Vishal M. Patel, Glenn R. Easley, Dennis M. Healy, "Shearlet-based deconvolution," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp 2673-2685, 2009. Article (CrossRef Link)
[28]. Viola, P. & Jones, M, "Rapid object detection using a boosted cascade of simple feature," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp 83-87, 2001 Article (CrossRef Link)
[29]. Zhu, Z., Zou, H., Rosset, S., Hastie, T, "Multiclass adaboost," *International Journal of Statistics and its Interface*, pp 349-360, 2009. Article (CrossRef Link)
[30]. Renno, J. P., Makris, D., Jones, G. A, "Object classification in visual surveillance using adaboost," in *Proc. of International Conference on Computer Vision and Pattern Recognition*, pp 1-8, 2007. Article (CrossRef Link)
[31]. M.Khare, Nguyen Thanh Binh, R.K. Srivastava, A. Khare, "Vehicle identification in traffic surveillance – complex wavelet transform based approach," *Journal of Science and Technology*, vol 52, no 4A, pp 29-38, 2014.
[32]. Castleman K. R., "Digital Image Processing," *Prentice Hall, Englewood Cliffs*, USA, 1996.
[33]. Manish Khare, Nguyen Thanh Binh, Rajneesh Kumar Srivastava, "Human object classification using dual tree complex wavelet transform and zernike moment," *Transactions on Large-Scale Data- and Knowledge-Centered Systems XVI*, pp 87-101, 2014. Article (CrossRef Link)
[34]. Zhihua, Li , Fan Zhou , Xiang Tian, Yaowu Chen, "High efficient moving object extraction and classification in traffic video surveillance," *Journal of Systems Engineering and Electronics*, vol. 20, no 4, pp. 858-868, 2009. Article (CrossRef Link)
[35]. Nguyen Thanh Binh, "Human object detection based on context awareness in the surroundings," *EAI Endorsed Transactions on Context-aware Systems and Applications*, vol. 2, issue 4, 2015. Article (CrossRef Link)
[36]. Tang Sze Ling, Liang Kim Meng, Lim Mei Kuan, Zulaikha Kadim, Ahmed A. Baha Al-Deen, "Colour-based object tracking in surveillance application," in *Proc. of International Multi Conference of Engineers and Computer Scientists*, vol. 1, 2009. Article (CrossRef Link)
[37]. PESTS2009 : http://ftp.pets.rdg.ac.uk/pub/PETS2009/
[38]. CAVIAR: http://groups.inf.ed.ac.uk/vision/caviar/
[39]. SUN: http://vision.princeton.edu/projects/2010/SUN/
[40]. Lucas, B.D Kanade T. "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Understanding Workshop*, pp 121-130, 1981. Article (CrossRef Link)

[41]. Tsuchiya M, Fujiyoshi H, "Evaluating feature importance for object classification in visual surveillance," in *Proc. of The 18th IEEE International Conference on Pattern Recognition*, pp. 978–981, 2006. Article (CrossRef Link)
[42]. Zhang T, Ghanem B, Liu S, Ahuja N, "Robust visual tracking via structured multi-task sparse learning," *Journal of Computer Vision*, vol. 101, pp 367–383, 2013. Article (CrossRef Link)
[43]. Zhong W, Lu H, Yang MH, "Robust object tracking via sparsity-based collaborative model," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1838–1845, 2012. Article (CrossRef Link)
[44]. Manish Khare, Rajneesh Kumar Srivastava, Ashish Khare, "Object tracking using combination of daubechies complex wavelet transform and zernike moment," *Multimedia Tools and Applications*, doi:10.1007/s11042-015-3068-5, pp 1–44, 2015. Article (CrossRef Link)

Nguyen Thanh Binh received the Bachelor of Engineering degree from Ho Chi Minh City University of Technology -Vietnam National University at Ho Chi Minh City (VNU-HCM) in 2000, the Master's degree and Ph.D degree in computer science both from University of Allahabad, India, in 2005 and 2011 respectively. Now, he is a lecturer at Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology, VNU-HCM. He has published one book, one book chapter and more than 50 research papers. His research interests include recognition, image processing, multimedia information systems, decision support system, and time series data.

Ashish Khare received M.Sc. (Computer Science) and Ph.D. degree both from University of Allahabad, Allahabad, India, in 1999 and 2007 respectively. He did his post-doctoral research from Gwangju Institute of Science and Technology, Gwangju, Korea, in 2007-2008. Presently he is working as an Associate Professor (Computer Science) in Department of Electronics and Communication, University of Allahabad, Allahabad, India. 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. Six Ph.D. students have completed their research work under his supervision. He has published one book, two book chapters and more than 100 research papers. He has also served as Guest Editor of reputed SCI and Scopus indexed international journals.

Nguyen Chi Thanh received the Bachelor degree in information system from University of Science-Vietnam National University at Ho Chi Minh City in 2007, the Master's degree in information system from Le Qui Don Technical University in 2015. Now, he is a lecturer at Faculty of Electronics and Computer Science Engineering, Cao Thang Technical College, Ho Chi Minh city, Vietnam. His research interests include recognition, image processing.