Online Learning Classification for Video Monitor

Zhiyuan Li, Chao Wang[a] [*], Xiaoduo Zhang

Zhengzhou Electromechanical Engineering Research Institute, China

[*a]854909839@qq.com

Key words: Online learning, Unsupervised, Objects classification, Video Monitor.

Abstract. This paper presents an online unsupervised learning classification of pedestrians and vehicles for video Monitor. Different from traditional methods depending on offline training, our method adopts the online label strategy based on temporal and morphological features, which saves time and labor to a large extent. It extract the moving objects with their features from the original video. An online filtering procedure is adopted to label the moving objects according to certain threshold of speed and area feature. The labeled objects are sent into a SVM classifier to generate the pedestrian & vehicle classifier. Experimental results illustrate that our unsupervised learning algorithm is adapted to polymorphism of the pedestrians and diversity of the vehicles with high classification accuracy.

1. Introduction

During the past decades, classifying moving objects in traffic scenes for security monitoring has been an emerging research. Object classification tends to provide basis for subsequent video analysis and processing (e.g.

public safety application, video synopsis and video semantic search). If a suspect break into particular area with camera, the object classification system can classify the object and call the police. Obviously, object classification removes large amount of redundant information for human labor saving.

Due to its importance, shape-based classification has been significant research in the recent past[1]. However,there are still some challenges to increase the accuracy and

adaptability of classification, due to the following reasons:1)The diversity of outdoor environment conditions (e.g illumination, camera position and occlusion) [2,] which is shown in Fig.1 (a). 2) The varying shapes and views of objects. Objects in the same class could have different shapes, which are shown in Fig.1 (b) and (c). On the other side, pedestrians and vehicles may have similar shape in special view.3) Incomplete appearance details of moving objects, for the input objects is based on the result of detection.

Supervised object classification labeled by morphological features in video has been addressed with silhouette to solve the problem, which is commonly used in surveillance systems. Collins R et al. proposed the morphological information based classification to separate the area of people and car[3], but the accuracy performance was not robust. L.Zhang et al. proposed the appearance-based method to achieve robust classification in diverse camera-

viewing angles[4]. However, classifying object

Fig. 1 Scenes and objects in our database.

(a) Four dialy scenes in surveillance video, the difference between them mainly acts on camera position and objects occlusion; (b) Pedestrians belong to scenes in (a), and shapes of these pedestrian objects are almost similar; (c) Vehicles belong to scenes in (a), but shapes between these vehicles have large difference

with only one kind morphological feature is not sufficient in low resolution surveillance videos. Boosting methods have been researched extensively. Viola et al. proposed a novel framework for feature selection by boosting[5]. However,these methods require a large number of heavy computation and human labor to handle on samples labeling.

In this case, the popular method is to adopt an online learning algorithm to reduce the number of labeling samples. Semi-supervised learning classification has emerged as an important learning framework in recent years. Methods based on semi-supervised classification exploit the information from a combination of both labeled and unlabeled data. A typical learning approach is the co-training algorithm proposed by Blum and Mitchell[6], which the classifiers train each other using the unlabeled data. Fanet al. proposed an improved EM-based semi-supervised learning method to reconstruct the labeled training set[7].Xin Zhao et al. presented a global and local training algorithm which trained local and global classifier to vote the most reliable feature clusters[8]. Other works have proved that semi-supervised learning strategy can realize very accurate classification[9−11].

However, most of existing semi-supervised methods are batch, which lead to a large amount of memory and heavy computation. If the batch learning carried out, retraining is the only way to update the model. Moreover,semi-supervised approaches are not appropriate to be applied when the unlabeled examples come in an online manner. In these situations, unsupervised learning algorithms are quite suited to surveillance application.

This paper proposed an online unsupervised learning classification of pedestrians and vehicles in surveillance video. As shown in the block diagram in Fig. 2, the objects classification has different modules: objects detection and tracking, feature extraction, online labeling, online training, and objects classification. The module of objects detection, tracking and feature extraction insure that we got all the people and car from original video, and also store the temporal and morphological features of them.Then we proposed an online labeling strategy. Two types of features are implemented: label features and classification features. By thresholding label features to screen and label the moving objects online, the labeled objects and classification features of them are then input into SVM classifier to learning the pedestrian & vehicle classifier.Finally, we input all of the moving objects into the pedestrian & vehicle classifier to finish the classification.
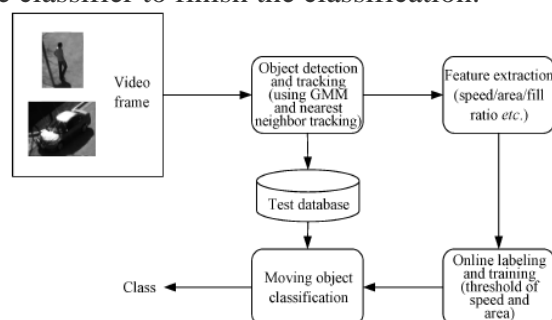

Fig. 2 The online learning classification framework

The contribution of our method is presenting an objects classification algorithm of pedestrians and vehicles by online unsupervised learning, which has strong self-adaptability to the polymorphism of pedestrians and the diversity of vehicles, also leads to a high classification accuracy.

2. Online Labeling Strategy

As we know, the moving objects have various view and shape, which usually caused by the different angles of camera. For this reason, training the pedestrian classifier and vehicle classifier based on sample database, which have been proposed in many practical classification tasks,would be impractical in surveillance videos. For those reason, we proposed an online labeling and training algorithm, which label the moving objects from the original video and training the labeled objects to SVM classifier.The framework of this section is shown in Fig.3.

For the speed feature is robustness to the change of the objects shape and the background brightness, We ranked all the moving objects by the speed feature, then set the Low speed threshold (ThL) and High speed threshold (ThH) for first filtering the moving objects after ranking.If the object speed is larger than high speed threshold,label that object to vehicle. For those objects move speed are less than low speed threshold, they can't be labeled to pedestrian easily due to slow movement (e.g. reversing,brake, etc.) of vehicle. To solve this speed labeling problem, we label the object which speed of it is less than low speed threshold to prior-pedestrian, which will be second filtering. Label the other objects to unclassified samples.Considering that vehicle is much bigger than pedestrian in neighboring pixel area, which makes object area become the second filtering threshold. We ranked the prior pedestrian objects by the area feature, and then we set the area threshold for second filtering: Label the object with area less than area threshold to pedestrian and the other objects to unclassified samples to finish our online labeling process.
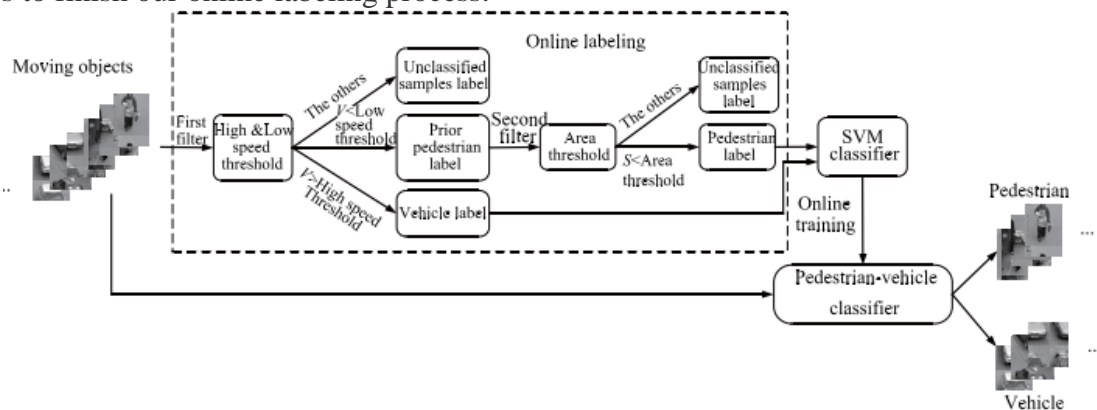

Fig. 3 The online labeling strategy framework

The value of low speed threshold and high speed threshold and area threshold should depend on the classification accuracy. We set the threshold to be percentage, for example, the computation formula of high speed threshold is $\text{ThH} = \dfrac{N_{VH}}{N_{ALL}}$, which the N_{VH} denote the number of vehicle objects, and the N_{ALL} denote the number of all the moving objects. This formula means we label the object that speed is large than $N_{V\,H} = N_{ALL} \times \text{ThH}$ to vehicle. The definition of low speed threshold and area threshold is similar to high speed threshold.

After a plenty of experiments, we found the label accuracy and classify accuracy of vehicle is the highest under the high speed threshold equal to 0.2. And the label accuracy and classify accuracy of pedestrian is the best under denote the low speed threshold to 0.2 and area threshold

to 0.5, which will be shown in our experiment result on section 4.

3. Unsupervised Moving Objects

Classification System

We proposes a moving objects classification system,including three phases: moving objects detection and tracking, feature extraction, online labeling and training based on speed and area threshold, and input the test sets into classifier for classification results.

3.1 Moving Objects Detection And Tracking

Our objects detection is based on Gaussian mixture model, which is very popular in objects detection. Gaussian model recorded the pixel values distribution feature of each pixel. It was realized based on the following two facts: background holds for most of the time in long-term observation, and more data support the distribution of the background; Even the relatively color moving objects also may produce more changes than the background, and objects usually with different colors. It is adaptive to solve the multi-mode changes of background in the dynamic environment. And objects tracking method is based on tracking the location feature of the nearest neighbor. To assume the position of the moving objects is denoted by the center of its connected domain external rectangular. We adopted the linear prediction algorithm, to predict the location T_i^t of the objects which is the number i of objects queue in frame t, through the coordinate's trajectory of the No.i objects in last few frames:

$$T_i^t = P_i^{t-1} + \sum_{k=t-24}^{t-1} \lambda^{k-t-1}(P_i^k - P_i^{k-1}) \tag{1}$$

Where λ is a parameter less than 1, which support $\sum_{k=t-24}^{t-1} \lambda^{k-t-1} = 1$.

Suppose there are numbers m of moving objects is detected in the current frame t (i.e., numbers m of connected domains), the center location vector is O_1, O_2. Om. Calculate the distance between T_i^t and O_1, O_2, Om, the position of the shortest distance point corresponding the location of the objects in frame t.

From the objects detection based on mixture Gaussian model, we can get all the moving objects from the source video and its feature like area, empty ratio, aspect ratio, dispersion ratio, Histogram of oriented gradients (HOG)features. From the algorithm of objects tracking based on

Nearest neighbor correlation, we can obtain the dynamic speed feature of moving objects and the objects queue.

3.2 Feature Extraction

The feature extraction phase is a crucial step in objects classification. Therefore, investigating different types of features and combining them is one of the important aspects of this study. Currently, there are two kinds of feature: static feature and dynamic feature [12]. Static feature containing grayscale features, color features, shape feature and texture feature. Extract static feature do not need to consider the continuity of frame, so this feature has a good real-time performance; Dynamic feature including the location of the objects, motion vector and periodic of the motion. Dynamic feature have certain robustness to the change of the objects shape and the background brightness.

3.3 Label Feature Extraction

Since the positive and negative samples should be filtered by the label feature, and the label feature should achieve high label accuracy. The label feature should be robust to the change of the objects shape and the background brightness. For these reason, we choose the speed feature as preliminary feature. In this paper, we calculated the average speed:

$$V_{nt} \frac{\Delta S}{\Delta T} = \sqrt{\frac{(xn_t + 2 - xn_t)^2 + (yn_t + 2 - yn_t)^2}{2T_n}} \tag{2}$$

It result to the speed V nt of the objects which is Number n in frame t, (xn_t, yn_t) is the center coordinate of the nth objects in frame t, (xn_{t+2}, yn_{t+2}) is the center coordinate of the nth objects in frame t + 2, T_n is the time consumption between each frame of nth objects.

But, there are usually some low speed vehicle which appears in the surveillance video, which may cause by scene constraints and view of the surveillance videos. This paper increases the area feature to be the secondary filter label feature, it promote the classification accuracy of pedestrian objects. We extracted area feature by calculate the total number of pixels of foreground objects, which have the characteristics of simple and intuitive. We initialed the counter to be 0, from the top left corner of minimum bounding rectangle of objects, scan pixels from top to bottom and from left to right, accumulate the counter if meet a foreground pixel.

3.4 Classification Feature Extraction

We extract the classification features by the morphological features, including empty ratio, aspect ratio, dispersion ratio and HOG features. These features have good effect to the classification of pedestrian and vehicle objects.

Aspect ratio: It is the ratio of the MBR (Minimum bounding rectangle) Height (Rect H) and MBR Width (Rect W) of the binary blobs. Aspect =Rect W/ (Rect W + Rect H).Although this feature can well describe the shape of the objects information, but it will change when the objects rotates, which affecting the accuracy.

Empty Ratio: Empty ratio is the ratio of size of the blob Segment Area (S) and the total MBR area (Rect S).Empty =S/Rect S. This feature has translation and rotation invariance.

4. Experimental Results

In the experimental section, we collected six video sequences to perform moving objects classification, which are no open data sets for moving objects classification. We performed the following objects classification experiments: 1) Setup the parameter mainly by confirming

The value of speed and area threshold, 2) Verify the self-adaptation of proposed classification in several scenes, 3) Compare and analyze our algorithm with current approaches.

4.1 Parameter Setup

We compares several groups of experiment to setup the value of the label feature threshold, including speed high threshold, speed low threshold and area threshold. The experiment tested the label and classification accuracy in different cases of threshold values. Consider the speed and area of objects may have large difference between different movements. We choose two traffic scenes respectively contain horizontal and vertical road to confirm the most robust threshold. Due to the speed threshold and area threshold are percentage of objects number in foreground sequence.

4.2 Classification Results and Analysis

We adopt six videos from real-world traffic scenes to evaluate our methods. To demonstrate the effectiveness of motion pattern learning, we run our algorithm on many videos from different traffic scenes. In this work, the results are shown in few typical scenes, which include: straight road, crossroad, the school gate, the parking lot, etc. All of these videos are from real-world traffic scenes and include illumination changes, occlusions, shadows, and different environmental effects. We picked up some classification results of the few scenes. The results of pedestrian classification are shown in Fig.5, and Fig.6 shows the result of vehicle classification.
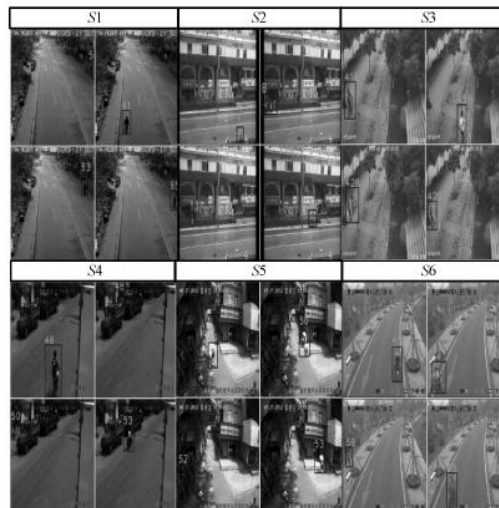
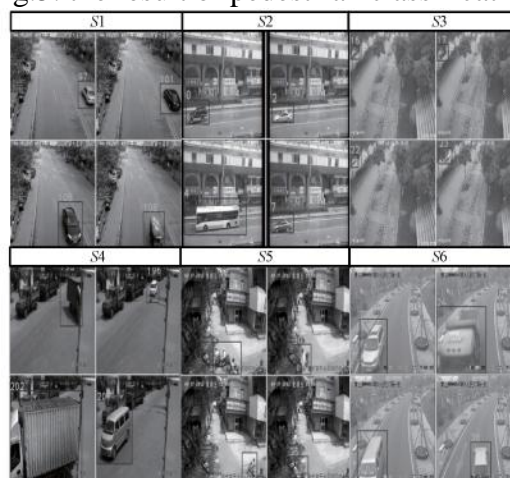Fig.5. the result of pedestrian classification

Fig. 6. The result of vehicle classification

5. Conclusion

As we proposed, in multi-scene surveillance videos, it is reliable to handle the problem by online unsupervised learning classification of pedestrians and vehicles. We obtain the moving objects from original video and extract the features. Then, we define the threshold of speed and area to filter and label the moving objects online. After that, we input the labeled objects into SVM classifier to generate the pedestrian & vehicle classifier. And the pedestrian & vehicle classifier is experimented with high accuracy: 91.6% for pedestrian and 93% for vehicle.

References

[1]. Zhi Kai Ku, "Shape-based recognition and classification for common objects – An application in video scene analysis", International Conference on Computer Engineering and Technology, Chengdu, China, pp.16–18, 2010.

[2]. Mohamed Elhoseiny, "Multi Class object class cation in video surveillance systems experimental study", IEEE Conference on Computer Vision and Recognition Workshops, Oregon, Portland, pp.788–793, 2013.

[3]. Xin Zhao, "Global and local training for moving classification in surveillance-oriented scene", Asian Conference on Pattern Recognition, Beijing, China, pp.681–685, 2011.

[4]. T.Z. Zhang, "Mining semantic context information for intelligent video surveillance of traffic", IEEE Transactions on Industrial Informatics, Vol.9, No.1, pp.149–151, 2013.

[5]. R.N. Hota, V. Venkoparao, A. Raja Gopal, "Shape based object classification for automated video surveillance with feature selection", International Conference on Information Technology, Roukela, India, pp.97–99, 2007.

[6]. Navneet Dalal, Bill Triggs, "Histograms of oriented gradients for human detection", IEEE Conference on Computer Vision and Pattern Recognition Workshops, San Diego, CA, USA, pp.886–893, 2005.

[7]. J. Xu, M.H. He and J. Han, "A comprehensive estimation method for kernel function of radar signal classifier," Chinese Journal of Electronics, Vol.24, No.1, pp.218–220, 2015.