Human Object Detection based on Context Awareness in the Surroundings

Nguyen Thanh Binh[1,*]

[1]Faculty of Computer Science and Engineering,   Ho Chi Minh City University of Technology, Vietnam

Abstract

Surveillance system has been applied in providing public security for many complex places like railway stations, bus stops, etc. In most cases, human object detection is an important task in surveillance system. In the case that human objects are occlusion or outdoor environment, human objects detection is a challenging problem. In this paper, we propose a method to implement for human object detection based on context awareness in new wavelet generation domain in outdoor environment. We use curvelet transform based on context awareness combined with support vector machines as a classifier for human detection. The proposed method was tested on a standard dataset like PEST2001 dataset. For demonstrating the superiority of the proposed method, we have compared the results with the other recent methods available in literature.

1. Introduction

Surveillance system has been applied for providing public security in many complex places like railway stations, bus stops, etc. However, most of the advancements in computer vision are less to apply in actual implementation of surveillance system. Human object detection is an important task in surveillance system. In the case that the human objects are occlusion or outdoor environment, human objects detection is a challenging problem. Human object detection algorithm must work under real-time constraints, natural conditions, different sizes of human objects, etc [10, 23]. Feature selection and machine learning are the key components in any detection algorithm. Most object detection algorithms developed base on machine learning methods [3].

In the applications of computer vision, the problem is how we can know which objects are moving, and which objects are background. There are many methods for solving this task such as: temporal median filter, mixture of Gaussian, kernel density estimation, statistical methods, temporal differencing method, etc. The major problem existed in these methods is the computational cost as well as high memory requirements.

Temporal median method involves calculating the median value of the previous frame in a video sequence to establish a background model for background subtraction methods. The drawback of the method temporal median filter is the relatively low accuracy. To overcome this drawback, the method proposed temporal median filters the solutions that build the background model. Then this

---

, Corresponding Author. E-mail:"pvdkpj B j eo w…Qgf w0xp"

background model is used to find out the foreground. This background model was built by learning through n consecutive frames. Pixel value at position (x, y) of the background model is built by selecting the median of n frame at position (x, y). Background subtraction methods detect moving object from the difference between the current image and a background image model.

In the past, many algorithms have been built for object detection. Lowe [4] used scale invariant feature transform as a feature descriptor for object recognition. Lu [5] proposed a visual feature for object classification based on binary pattern. Dalal [6] proposed histogram of oriented gradient (HoG) as a feature descriptor for object detection. Cao [7] proposed a method by extending the HoG to boosting HoG feature. All the methods discussed above have local advantages or disadvantages depending on the features they have used [10, 23].

Yu [8] proposed wavelet method for visual classification. This method uses real valued discrete wavelet transform. Real valued wavelet transform has three major problems: lack of shift sensitivity, poor directionality and lack of strong edge detection [11]. This drawback affects the process of feature selection. To increase the ability to identify objects, we use curvelet transform to overcome these problems.

In this paper, we propose a method to implement for human object detection based on context awareness in new wavelet generation domain in an outdoor environment. We use curvelet transform based on context awareness combined with support vector machines as a classifier for human detection. The proposed method was tested on a standard dataset like PEST2001 dataset. For demonstrating the superiority of the proposed method, we have compared the results with the other recent methods by Lu [5] and Renno [9]. We use three different performance metrics: average classification accuracy, true positive rate (recall), and predicted positive rate (precision) for this comparison. The rest of this paper is organized as follows: in section 2, we described the basic of curvelet transform and its advantages for human detection. And details of the feature selection, support vector machine classifier for human detection are presented in section 3. In the section 4, we proposed the method for human object detection based on context aware in curvelet domain. The results of proposed method are given in section 5 and conclusions in section 6.

2. Curvelet transform and its advantages for human detection

2.1. Curvelet transform

Real valued wavelet transform suffers from three major problems: lack of shift sensitivity, poor directionality and lack of strong edge detection. E Candès [17] proposed a solution to overcome these problems by curvelet transform

(CT). In this section we explain what curvelets are, how they are constructed, and what their main properties are.

Curvelets are basically 2D anisotropic extensions to wavelets that have a direction associated with them. Similar to wavelets, curvelets can be translated and dilated. The anisotropic scaling relation is a key difference between wavelets and curvelets. The parabolic scaling is also a key ingredient to prove that curvelets remain localized in phase-space under the action of the wave operator provided the medium is smoothed appropriately prior to propagation [6].

The idea of curvelets [11, 17] is to represent a curve as a superposition of functions of various lengths and widths obeying the scaling law $width \approx length^2$. This can be done by first decomposing the image into subbands, i.e separating the object into a series of disjoint scales; then, each scale is analyzed by means of a local ridgelet transform.

Curvelets are based on multiscale ridgelets combined with a spatial bandpass filtering operation to isolate different scales. The bandpass is set so that the curvelet length and width at fine scales are related by the scaling law $width \approx length^2$ and so the anisotropy increases with decreasing scale like a power law. There is a very special relationship between the depth of the multiscale pyramid and the index of the dyadic subbands. The side length of the localizing windows is doubled at every other dyadic subband, hence maintaining the fundamental property of the curvelet transform which says that elements of length about $2^{-j/2}$ serve for the analysis and synthesis of the jth subband $[2^j, 2^{j+1}]$.

Like ridgelets, curvelets occur at all scales, locations, and orientations. However, while ridgelets have global length and variable widths, curvelets in addition to a variable width have a variable length and so a variable anisotropy does.

The length and width at fine scales are related by the scaling law $width \approx length^2$ and so the anisotropy increases with decreasing scale like a power law. Recent work [17] shows that the thresholding of discrete curvelet coefficients provided near optimal N-term representations of otherwise smooth objects with discontinuities C^2 along curves.

The curvelet dictionary is a subset of the multiscale ridgelet dictionary, which allows reconstruction. The "à trous" subband filtering algorithm [11, 17] is especially well-adapted to the needs of the digital curvelet transform. The algorithm decomposes an n by n image $f(x, y)$ as a superposition of the form

$$f(x, y) = c_J(x, y) + \sum_{j=1}^{J} w_j(x, y) \qquad (1)$$

where c_J is a coarse or smooth of the original image $f(x, y)$ and w_j represents the details of Im at scale 2^{-j}.

In wavelet theory, one uses a decomposition into dyadic subbands $[2^j, 2^{j+1}]$. The basic process of the digital realization for curvelet transform is given as follows [17]:

(1) *Subband Decomposition.* We define a bank of filters P_0, $(\Delta_s, s \geq 0)$. The image f is filtered into subbands with à trous algorithm

$$f \rightarrow (P_0 f, \Delta_1 f, \Delta_2 f, \ldots \ldots) \qquad (2)$$

The different subbands $\Delta_s f$ contain details about 2^{-2s} wide.

(2) *Smooth Partitioning*. Each subband is smoothly windowed into "squares" of an appropriate scale.

$$\Delta_s f \rightarrow (w_Q \Delta_s f)_{Q \in Q_s} \qquad (3)$$

where w_Q is a collection of smooth window localized around dyadic squares.

$$Q = [k_1 / 2^s, (k_1 + 1) / 2^s] \times [k_2 / 2^s, (k_2 + 1) / 2^s] \qquad (4)$$

(3) *Renormalization*. Each resulting square is renormalized to unit scale

$$g_Q = (T_Q)^{-1}(w_Q \Delta_s f), \qquad Q \in Q_s \qquad (5)$$

where $(T_Q f)(x_1, x_2) = 2^s f(2^s x_1 - k_1, 2^s x_2 - k_2)$ is a renormalization operator.

(4) *Ridgelet Analysis*. Each square is analyzed in the orthonormal ridgelet system. This is a system of basis elements p_λ making an orthonormal basis for $L^2(R^2)$:

$$\alpha_\mu = \langle g_Q, p_\lambda \rangle \qquad (6)$$

We see that the performance of vehicle tracking will increase if the correct feature is selected for tracking algorithm. In our proposed work, we have used curvelet coefficients as a feature set.

Human object detection is a problem where the objects may present in translated as well as rotated form among different scenes. Curvelet transform has the time-frequency-localization and multiscale properties of wavelets. It offers a high degree of directionality and anisotropy. Therefore, the properties of curvelet transform will be useful for human detection.

2.2 Advantages of curvelet transform for human detection

Curvelet transform is useful for human detection due to its following properties:

The curvelet transform is a multiscale transform with frame elements indexed by location, scale and orientation parameters, and have time-frequency localization properties of wavelets but also shows a very high degree of directionality and anisotropy. Curvelets provide optimally sparse representations of objects which display curve-punctuated smoothness except for discontinuity along a general curve with bounded curvature. Such representations are nearly as sparse as if the object were not singular and turn out to be far sparser than the wavelet decomposition of the object.

Boundary curvelets that are aligned with a boundary edge mostly respond to the artificial discontinuity created by periodization. Boundary curvelets misaligned with respect to the boundary edges are assigned big coefficients when they respond to geometrical structure on the opposite side of the image, across the edge. The curvelet coefficients are directly calculated in the Fourier space. In the context of the ridgelet transform, this allows avoiding the computation of the 1-D inverse Fourier transform along each radial line.

3. Feature Selection and support vector machine classifier for human detection

3.1 Feature Selection

Feature selection is to select a subset of input variables with no predictive information by eliminating features. It can significantly improve the comprehensibility of the resulting classifier models. A feature is a function of one or more measurements computed so that it quantifies some significant characteristics of objects [15]. In any object detection algorithm, the selection of appropriate feature is very important. We see that the performance of detection will increase if the correct feature is selected for detection algorithm. In our proposed work for human, we have used curvelet transform coefficients as a feature set. We have taken combination of two different features – curvelet transform and support vector machines. A brief description of these two features and why they are useful for human object detection are described in subsection 2.2 and 3.2 respectively.

3.2 Support vector machine classifier for human detection

Support vector machines (SVM) include associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis in machine learning. SVM can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.

SVM is a popular classifier. The classifier objects are into two categories: object and non-object data [13]. In here, we detect two types: human and car object.

An n-dimensional object x has n-coordinates.

$$x = (x_1, x_2, x_3, \ldots, x_n),$$

where, each $x_i \in R$ for i=1, 2, 3,….,n .

Each object x_j belongs to a class $y_j \in \{-1, +1\}$. Consider a training set T of m patterns together with their classes,

$$T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$$

and a dot product space S, in which the objects are embedded: $x_1, x_2, \ldots, x_m \in S$.

Any hyperplane in the space S can be written as:

$$\{x \in S \mid w.x + b = 0\}, w \in S, b \in R$$

The dot product w.x is defined by [10]:

If there exists at least one linear classifier defined by the pair (w, b) which correctly classifies all objects as shown in Figure 1 then a training set of objects is linearly separable [10]. The linear classifier is represented by the hyperplane H (w.x + b = 0) and defines a region for class +1 and another region for class -1 objects (figure 1).
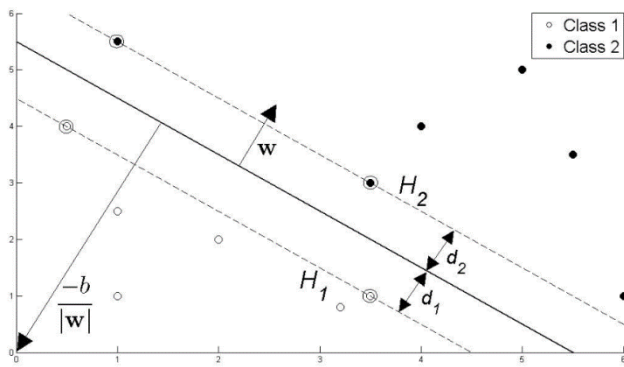
Figure 1: Linear classifier [10] defined by the hyperplane H

After training, the classifier is ready to predict the class membership for new objects, different from those used in training. The class of object x_k is determined with the equation [10]:

$$class(x_k) = \begin{cases} +1 & if \quad w.x_k + b > 0 \\ -1 & if \quad w.x_k + b < 0 \end{cases}$$

4. The method for human object detection based on context aware in new wavelet generation domain.

In this section, we propose a method for human object detection. Object detection is hard work. In the past, there are many methods for this work. Every method has particular strength and drawback depending on the scenes.

Most of the last methods only used a feature. This work is not efficient because it is difficult to use simple feature to present the difference types of object such as: shape object, moving object, etc. To increasing the accuracy and performance of object classification, we can use many object features combined together. In here, we used curvelet filter based on context awareness and combined shape-based features with motion-based features method for human object detection. Most previous definitions of context are available [18] 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, our 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. These maps can encode high-level semantic features or low-level scene features. The low-level features are image gradients, texture descriptors, and shape descriptors information [19].

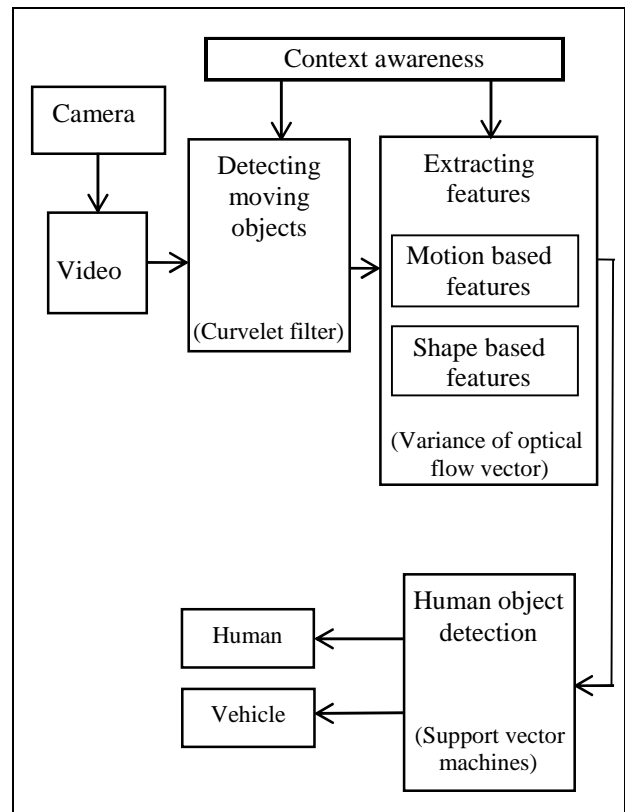A contextual feature vector can be extracted for each object in a training set.

Figure 2. The overall of the proposed method for object detection

The overall of the proposed method for object detection is described as figure 2. In the figure 2, the proposed method for human objects detection includes three periods: Firstly, detecting moving object. In this period, we use curvelet filter combined with context awareness for detecting moving objects. After that, we extract the objects from the scenes. These objects are saved as the blob. Secondly, extracting the features. We calculate the features which depend on aspect ratio, dispersedness and motion-based features such as variance of optical flow vector and context awareness. Finally, classifying the objects. We apply support vector machines to training the classifier in two classes which have been defined: the human and vehicle.

4.1 Detecting moving objects

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 steps of this period are described as figure 3:
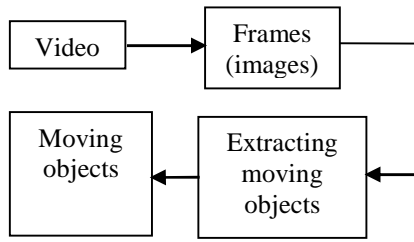
Figure 3: The process of the detecting moving object step.

We assume 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 we are pre-defined. If the difference of a pixel is smaller than T, then it is background, otherwise, it is foreground. To detect objects, the curvelet coefficients and their statistical values were extracted as the features of object images. We define a discrete warped curvelet transform which goes across the region boundaries based on context awareness. We compute the image sample values in each region of the partition and also describe its implementation together with the inverse resampling. A warped curvelet transform with a sub-band filtering along the flow lines is implemented. At the boundaries, warped curvelet still have two vanishing moments. The curvelet coefficients of a discrete image are computed with a filter bank [20]. 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 curvelet element of the consecutive frame will be equal.

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

4.2 Extracting features

The features can base on the shapes, colors, texture and motion. The steps of extracting features are described as figure 4:
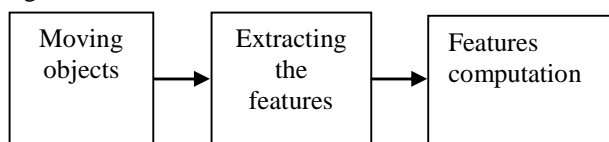

Figure 4: The steps of extracting features

After extracting moving objects, we extract the features of objects. In here, we use two features: the shape-based features and the motion-based features

4.2.1 The shape-based feature

We calculate the bounding rectangle. There are minimum rectangle which contain the object area. We save the location coordinates of the left upper and right lower position, the shape-based features are extracted as follows [21]:

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

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

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

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

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

The above shape-based features are useful feature to discriminate a human from a car. Because the human has the more complex shape then they will have a greater dispersion medium.

4.2.2 Motion-based feature

The objects in the context are the dynamic objects. The behavior of the object must be determinated based on the features of the object. For example, the human behavior and the behavior of the car are not the same. Here, we use variance of optical flow vectors. The main idea of the method is to use optical flow vector direction of moving objects over time. Each pixel in a blob will correspond to each vector. 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 into the body parts to analyze the motion of an object. Obviously, the human will have symmetrical limb movement while the movement of the whole body of the car is not.

Optical flow methods are used to distinguish which objects are flexible and not flexible (rigid and non-rigid). We see that: a moving of car is hard, and not flexible while the human walk action is soft, and mobile. The flexible objects as human [22] have the parts like arms, legs, etc. moving in different directions so each angle greater than the average vectors. Therefore, the feature average gradient G of human is higher average gradient G of vehicle. Because the vehicle is not flexible, they will have typical low gradient G. Because of the vehicles contain pixels nearly the same. The motion vectors are the same then angle smaller than average vectors (nearly 0). Accordingly, the characteristic gradient G, generated by the movement of human, will be cyclical. Moreover, depending on the context awareness in the frame, we can use this method, the moving of people can be distinguished from other objects such as vehicle.

4.3 Human object detection

After feature extraction for positive and negative datasets, we will train for the SVM classifier. SVM is a

supervised learning algorithm which is used for data classification. SVM is very effective to solve large data dimensionality and solve overfitting problem very well such as data contain noise and displaced groups, etc. The steps of objects detection are described as figure 5.
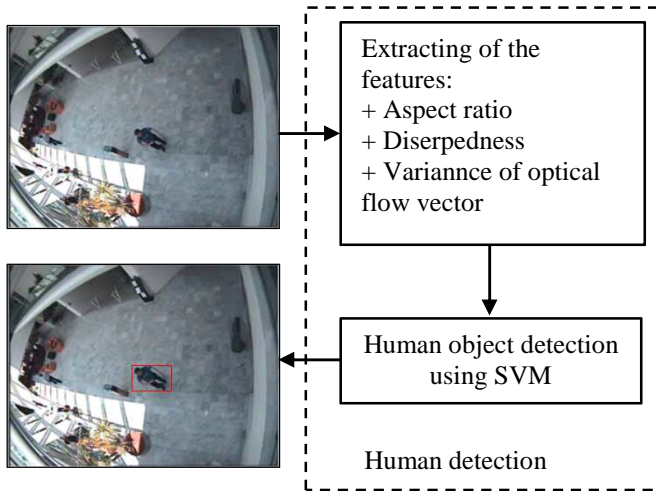

Figure 5: The process of objects detection using SVM and features based on shape and motion

The main steps in the classification of moving objects will be implemented as follows:

(i) Training for SVM classifier: from the training dataset containing positive and negative images of the object, we extract the features based on the shapes (aspect ratio, dispersedness), features based on motion (optical flow vector) combine with context aware, and from there to train SVM classifier.

(ii) Object Classification: from the detected object, we extract the above features and use SVM classifier to determine whether the object is the vehicle or the human.

4.3.1 Training for SVM classification

For performing classification using SVM, initially, we must build the model training.

Input:

+ The feature vector of objects in the training dataset. In here, we use 1420 x 3 matrix, where, 1420 is the number of feature vectors in the training dataset, 3 is the characteristic of the vector.

+ Volume label / class for each feature vector of the training dataset.

+ The parameters for SVM model: C, γ (the parameter of the kernel function, usually a Gaussian function)

Output:

+ Model SVM (Support Vectors, Lagrange multipliers a, parameter b).

The data of this training consist of two files: a data file containing the images of the human and the data set containing the image of the car. The data set includes a total of over 1420 images. This dataset is created by cutting the area as human and vehicles in videos for training. Here,

some images in the collection of data are as follows in figure 6.

Figure 6: Some images [14] in the collection of data.

For each dataset, we will extract the features based on shapes and motion as the above mention. The process of implementing the training is conducted as follows: select each rectangle which surrounds the objects detected. The result is a red rectangle, as shown below (figure 7), we will proceed to extract the feature vectors in this blob.

Figure 7: Rectangle surrounding the object detected

The feature extraction is carried out as follows:
(i) Adjust the threshold parameter of optical flow:
+ Human (Optical_Flow_Ths = 0.9) and vehicles (Optical_Flow_Ths = 1.2)
+ Two thresholds value are used to eliminate small motions due to noise. So, the threshold to choose depends on the outside environments (context aware), the different part of video.
(ii) Extracts features: aspect ratio, dispersedness, gradient G in the blob.
After the training phase, we obtain the SVM model.

4.3.2 Human object detection

After the training step for the SVM, we use this model to conduct human object detection.

Input:

+ Vector feature of the object to classify.

+ SVM model

Output:

+ Label / class of the object to be classified.

After extracting the feature vector of the object, we put this feature into the SVM model to perform object detection. If the label SVM detection is 1, we will conclude that the object is vehicle. Otherwise there label is -1, we will conclude that the object is the human. The results of the human object detection are displayed in white letters in the red rectangle (figure 8).

Figure 8: The result of human object classification

5. Experimental and Evaluation

In this section, we apply the procedure described in section 4 and achieved a superior performance in our experiments as demonstrated in this section. For performance evaluation, we compare the results of the proposed method with the methods: method proposed by Lu [5] and Renno [9].

We use the standard dataset PETS 2001 to experiment and evaluate. PESTS2001 is the video dataset in computer vision field. The dataset is divided into two groups: Training and Testing. Here, we use two video in Training1 (PETS D1Trai1.avi and PETS D1Trai2.avi) to trainning. Every video training has length 181s. And using 4 video in Testing 1 (PETS D1Tes1.avi, PETS D1Tes2.avi) and Testing 2 (PETS D2Tes1.avi, PETS D2Tes2.avi) to human detection. Every video testing has length 162s. These videos have different views of the same scene surveillance. Two training videos in Training 1 set the camera angle in two different locations. Some training video presents as figure 9.


Figure 9: Some training videos with different views of the same scene surveillance [14].

Some video testing in Testing1 and Testing2 set with the camera angles in four different positions (figure 10).

Figure 10: Some videos testing with the camera angles in four different positions [14].

We test on the data set with the different camera angles, so it may review all case objects with different shapes. The video test includes multiple videos with many different contexts as figure 11.
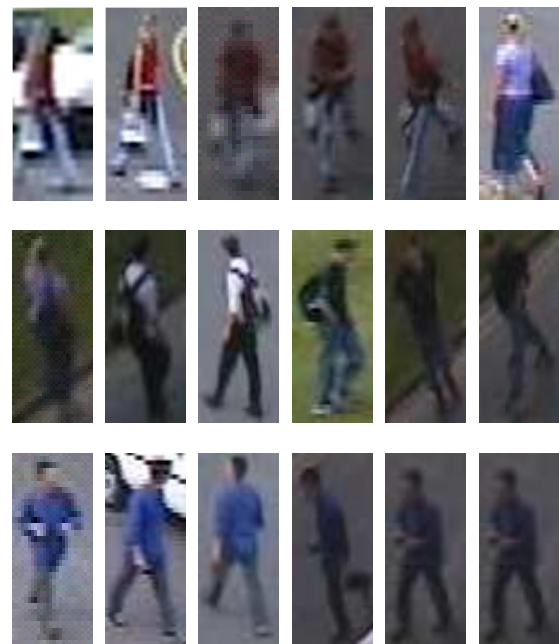
a.   Human object

b.   Car object

Figure 11: Some objects with different observation angles [14]

The program is implemented in Matlab. We use curvelet filters to detect motion objects with soft threshold. The time processing is low.

Table 1. The result of human detection in real scene

Video	The number human in real scene	The number human detection in real scene	Precision
D1Trai1	31	30	96,78%
D1Trai2	34	35	97.08%

The results of human detection in real scene present in table 1. There are the results of detecting moving objects with training videos PETS D1Trai1 and PETS D1Trai2. The results obtained in the detection of moving objects is relatively good. All objects are detected. However, some cases the detection of the object will be incorrect such as: for some objects move under the influence of the environment: air, light, etc.. may be false such as: the trees have leaves flickering, light changed suddenly,..etc. Moreover, the system is also difficult to detect objects very far from surveillance cameras. Because these objects are little change in motion,...

The different performance metrics, such as Average classification accuracy (ACA), True positive rate (TPR) (Recall) and Predicted positive rate (PPR) (Precision), are dependent on four values: True Positive (TP), TN (True Negative), FP (False Positive) and False Negative (FN), where [10]:

+ TP is the number of images, which are originally positive images and classified as positive images.

+ TN is the number of images, which are originally negative images and classified as negative images.

+ FP is the number of images, which are originally negative images and classified as positive images.

+ FN is the number of images, which are originally positive images and classified as negative images.

All above three performance metrics are defined in [10]. In here, we review parameters following:

+ ACA is defined as the proportion of the total number of prediction that was correct: $ACA = \dfrac{TP + TN}{TP + TN + FP + FN}$

+ TPR is defined as the proportion of positive cases that were correctly classified as positive:

$$TPR\,(Recall) = \dfrac{TP}{FP + FN}$$

+ PPR is defined as the proportion of the predicted positive cases that were correct: $PPR(Precision) = \dfrac{TP}{FP + TP}$

Table 2: The results of the classification of objects by SVM classifier for datasets

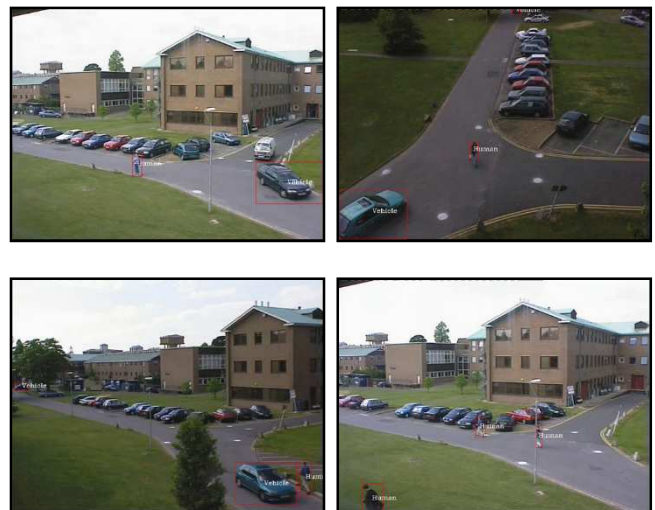| Video	GroundTruth		Human classification		Vehicle classification	
	Human	*Vehicles*	*True positives*	*False positives*	*True positives*	*False positives*
D1Tes1	9	3	8	1	3	1
Accuracy	91,66%					
D1Tes2	11	3	6	3	7	0
Accuracy	92,85%					
D2Tes1	10	5	7	2	9	4
Accuracy	93,33%					
D2Tes2	13	3	12	1	3	1
Accuracy	93.75%					
Tes3	31	4	28	1	6	0
Accuracy	97,14%					

Figure 12: Some results of classification

Tes3 is scenes shot from cameras on the streets. As the above mention, the paper classifies two objects: human and car. In comparison, table 2 presents object detection in test dataset. In the table 2, Groundtruth of the dataset and result of classifies for human and vehicles include: true positives and false positives.

The measurement accuracy is defined as follows:

$$\text{Accuracy} = \frac{\text{Total number of True Positives}}{\text{Total number of Groundtruth}}$$

Figure 12 presents some results of classification.

Values of all above performance measures have been shown in Table 3, for the proposed method and all other state-of-the-art methods.

Table 3: Values of Performance Measures of the other state-of-the-art methods

Methods proposed	TPR (Recall) (%)	TNR (%)	FPR (%)	FNR (%)	PPR (Precision) (%)	Average Accuracy (%)
Lu [5]	90	82	15	12	90.5	91.25
Renno [9]	90.5	83	20	08	91.2	93.45
Wavelet transform as a feature	85	76	17	6	82.13	81.21
Curvelet as a feature	92	89	16	11	92.15	92.50
Curvelet combined with context aware as a feature	94	91	16	13	94.29	96.12

The experiments are performed with SVM, classification results obtained from other state-of-art methods are given in Table 3. From Table 3, other experiment, we observe that the proposed method yields better performance than the other state-of-the-art methods for human object classification.

In summary, the classification of objects in video images is a relatively difficult problem. Each method has advantages and disadvantages to suit each context and different problems. To achieve good results, we need to analyze the problem, the specific circumstances and choose the appropriate method. The context awareness was applied to process detecting moving objects and object features more accurate. Therefore, they make the results of the process classification higher than the other proposed methods.

6. Conclusions

In the present work, our aim is to classify objects into two types of classes: human and car. We develop a method for object classification in real scenes using curvelet transform as a feature set. Curvelet allows for a different number of directions at each scale and aspect ratios. This feature allows an efficient curvelet to have based approximation of a smooth contour at multiple resolutions. Human and car object classification is a problem where the objects may present in translated as well as rotated form among different scenes. Curvelet transform has the time-frequency-localization and multiscale properties of wavelets. It offers a high degree of directionality and anisotropy. Therefore, the properties of curvelet transform will be useful for classification of human and car objects.

In this paper, we apply the SVM classifier with features based on shape combined with motion-based features. Classification results have showed high accuracy. However, the processing speed is rather slow due to the cost in the calculation of the characteristics, and training is relatively large. For the SVM model is accurate and efficient, it depends a lot on the training data set. This dataset is large enough to be collected objectively. Another important thing is that the result of this method depends largely on the steps of moving object detection. Thus in the case where the object is detected, the error object classification will be no longer accurate.

The proposed approach first trains SVM classifier by using curvelet coefficients of data as a feature set and then classifies testing data into one of the two categories: human and car objects. The proposed method is compared with other methods proposed by Lu [5] and Renno [9]. Experiments show that the proposed method gives better classification results at higher levels of curvelet transform and provides better results than other methods. The proposed method can detect human objects in a complex background.

References

1. Hu, W.Tan, T., A Survey on Visual Surveillance of Object Motion and Behaviors, IEEE Transaction on System, Man, and Cybernetics, 34 (3), pp. 334-352, 2006.
2. Zheng Song, Qiang Chen, Zhongyang Huang, Yang Hua, Shuicheng Yan, Contextualizing object detection and classification, IEEE conference on computer vision and pattern recognition, pp 1585 – 1592, 2011
3. Wang, L., Hu, W., Tan, T., Recent Developments in Human Motion Analysis, Pattern Recognition, 36(3), pp. 585-601, 2003.
4. Lowe, D. - Object recognition from local scale invariant features, in proceeding of 7th IEEE International Conference on Computer Vision, pp. 1150-1157,1999.

5. Lu, H., Zheng, Z., Two novel real-time local visual features for omnidirectional vision, Pattern Recognition. 43 (12), pp. 3938-3949, 2010.

6. Dalal N., Triggs, B. - Histograms of oriented gradients for human detection, in proceeding of IEEE International Conference on Computer vision and Pattern Recognition, pp. 886–893, 2005.

7. Cao, X., Wu, C., Yan, P., Li, X., Linear SVM Classification Using Boosting HoG Features for Vehicle Detection in Low-Altitude Airborne Videos, in proceeding of IEEE International Conference on Image Processing, pp. 2421-2424, 2011.

8. Yu, G., Slotine, J. J., Fast Wavelet-Based Visual Classification, in proceeding of IEEE International Conference on Pattern Recognition (ICPR), pp. 1-5, 2008.

9. Renno, J. P., Makris, D., Jones, G. A., Object classification in visual surveillance using Adaboost, in proceeding of IEEE International conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

10. M.Khare, Nguyen Thanh Binh, R.K.Srivastava, "Dual tree complex wavelet transform based human object classification using support vector machine", Journal of Science and Technology, Vietnam Academy of Science and Technology, Vol. 51, No.4B, pp. 134 -142, 2013

11. Nguyen Thanh Binh, A.Khare, Image Denoising, Deblurring and Object Tracking, A new Generation wavelet based approach, LAP LAMBERT Academic Publishing, 2013.

12. M. N. Do, M. Vetterli, The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, IEEE Transactions on Image Processing, vol. 14, pp. 2091–2106, 2005.

13. Noble, W. S, What is Support Vector Machine, Nature Biotechnology, 24(12), pp. 1565-1567, 2006.

14. PEST2001 Dataset http://www.cvg.rdg.ac.uk/PETS2001/pets2001-dataset.html [Last Accessed 11/09/2014]

15. Castleman K. R., Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.

16. N.T.Binh, T.A.Dien, "Object Detection and Tracking in Contourlet Domain", Context-Aware Systems and Applications, Vol 109, 192–200, ISSN 1867-8211, 2013

17. E. J Candes and L Demanet, "The curvelet representation of wave propagators is optimally sparse" Communications on Pure and Applied Mathematics, Vol 58, pp. 1472-1528, 2004

18. Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, Pete Steggles, Towards a Better Understanding of Context and Context-Awareness, Lecture Notes in Computer Science Volume 1707, pp 304—307, 1999

19. Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, Shipeng Li. Automatic Salient Object Segmentation Based on Context and Shape Prior. British Machine Vision Conference, pp 1-12, 2011

20. Nguyen Thanh Binh, "Vehicle Tracking in Outdoor Environment Based on Curvelet Domain", Nature of Computation and Communication, Vol 144, pp 360-369, 2015

21. Li Zhihua et al, "High efficient moving object extraction and classification in traffic video surveillance", Syst. Eng. Electron. 20(4), pp. 858-868, 2009

22. Tsuchiya M and Fujiyoshi H, "Evaluating feature importance for object classification in visual surveillance", Proc. of the 18th IEEE International Conference on Pattern Recognition, pp. 978–981, 2006

23. 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, DOI 10.1007/978-3-662-45947-8_7, ISSN: 0302-9743