Research on Moving Target Recognition and Tracking Technology Based on Machine Learning

Ying Zhang¹, Jie Li¹, Yi Wang¹,*
¹College of Informatics, Huazhong Agricultural University, Wuhan, 430070, Hubei, China
*Corresponding author e-mail: zy@mail.hzau.edu.cn

Abstract: In this paper, the SIFT feature, HOG feature and SVM are adopted to design a movement vehicle recognition system, when the vehicle is relatively close to the camera and it backs to the camera, also the vehicle can be implemented to track. In this paper, two feature extraction methods, SIFT feature and HOG feature are respectively used to extract the features from 6000 images of vehicles with back facing the cameras and 6000 images without vehicles. Then SVM is used to train them to obtain a model. Then, we use the moving object recognition algorithm to extract the moving object in the given video, and put each moving object into the trained model to judge. If the model is judged as a vehicle, it will be tracked.

1. Introduction
Moving object recognition refers to the recognition of objects with physical motion in video. This technology has great significance in video monitoring, human behaviour analysis, road monitoring, airport security, intelligent security and so on[1].

Background subtraction, inter-frame difference method, and optical flow are the more commonly used algorithms for motion target recognition. Machine learning for artificial intelligence focuses on three areas of research: learning methods, learning mechanisms, and face to the missions. Through machine learning, motion target recognition is of great significance in driverless area.

In this paper, a simple motion vehicle recognition system is designed by machine learning. The system is able to train a model that can recognize the vehicle by using a large number of images as a sample set, then extracting the motion targets in the video sequence by a motion target recognition algorithm, finally, the model is used to predict whether the moving target is a vehicle or not. If it is a vehicle, you can record the position of the vehicle's coordinate points in the video, Through these positions, we can get the approximate movement locus of the vehicle, thus realizing tracking.

2. Data processing
The system firstly uses the SIFT algorithm to extract the SIFT feature points of the sample set images, and then uses the clustering algorithm KMeans to divide these points into categories, which can be used as parameter inputs to the system. Finally, using the bag-of-words model and the matcher, the images are mapped into feature vectors.

2.1 SIFT algorithms
SIFT is short for Scale Invariant Feature Transform. A significant difference from other features is that the SIFT feature is not affected by some factors, such as the pictures of scaling, rotation and brightness changes. We can take a simple example. There is a corner in the original image, but after the image
magnified several times or rotated, the computer does not necessarily recognize that original corner. In fact, this corner itself is only magnified or rotated. The SIFT algorithm can solve this problem. Based on the matching method of SIFT feature, it is widely used in the fields of target recognition, mosaic panorama, and recovering structures of motion[2-3].

The feature points extracted by using SIFT in this system, which are shown in Fig 1.

![sift features keypoint](image)

Figure 1. sift features keypoint

2.2 KMeans algorithm
Cluster learning is one of the methods of pattern recognition and data mining, the purpose of this algorithm is to divide a number of unlabeled objects into K categories, the distance sum which from the objects in the same category to the center of that category, is as small as possible. since the sample of the KMeans algorithm and original sample are unlabeled, it is an unsupervised learning algorithm[4].

2.3 Bag-of-words model
The bag-of-words model was originally a model of expression, under the domains of natural language processing and information retrieval, bag-of-words model, just as its name implies, we can be thought of as a bag containing words, so that we can use the bag-of-words model to represent the text if we do not consider the order in which the words appear in the text, but only the frequency of their occurrence[5].

3. SVM
Support Vector Machine (Support Vector Machine) is a dichotomy classifier, it is a new approach to data mining and a significant application in face recognition, handwriting recognition and other aspects, there are two kinds of training data labels, SVM algorithm finds a division of the hyperplane and separates the two kinds of data, through this division of the hyperplane, the new data can be predicted for the label or category of the data. So the main factor that affects the prediction accuracy is whether the selection of division of this hyperplane is suitable or not[6].

4. Motion Target Recognition Algorithm

4.1. Background differencing
Background differencing is the identification of motion targets by comparing the difference between the current frame and the background model. Therefore, an important task for us is to find an accurate background model. However, the complexity of the scene and various disturbances, such as the sudden changes in lighting and the shaking of camera, make it difficult to find an accurate background model. There are two common and relatively simple methods of background modeling, the median and averaging methods.

4.2 Inter-frame difference method
When there are moving objects in the video, the grayscale between the two connected frames will be different, according to this feature, we can get the difference of two connected frames and have the
absolute value, so after the difference of the still part of the image is 0, after the difference of the moving part is non-0, when the non-0 difference is greater than a certain threshold value, we can determine that this part is the moving object.

5. HOG’s characteristics

5.1 Introduction of HOG
In terms of image feature extraction, in addition to the above-mentioned SIFT feature, there is another more commonly used HOG feature, namely the gradient histogram. The main idea is selecting some areas in an image, then counting the gradient directions and sizes of the pixels in the region, constructing the histograms for several directions, which in a voting way and it is similar to compute SIFT feature point descriptors. Then a number of histograms are used to construct the final HOG feature[7].

5.2 Extract the HOG’s features
There are four steps to extract the HOG’s features of an image: calculating the image gradient, statistics of gradient direction, feature standardization and generating the feature vector. (1) computing the image gradients; (2)statistics of gradient direction; (3)feature standardization; (4) generating the feature vectors[8].

6. Systematic design

6.1. System framework
The system is mainly divided into three modules, the first module is the image processing module, which is responsible for processing the image into a one-dimensional vector, and its main steps are: firstly extract the SIFT features of the image, and then use the KMeans algorithm to cluster these points, and finally generate BoW image feature extractor. The second module is the machine learning module, this module is divided into two parts, the first part is the learner, which is responsible for learning an SVM model by the sample set, the second part is the decision-making, which is responsible for importing the previously trained model, input the new data into the model and predict the category of new data. The third module is the moving target extraction and tracking module, which is also divided into two parts, the first part is the moving target recognizer, which is responsible for detecting and identifying the moving objects in the video, and the second part is the target marker, which is responsible for marking the specific moving objects so as to achieve tracking.

6.2 Results

6.2.1 Presentation environment
The system runs in the following hardware and software environment: operating system: windows 10; processor: inter Core i7-8700k; development tools: psycharm community version; Python version: python3.7.

6.2.2 Interface
Positive Sample Directory, Negative Sample Directory, Test Set Directory are the names of the folders that store the Positive Sample Picture Set, Negative Sample Picture Set, Test Set Picture Set, but Flann Algorithm, SVM C, Flann Trees are the some parameters of the SVM. Clusters represents that the SIFT feature points in an image, which divided into Clusters category by KMeans algorithm. K-Means Samp indicates that how many samples are selected from the sample set for training the KMeans Cluster. The last SVM Kernal indicates that whether a SVM of Linear type or a RBF type is used. In addition, the program can also save the trained SVM model as a mat type file, and load it directly if you continue to use the model next time.
6.2.3 The effect picture of presentation

There are over 6,000 samples, in each of positive and negative samples from the entire picture set, 70% of which are used as training sets and the remaining 30% as test sets.

Figure 2. test set 1

Figure 3. test set 2

In this paper, a video of a traffic accident scene is selected as a test video, and the accuracy of the SVM model trained by SIFT features, bag-of-words model and KMeans is 83.49%. This is why there are cases where the test mistakenly identifies a non-vehicle as a vehicle.

Figure 4. SIFT Demo

The accuracy of the SVM model trained by the HOG features is 98.67%, which is relatively high. Therefore, it also performs better in test videos, accurately identifying the moving vehicles in the test video, rarely and mistakenly identifies a non-vehicle as a vehicle.
Figure 5. HOG Demo

The comparison between them is as follows:

| Method | Extraction algorithm of feature vector | Machine learning algorithm | Accuracy of the model |
|--------|----------------------------------------|---------------------------|----------------------|
| 1      | SIFT, Kmeans, bag-of-words              | SVM                       | 83.49%               |
| 2      | HOG                                    | SVM                       | 98.67%               |

7. Summary
In this paper, the SIFT feature, HOG feature and SVM are adopted to design a movement vehicle recognition system, when the vehicle is relatively close to the camera and it backs to the camera. In the design process of whole system, the SIFT feature, KMeans algorithm, bag-of-words model and HOG feature are used to extract the feature vector of the image, at the same time, the two mainstream feature extraction methods are compared with the effect of SVM collocation in detail. It also uses the processing training set of SVM in opencv and thus obtain the training model. Finally it is also implemented to extract the moving objects in the video by background subtraction. By recording the coordinate points in the vehicle's movement, it is possible to store the movement locus of the vehicle, thus realizing a simple tracking function.

Regardless of which method is used, although the final trained model has more than 80% accuracy. Since the pictures of the vehicles in the training data set are the back of the vehicle, they are not well identified when the vehicle is facing to the camera. In addition, this system firstly extracts the moving objects by background subtraction, for the moving objects that are farther away from the camera, the extracted image is much smaller than the size of 64 * 64, it will mistakenly identified the non-vehicle moving objects as vehicles.

Acknowledgements
We would like to thank the HZAU funding project(Ying Zhang’s project in 2020).

References
[1] Zhang Qian, Zhao Xinxue, Review of the detection methods for moving target [J]. Electronics World, 2019(04):65.
[2] Hu Ruiqing, Tian Jierong, Research on Detecting Algorithm of Moving Target, Based on Optical Flow [J]. Electronics World, 2019(05):58-61.
[3] Lv Wei, Zhong Zhenyi, Zhang Wei, Overview of artificial intelligence technology [J]. Shanghai Electrical Technology, 2018, 11(01):62-64.
[4] He Qing, Li Ning, Luo Wenjuan, Shi Zhongzhi, The machine learning algorithm of big data is summarized [J]. Pattern recognition and artificial intelligence, 2014, 27(04):327-336.
[5] Bai Yuxuan. Kmeans application and feature selection [J]. Electronic Technology and Software Engineering, 2018(01):186-187.
[6] Cui Fangyuan. Video-based vehicle inspection research and application [D]. University of Electronic Science and Technology of China, 2013.
[7] Lowe D G. Distinctive Image Features form Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
[8] Brown M, Lowe D G. Recognising Panoramas[A]. In Proceeding of the 9th international Conference on Computer Vision(ICCV03)[C]. Nice, October, 2003.