

# Human Object Tracking in Nonsampled Contourlet Domain

Nguyen Thanh Binh

Faculty of Computer Science and Engineering,  
Ho Chi Minh City University of Technology, VNU-HCM  
Ho Chi Minh City, Vietnam

**Abstract**—The intelligent systems are becoming more important in life. Moving objects tracking is one of the tasks of intelligent systems. This paper proposes the algorithm to track the object in the street. The proposed method uses the amplitude of zernike moment on nonsampled contourlet transform to track object depending on context awareness. The algorithm has also been processed successfully such cases as the new object detection, object detection obscured after they reappeared, detecting and tracking objects which successfully intertwined and then separated again. The proposed method tested on a standard large dataset like PEST dataset, CAVIAR dataset and SUN dataset. The author has compared the results with the other recent methods. Experimental results of the proposed method performed well compared to the other methods.

**Keywords**—object tracking; zernike moment; nonsampled contourlet transform; context awareness; extracting features

## I. INTRODUCTION

The intelligent systems are becoming more important in life. Building an intelligent surveillance system can be split into four main challenges: moving object detection, object classification, object tracking and behavior recognition. Moving object tracking is one of the tasks of intelligent systems. Object tracking can be applied in many place such as security surveillance systems for airports, train stations, schools, or buildings of government etc.; or traffic control systems like automatic traffic signal systems, street-crossing safety systems, traffic density statistics systems etc.

In past time, there are many researchers who proposed the methods to track moving objects. Most of these methods are divided into four groups such as contour-based [1], region-based [2], feature-based [3] and model-based [4] algorithms. Moving object detection used common techniques such as: background subtraction, statistical models, temporal differencing and optical flow [11, 12]. Algorithm based on background subtraction utilizes the current image to compare it with the background image and detect the moving scene. Most methods of background subtraction are median filter, mean filter, temporal median filter, Kalman filter, sequential kernel density approximation and Eigen backgrounds [11, 13]. The mean-shift algorithm using colors has been used to track the objects in video. This method has improved tracking results [5, 6, 7]. However, the implementations of them for object blur are complex. The object tracking algorithms using feature are performed based on the point, shape and contour in many domains [8, 9, 10]. Johnsen [14] gave the model to track

objects through static camera with motion detection technique using background subtraction in which the background generated using filters approximated Median Filter. After detecting the moving area, he uses two-pass connected component labeling for identifying and selecting the motion area to make for object tracking step. Next, the model boundary objects are separated using standard RGB color and use the Kalman filter to predict the next location of the object. Finally, using apriori assignment combined with Euclidean metric distance, Bhattacharya to track objects. The drawback of this method is to require the appropriate reference background image. If there are changes of light compared to the reference background image, the current area will be considered as the motion that should not be able to detect moving objects. The object tracking is hard work. This task has many challenges.

In this paper, the author proposes a method to implement for human tracking based on their contour. The proposed method uses the amplitude of zernike moment on NonSampled Contourlet Transform (NSCT) to track object depending on context awareness. The algorithm has also been processed successfully such cases as the new object detection, object detection obscured after they reappeared, detecting and tracking objects which successfully intertwined and then separated again. The proposed method tested on a standard large dataset like PEST (Performance Evaluation of Tracking and Surveillance) dataset, CAVIAR (Context Aware Vision using Image-based Active Recognition) dataset and SUN (Scene UNDERstanding) dataset. The author has compared the results with the other recent methods. Experimental results of the proposed method performed well compared to the other methods such as Kernel Filter method [14], Particle Filter method [15], curvelet method [16] and contourlet method [17]. The rest of this paper is organized as follows: in section 2, the author described the basic of NSCT, zernike moment and its advantages for human object tracking. And details of the proposed method for human object tracking are presented in section 3. The result and conclusion of the paper are presented in section 4 and 5 respectively.

## II. SELECT A NEW GENERATION WAVELET TRANSFORM FOR TRACKING

The Discrete Wavelet Transform (DWT) provides a fast, local, sparse, multiresolution analysis of real-world signals and images. Although DWT is a powerful tool for signal and image analysis, it has three serious disadvantages: shift

sensitivity, poor directionality and lack of phase information. To improve these drawbacks, the new generation wavelet transforms such as the ridgelets, curvelets, contourlet transform and NSCT have been proposed.

Curvelets and ridgelets take the form of basic elements which exhibit very high directional sensitivity and are highly anisotropic. In two-dimensions, for instance, curvelets are localized not only in position (the spatial domain) and scale (the frequency domain), but also in orientation. Unlike wavelet transforms, the ridgelet transform processes data by first computing integrals over lines with all kinds of orientations and locations. The curvelet transform, like the wavelet transform, is a multiscale transform, with frame elements indexed by scale and location parameters. Unlike the wavelet transform, it has directional parameters, and curvelet pyramid contain elements with a very high degree of directional specificity. In addition, the curvelet transform is based on a certain anisotropic scaling principle which is quite different from the isotropic scaling of wavelets. However, curvelet also has two drawbacks: first, not optimal for sparse approximation of curve beyond  $C^2$  singularities and second, highly redundant [18].

Contourlet transform [19] is built from a discrete domain first, then extend to the continuous domain, it has lower redundancy and a faster discrete implementation version than curvelet. But contourlet is just multidirectional and multi-scale but not shift - invariant, which causes pseudo-Gibbs phenomena visible on the decoded image by high compression ratio. NSCT [20] brings shift - invariance for contourlet with the trade-off of more redundancy. As NSCT is used for contour detection, this redundancy is not a drawback but even gives better results.

The important feature of the proposal bases on object's contour and NSCT is chosen to extract the object's contour. NSCT belongs to the family of the new generation wavelet transform. The NSCT development through its main predecessors may start at wavelet transform [21] as in [22]. The properties of NSCT like multi-scale, multi-direction, shift-invariant make it very suitable for contour detection.

#### A. Nonsubsampling contourlet transform

NonSubsampled Pyramid (NSP) is similar to Laplacian pyramid. This algorithm uses two-channel nonsubsampling filter banks and decomposition is realized with level  $j = 3$  stages. The filters of subsequent stages are the result of upsampling filters of the first stage. The subbands are divided into high-pass and low-pass filter. The region of low-pass filter is  $[-\frac{\pi}{2j}, \frac{\pi}{2j}]^2$ , and the complement of the low-pass filter is the high-pass filter which is solved by the region  $[-\frac{\pi}{2j-1}, \frac{\pi}{2j-1}]^2 \setminus [-\frac{\pi}{2j}, \frac{\pi}{2j}]^2$ .

Directional Filter Bank (DFB) eliminates the down-samplers and up-samplers in each two-channel filter bank that is NonSubsampled Directional Filter Bank (NSDFB). NSCT is the process that combines NSP and NSDFB. Ping [23] also asserts that NSCT is implemented by two major steps: when multi-scale image multi-scale decomposition is NSP filter to be used in a low-pass subband and a band-pass subband. Then,

each level of band pass subbands directional decomposition of band-pass subbands will continue with NSDFB. And image decomposition process is complete with two steps: NSP and NSDFB.

#### B. Zernike Moment (ZM)

ZM [24] was firstly introduced by Teague [25] to overcome the shortcomings of information redundancy present in geometric moments [25]. ZM can represent the properties of an image with no redundancy or overlap of information between the moments [26]. Due to these characteristics, ZM has been utilized as a feature set in different applications such as object classification, shape analysis, content based image retrieval etc.

ZM owns the following properties [24, 27, 28]:

- (i) ZM is rotation, translation and scale invariant [29].
- (ii) ZM is robust to noise and minor variation in shape.
- (iii) Since the basis of ZM is orthogonal, therefore they have minimum information redundancy.
- (iv) ZM can characterize the global shape of pattern. Lower order moments represent the global shape pattern and higher order moment represents the detail.
- (v) An image can be better described by a small set of its ZM than any other types of moments [24].

ZM is a set of complex polynomial which forms a complete orthogonal set over the interior of the unit circle of  $x^2 + y^2 \leq 1$  [30, 31]. These polynomials are of the form,

$$V_{mn}(x, y) = V_{mn}(r, \theta) = R_{mn}(r) \cdot \exp(jn\theta) \quad (1)$$

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

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

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

The above-mentioned polynomial in equation (2) is orthogonal and satisfies the orthogonality principle:

$$\iint_{x^2+y^2 \leq 1} V_{mn}(x, y) \cdot V_{pq}(x, y) dx dy = \frac{\pi}{n+1} \delta_{mp} \delta_{nq} \quad (3)$$

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

ZM is 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 [24, 32].

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

$$Z_{nm} = \frac{m+1}{\pi} \iint_{x^2+y^2 \leq 1} I(x, y) [V_{nm}(r, \theta)] dx dy \quad (4)$$

In case of digital image, the integral are replaced by summation, given as follows

$$Z_{nm} = \frac{m+1}{\pi} \sum_x \sum_y I(x, y) V_{nm}(r, \theta), \quad x^2 + y^2 \leq 1 \quad (5)$$

ZM has its own properties such as: shift-invariant [33], translation invariant [28], rotation invariant [29] etc. So if the author use the combination of NSCT and ZM features in one methodology then the author can expect for more accurate object tracking results.

### III. OBJECT TRACKING BASED ON NSCT COMBINED WITH ZERNIKE MOMENT

In this section, the author proposed the method for object tracking based on NSCT combined with ZM depending on context awareness. The overall of the proposed method is to present as figure 1.

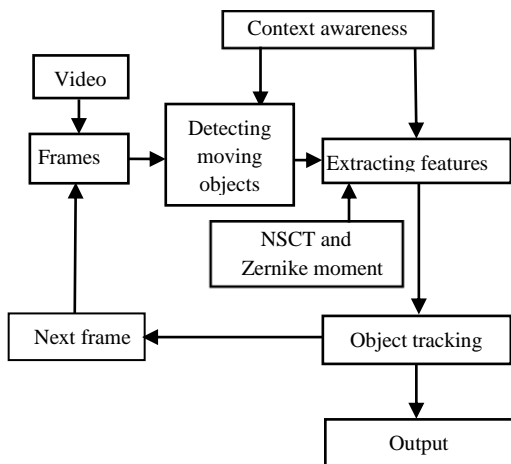


Fig. 1. The overall of proposed method

The proposed method has three main stages: moving object detection, features extraction and object tracking. Firstly, the input data is videos that have the same serial frames. The author detects moving objects. Secondly, the extracting features of object. In this stage, the author uses NSCT and zernike combined with context awareness to extract features. Finally, object tracking. The author uses a contour object for object tracking from frame to frame.

#### A. Moving object detection

Most of the moving object detection methods take two consecutive input images and return the locations where differences are identified. The motion of an object can cause these differences.

A video sequence contains a series of frames. Each frame can be considered as an image. The common approach for detection of objects consists of three steps: background modeling, foreground detection and data validation. The

author assumes there are only two modes for each pixel in a single frame: background and foreground. The basic of background subtraction method is to compare the frame background with a threshold  $T$  which the author is pre-defined. If the difference of a pixel is smaller than  $T$ , then it is background, otherwise, it is foreground. To detect objects, the NSCT coefficients and their statistical values were extracted as the features of object images. The author defines a discrete warped NSCT which goes across the region boundaries based on context awareness. The author computes the image sample values in each region of the partition and also describes its implementation together with the inverse resampling. A warped NSCT with a sub-band filtering along the flow lines is implemented. At the boundaries, warped NSCT still have two vanishing moments. The NSCT coefficients of a discrete image are computed with a filter bank. This method reduces computation time significantly by utilizing the characteristics of high correlation between adjacent frames. Because the data are highly correlated pixels in each frame, there are two possibilities for the NSCT element of the consecutive frame will be equal.

The algorithm uses a diagram to check the repetitive element NSCT between two consecutive frames depending on the context awareness in video, thereby reducing the frequency of calculation of the NSCT calculation. The results have showed that this method improved significantly reduces the computation time, and it goes beyond real-time requirements.

#### B. Feature Extraction

Most previous definitions of context are available that context awareness look at who's, where's, when's and what's of entities and use this information to determine why the situation is occurring. Here, the author's definition of context is:

“Context is any information that can be used to characterize the situation of an image such as: pixel, noise, strong edge, weak edge in image that is considered relevant to the interaction between pixels and pixels, including noise, weak and strong edge themselves.”

In video processing, if a piece of information can be used to characterize the situation of a participant in an interaction, then that information is context. Contextual information can be stored in feature maps on themselves. Contextual information is collected over a large part of the scene.

As [18], the author will be calculated two values as a feature vector: Aspect Ratio (AR) and zernike NSCT value for each bounding boxed object. AR is defined by:

$$AR = \text{width of bounding box} / \text{weight of the bounding box} \quad (6)$$

The bounding box of human usually has less in width than height, and the opposite for car. This feature may fail in cases that human sits down and his height is just a half of the standing pose or poses with two raising arms may increase the width of the bounding box and break the assumption. For the car, that is the perspective of the camera. If the camera view is in the same line of a moving car, it will show the longer

dimension in height not width. But in many of the outdoor cases, the assumption is true. The author chooses AR feature because it is simple, fast and its nature is completely different so that the union of failed case set of them is smaller.

Zernike NSCT value is a value that represents the contour property of objects and used to differentiate between a human and a car. It is calculated as the amplitude of zernike moment on contour binary images of the bounding boxed object as followings: first, the bounding boxed object image is contour detected by applying NSCT [18, 20] decomposition on it. The n levels parameter is [0, 1, 3]. This parameter is has 3 numbers, which means using 3 pyramidal levels (from coarser to finer scale). The first number – zero means at level 1 of pyramid, the level of directional filter bank decomposition will be 2 exponent zero to 1. It is the contour image received which is synthesized from two next levels [18]. The second level of pyramid is 2 exponents 1 to 2. It's  $\pi/2^1$  so the author may see it like the two images with horizontal and vertical ways of energy. Similarly, the third level is  $\pi/2^3$ , which is 8 images with rotation of energy. The synthesis of from the third through the second and to the first level gives us contour image with energy at all ways keeping. The number of levels and direction number at each level is chosen to be computationally efficient and good enough to reflex the contour of object. In my experiment, NSCT is quite slow, but three levels are also enough to do the job. Not all contour images are perfectly detected but it is good enough for the classification result. Detailed implementation of NSCT can be found in [20].

Second, the contour image is converted to the binary image based on threshold which is just simply the mean of the pixel values in image. This is just a preparatory step for zernike, which is done due to [18, 34] that the binary image with just contour point is faster and more accurate than the original image for zernike.

Third, the binary image is passed through zernike moment to get amplitude [18, 20]. Zernike moment is rotationally invariant, so it is suitable for characterizing contour of object, which may be changed because of the various activities and this property reduces effort in training many poses which are just the rotations of another. Detailed implementation of zernike and its amplitude can be found in [20].

### C. Object tracking

Tracking task is to identify an object in the current frame and the previous frame. There are many cases in the tracking task such as new objects appear, the object is obscured or disappear, and the object intertwines and then splits again. The tracking method is used optical flow [35] on the points of the contour of the object, to track the object through each frame.

To solve the problem of objects tracking, the author defines an equivalent coefficient  $p(m, m')$  between an object of the frame previous  $R_{prev}(m)$  and the object of the current frame  $R_{cur}(m')$  as follows [35]:

$$\{p_1, \dots, p_{n_i}\} \in S_i \quad (i = 1, \dots, m) \quad (7)$$

$$\{S_1, \dots, S_m\} \in R_{prev}(m) \quad (8)$$

and

$$\{p'_1, \dots, p'_{n'_j}\} \in S'_j \quad (j = 1, \dots, m') \quad (9)$$

$$\{S'_1, \dots, S'_{m'}\} \in R_{cur}(m') \quad (10)$$

The equivalent coefficient  $p(m, m')$  is defined as:

$$p(m, m') = \frac{\sum_{i=1}^m T_i}{m} \quad (11)$$

where  $T_i = \begin{cases} 1 & \text{if } S_i \text{ compatible with at least one } S'_j \\ 0 & \text{otherwise} \end{cases}$

Two contour  $S_j$  and  $S'_j$  called incompatible if the equivalent coefficient  $q(S_i, S'_j)$  met the following conditions:

$$q(S_i, S'_j) = \frac{\sum_{k=1}^{n_i} X_k}{n_i} > \beta \quad (12)$$

where,  $\beta$  is the threshold compatibility of the two contour with  $0 \leq \beta \leq 1$  and  $X_k = 1$  if  $p_k(x, y) \in S_i$  satisfy (13) and (14), otherwise  $X_k = 0$ .

$$\min_{p_i \in S_j} \{|x + u_k - x'_i|\} < \alpha_u \quad (13)$$

$$\min_{p_i \in S_j} \{|y + v_k - y'_i|\} < \alpha_v \quad (14)$$

where,  $(u_k, v_k)$  is the velocity vector of point  $p_k(x, y)$  obtained from the results of optical flow;  $(x'_i, y'_i)$  is the coordinates of points  $p'_i \in S'_j$  and  $\alpha_u, \alpha_v$  are the threshold constant [35].

After calculating the equivalent coefficient  $p(m, m')$ , two objects  $R_{prev}(m)$  and  $R_{cur}(m')$  are considered duplicate if

$$p(m, m') > \Omega \quad (15)$$

where  $0 \leq \Omega \leq 1$  is threshold constant.

The step of the tracking period as:

Initial:  
 $C_i$  is the contour of the  $i^{\text{th}}$  object in the previous frame.  
 $C_j$  is the contour of objects in the current frame number  $j$ .  
 $p(C_i, C_j)$  is the coefficient equivalents of  $C_i$  contour and  $C_j$  contour.  
 Step 1: take contour  $C_i$  and set  $n = 0$   
 Step 2: take contour  $C_j$   
     if  $p(C_i, C_j)$  satisfied of (15) then  
          $n = n + 1$ ;  
     else  
         if contour  $j$  is final contour then  
             if ( $n > 1$ )  
                  $C_i$  is separated into  $n$  object ( $C_j$ )  
                 goto step 3  
             if  $n = 1$  then  
                  $C_i$  coincides with  $C_j$   
             if  $n = 0$  then  
                 if  $C_j$  outside frame  
                     delete  $C_j$   
                 else

```

else
    goto step 2
Step 3: if contour  $i^{th}$  is final contour then
    Stop
else
    goto step 1
    
```

IV. EXPERIMENTS AND RESULTS

In this section, the author applies the procedure described in section 3 and achieved a superior performance in my experiments as demonstrated in this section. For performance evaluation, author compares the results of the proposed method with the methods such as Kernel filter method [14], Particle filter method [15], curvelet method [16] and contourlet method [17].

The author uses the standard dataset PETS dataset [36], CAVIAR dataset [37], SUN dataset [38], etc to experiment and evaluate. These video datasets are the video datasets in computer vision field. Hard thresholding is applied to the coefficients after decomposition in NSCT domain. For the tracking part, the author determines the object in each frame of the video. The object area is determined in the first frame; the tracking algorithm needs to track the object from frame to frame. Here, the author reports the results on some video clips. Figure 2 shows some scenes in PETS dataset. The experimental approach is as follows.

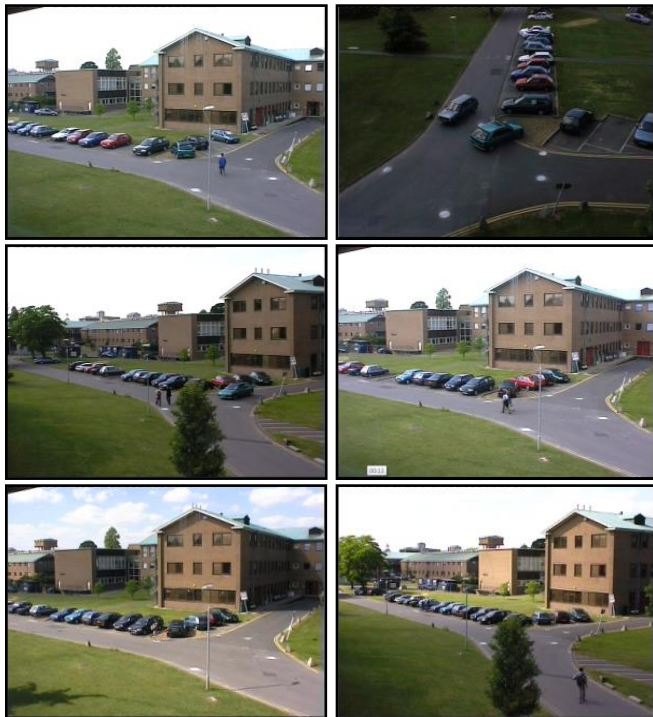


Fig. 2. Scenes in test dataset

The experiments are person video clips with the frame size 288 by 352. The proposed method processes this video clips at 28 frames/second. The author has experimented on the video up to 5000 frames. Here, the author reports the results up to 400 frames. Some results achieved as shown in figure 3.

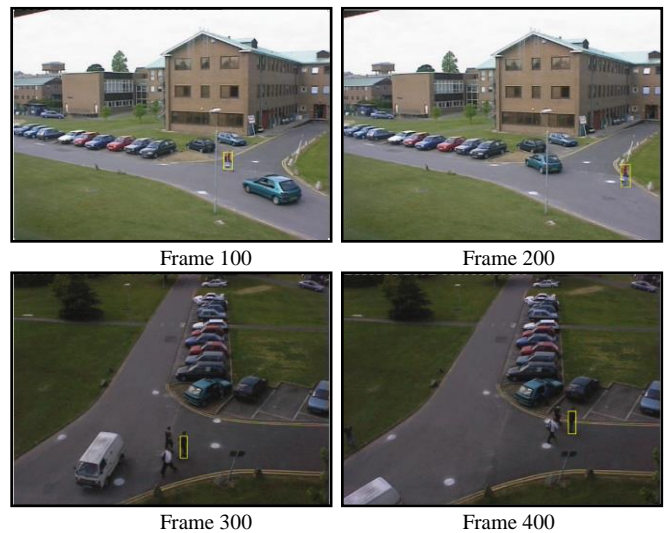


Fig. 3. Tracking in human video clips up to 400 frames

The other of the experiments is person video clips with the frame size 288 by 352. The proposed method processes this video clips at 25 frames/second. The author has experimented on the video up to 2000 frames. Here, the author reports the results up to 250 frames. Some results achieved as shown in figure 4.

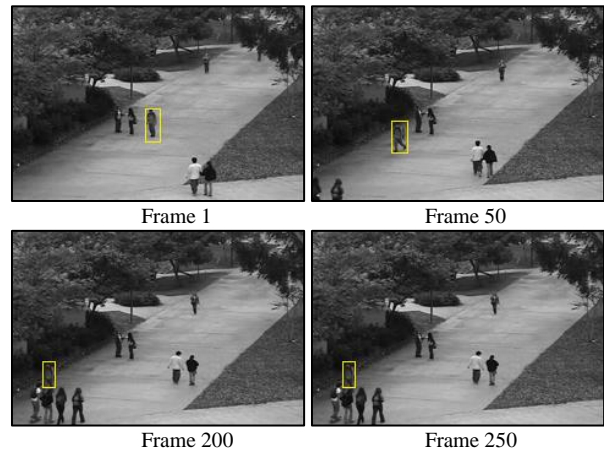


Fig. 4. Tracking in human video train clips up to 250 frames

In these figures, the author observes that the proposed method performs well.

TABLE I. COMPARING THE OBJECT TRACKING ERROR OF THE PROPOSED METHOD AND OTHER METHODS

Frame Number	Kernel method [14]	Partial method [15]	Curvelet tracking [16]	Contourlet Tracking [17]	Proposed method
100					
150					
200	Error				
250					
300		Error	Error		
350				Error	

Frame Number	Kernel method [14]	Partial method [15]	Curvelet tracking [16]	Contourlet Tracking [17]	Proposed method
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			

Table 1 compares the human tracking error between the proposed method and other methods. The error in this case means that the tracking algorithm could not track the correct object. In the table 1, the author sees that the number of errors of the proposed method is less than other methods.

Now, the author uses the Euclidean distance values to perform the evaluation of tracking. The Euclidean distance between the computed centroid  $(x_c, y_c)$  of tracked human window and the actual centroid  $(x_A, y_A)$  is followed:

$$K_c = \sqrt{(x_A - x_c)^2 + (y_A - y_c)^2} \quad (16)$$

Figure 5 shows the Euclidean distance of the proposed method and other tracking algorithms. 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. The ground truth values of centroid of the object were shown by the x-axis.

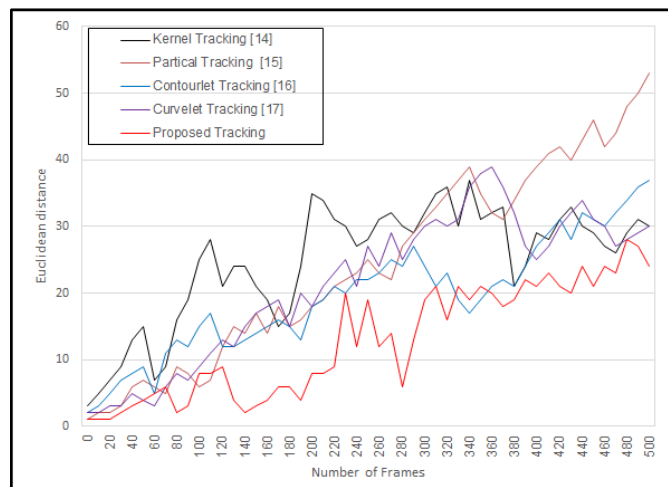


Fig. 5. The Euclidean distance of the proposed method and other tracking algorithms

## V. CONCLUSION

Object tracking is an important feature of intelligent vision systems. Tracking has many applications in intelligent surveillance systems such as in traffic system: counting the vehicles, measuring vehicle's velocity, and then it can detect faults about the velocity of vehicles automatically, etc. These tasks are hard because they depend on the context and environment. The results of the monitoring objects tremendously impact the analysis phase and identification of behavior. This paper proposes a new method to track the human based on their contour. The proposed method uses the amplitude of zernike moment on NSCT to track object depending on context awareness. The algorithm has also been processed which successfully such cases as the new object detection, object detection obscured after they reappeared, detecting and tracking objects successfully intertwined and then separated again. In the future work, the author should compare equivalent geometry contours of the object shape. This equivalent does not depend on the direction of motion, or the location of the object. So the objects are still tracked properly in the above case.

## ACKNOWLEDGMENT

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2015-20-08.

## 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, 13(4), pp 310–324, 2003
- [2]. Liu T. L., Chen H. T., "Real-time tracking using trust region methods", IEEE transactions on Pattern Analysis and Machine Intelligence, 26(3), pp 397–402, 2004
- [3]. Wang D., "Unsupervised video segmentation based on watersheds and temporal tracking", IEEE Transactions on Circuits and Systems for Video Technology, 8(5), pp 539–546, 1998
- [4]. Zhou S. K., Chellappa R., Moghaddam B., "Visual tracking and recognition using appearance-adaptive models in particle filters", IEEE Transactions on Image Processing, 13(11), pp 1491–1506, 2004
- [5]. Comaniciu D., Ramesh V. , Meer P., "Kernel-based object tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), pp 564–577, 2003
- [6]. Comaniciu D., Meer P., "Mean shift: a robust approach toward feature space analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp 603–619, 2002
- [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
- [8]. Yilmaz A., Javed O., "Shah M., Object Tracking: A survey", ACM Computing Surveys, 38(4), 2006
- [9]. Wang D., "Unsupervised video segmentation based on watersheds and temporal tracking", IEEE Transactions on Circuits and Systems for Video Technology, 8(5), pp 539–546, 1998
- [10]. Moeslund T.B., Granum E., "A survey of computer vision based human motion capture", Computer Vision and Image Understanding, 81, pp 231-268, 2001
- [11]. Ruolin Zhang, Jian Ding, Object tracking and detecting based on adaptive background subtraction, International workshop on information and electronics engineering, Vol. 29, pp1351–1355, 2012
- [12]. Wei Shuigen Chen Zhen, Dong Hua, Motion detection based on temporal difference method and optical flow field, Second international symposium on electronic commerce and security, Vol. 2, pp 85-88, 2009

- [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
- [14]. Comaniciu D., Ramesh V., Meer P., "Kernel based object tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), pp 564-577, 2003.
- [15]. Nummiaro K., Meier E. K., Gool L.J.V., "An Adaptive Color-based Particle Filter", Image and Vision Computing, 21(1), pp 99-110, 2003
- [16]. 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
- [17]. Nguyen Thanh Binh, T.A.Dien, "Object Detection and Tracking in Contourlet Domain", Context-Aware Systems and Applications, Vol 109, 192-200, 2013
- [18]. Luu The Phuong, Nguyen Thanh Binh, "Human Object Classification Based on Nonsampled Contourlet Transform Combined with Zernike Moment", 4th EAI International Conference on Context-Aware Systems and Applications, Vietnam, November, 2015
- [19]. Do, M.N., Vetterli, M., "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation". IEEE Transactions on Image Processing, Vol. 14, Issue 12, pp. 2091-2106, 2005
- [20]. Cunha, L.D., Zhou, J., Do, M.N., "The Nonsampled Contourlet Transform: Theory, Design, and Applications". IEEE Transactions on Image Processing, Vol. 15, No. 10, pp. 3089-3101, 2006
- [21]. Graps, A., "An Introduction to Wavelets". IEEE Computational Science and Engineering, Vol. 2, No. 2, pp. 50-61, 1995
- [22]. Ma, J., Plonka, G., "The Curvelet Transform a Review of recent Applications". IEEE Signal Processing Magazine, Vol. 27, Issue 2, pp. 118-133, 2010
- [23]. Ping Jiang, HaoSha. "Image De-Nosing Based on Non-Subsampled Contourlet Transform Domain in Multi-Bessel K Form Model". Research Journal of Applied Sciences, Engineering and Technology, 6(18), 3400 - 3403, 2013
- [24]. 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, Lecture Notes in Computer Science, pp 87-101, 2014
- [25]. Teague, M.: "Image analysis via the general theory of moments". Journal of Optical Society of America. 70(8), pp 920-930, 1980.
- [26]. Hwang, S. K., Kim, W. Y., "A Novel approach to the fast computation of Zernike moments". Pattern Recognition. 39(11), pp 2065-2076, 2006.
- [27]. Celebi, E. M., Aslandogan, Y. A., "A comparative study of three moment based shape descriptors". In proceeding of International Conference on Information Technology: Coding and Computing, Vol. I, pp. 788-793, 2005.
- [28]. Chong, C. W., Raveendran, P., Mukundan, R., "Translation invariance of Zernike moments". Pattern Recognition, 36(8), pp 1765-1773, 2003.
- [29]. Bin, Y., Xiong, P. J., "Invariance analysis of improved Zernike moments". Journal of Optics A: Pure and Applied Optics. 4(6), pp 606-614, 2002.
- [30]. Papakostas, G. A., Boutalis, Y. S., Karras, D. A., Mertzios, B. G., "A new class of Zernike moments for computer vision applications". Information Sciences. 177(13), pp 2802-2819, 2007.
- [31]. Farzem, M., Shirani, S., "A robust multimedia watermarking technique using Zernike transform". In proceeding of fourth IEEE workshop on Multimedia Signal Processing, pp. 529-534, 2001.
- [32]. Zhenjiang, M., "Zernike moment based image shape analysis and its application". Pattern Recognition Letters. 21(2), pp 169-177, 2000.
- [33]. Khare, M., Srivastava, R. K., Khare, A.: Moving Object Segmentation in Daubechies Complex Wavelet Domain, Signal Image and Video Processing, Doi: 10.1007/s11760-013-0496-5, pp. 635-650, 2013.
- [34]. Mukundan, R.: A Contour Integration Method for the Computation of Zernike Moments of a Binary Image. USM-Penang: National Conference on Research and Development in Computer Science and Applications - REDECS'97, pp. 188-192, 1997
- [35]. Takashi Fuse, Eihan Shimizu, Morito Tsutsumi "A Compare Study on Gradient-Based Approaches for Optical Flow Estimation", International Archives of Photogrammetry and Remote Sensing, Vol XXXIII, Part B5, 2000.
- [36]. PESTS2009 (Performance Evaluation of Tracking and Surveillance): <http://ftp.pets.rdg.ac.uk/pub/PETS2009/>
- [37]. CAVIAR (Context Aware Vision using Image-based Active Recognition) : <http://groups.inf.ed.ac.uk/vision/CAVIAR/>
- [38]. SUN (Scene UNDERstanding): <http://vision.princeton.edu/projects/2010/SUN>