Vehicle Recognition Based on Local Feature

Ying Qian & Yifan Zhang*
College of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

ABSTRACT: This paper introduced a vehicle recognition system based on Speed-up Robust Feature (SURF) and bag-of-features (BOF). In the system, we extract SURF features of vehicle images, and use k-means algorithm to analyze the features, clustering center of the features, so we can get some “visual words”. Then we describe all images use those “words”, and generate histogram to quantize the features. In this paper, term frequency-inverse document frequency (tf-idf) is used to weight the features to weaken the influence of useless features. We have 22 different type and brand of vehicle, including sedan, SUV, bus and so on. The experiment shows that this way can make the recognition accuracy over 90%.

Keywords: vehicle recognition; speed-up robust feature; bag of feature; term frequency-inverse document frequency

1 INTRODUCTION

Intelligent traffic monitoring has become important recent years, and vehicle recognition is a significant part in Intelligent Transportation System (ITS). Various equipment are applied to achieve the recognition such as infrared, magnetic induction, laser radar and so on. In all these ways, vision-based recognition system is low-cost, easy to maintenance, and it contains more important traffic information. It is of great value to develop and use.

Currently, research on vehicle recognition focuses on two aspects, vehicle structure classification (sedan, SUV, bus, truck) and vehicle type recognition (which type of which brand), and many researches have been done on this two parts.

Paper [1] presents a way to classify vehicle use length information and puts forward an image processing algorithm. They used streams of images captured from un-calibrated video camera, compared the length of different vehicles in order to estimate the truck volumes, and eliminated the needs of different complex calibration systems.

Aiqin Hu et al. proposed to learn manifold hierarchical features via a deep learning model Deep Boltzmann Machines (DBM)[2], and utilized the human-engineered descriptors such as Log-Gabor, HoG and Gist as the source data of deep learning architecture, and they classified all vehicle into four type: bicycle, motorbike, bus and car.

Wei Pei et al. utilized hessian-affine detector to detect the features of the input image and used SIFT descriptor to transform each feature into a 128-dimensional vector [3], after quantizing the features as a “visual word”. They used Support Vector Machine (SVM) to classify the vehicle into 28 types.

Thitiphat et al. extracted Eigenfaces and Pyramid Histogram of Oriented Gradients (PHOG) of vehicle image [4], and then used K-nearest neighbor (KNN) to train the features. They recognized the logos and frontal masks to identify the car type.

This paper takes the frontal picture of vehicle as the database and query set, and 22 types of vehicle are collected. We propose a vehicle recognition scheme based on a local feature of image named Speed-up Robust Feature. This paper is organized as follows. Section 2 describes the detail of our recognition scheme. In Section 3, experiment results are explained and analyzed. In the last section, the conclusion of experiment and further works are described.

2 THE PROPOSED SCHEME

Process of our vehicle recognition system is shown as Figure 1:
Step 1. Detecting interest points, extracting feature vector: we choose Speed-up Robust Features (SURF [5]) to describe vehicle images. Assuming there are $N$ vehicle images in the database, each image contains $n_i$ feature points, and then all feature of all image compose a $\Sigma_{i=1}^{N} n_i$-by-64 matrix.

Step 2. Building Bag-of-features: after extracting features from vehicle image, we quantize the feature descriptors into clusters which we called “visual words [6]”. This step can reduce the dimensional of feature matrix above and accelerate the recognition. The $\Sigma_{i=1}^{N} n_i$-by-64 feature matrix become a $\Sigma_{i=1}^{N} n_i$-by-1 matrix since we utilize “visual words” to replace the SURF descriptors. After that, we count a vector of word frequencies, suppose there is a vocabulary of $j$ words, and the dimensional of feature matrix will reduce to $N$ -by-$j$. The process above is shown in Figure 2.

Step 3. Visual indexing using tf-idf: consider of useless features will reduce the accuracy of recognition, and we apply a weighting to the frequency vectors.

Step 4. Vehicle recognition: use cosine of angle to measure the similarity between the query image and all images in vehicle database.

2.1 Feature detecting and extraction

Traffic images are always affected by illumination variation and weather change, in addition to the change of pose and scale caused by the different location when vehicle is shot, and occlusion occur frequently on the road. Taking into account those points above, this paper chooses SURF as the descriptor of vehicle features. SURF is invariant to rotation, brightness variations and scale changes, and it’s robust to perspective changes, affine transformation and noise. As a promotion of Scale-invariant Feature Transform (SIFT [7]), SURF attracted intensive attention and be used in various fields.

The interest point detection approach of SURF detects the locations where the determinant of Hessian matrix is maximum and those locations are blob-like structures in the image. The interest point description step of SURF exploits integral images, so that the speed of descriptor extraction can be boosted. By way of assign a main orientation to interest point, the algorithm achieved rotation invariant. Haar wavelet filter is utilized to decide the orientation, it is invariant to Brightness.

2.2 Building Bag-of-features

Bag-of-words (BOW) is one of the most popular methods for text categorization. The key idea is to ignore the grammar and contextual relation between words, just regard the article as a set of words, and then represent each text by a histogram of those words. Compare histogram of different text to retrieval the similar article. In recent years, this modal has been used on image categorization and image retrieval [8]. Neglect the contextual relation and location of local patches or features, describe images with those patch-es or features (visual words), and regard image as a set of visual words. Build histogram of visual words to describe an image.

In the previous step, a lot of feature points are extracted from each vehicle image, and each SURF feature point is described by a high-dimensional vector, thus a great deal of time will be spent when we doing vehicle retrieval. Hence we build bag-of-features (BOF) to reduce the dimensional. There are two steps to establish the bag-of-features: build visual vocabulary and count the histogram of visual words.

1) Build visual vocabulary (see Figure 4): extract SURF features from every image in the database. Assuming there are $m$ vehicle images in the database, each image contains $n_i (i = 1, 2, ..., N)$ feature points. Cluster those features to visual words using k-means. After encoding each feature vector by mapping it to a visual word, the dimensional of feature matrix in an image is reduced from $n_i$-by-64 to $n_i$-by-1.
2) Build histogram (see Figure 5): after quantizing the feature vectors with the \( j \) words in the visual vocabulary, calculate the frequency of each word in the image, and generate the histogram (bin= \( j \)).

2.3 Weighting features

In text retrieval, each document is represented by a vector of term frequencies. However, it is usual to apply a weighting to the components of this vector rather than use the frequency vector directly for indexing \([9]\), because different words in a text document should have different importance. Term frequency-inverse document frequency (tf-idf) is a well-known algorithm in field of text retrieval, it can solve above problem. Compute the term frequency, the bigger \( tf \) value of a term in a document demonstrate that this term is appear repeatedly in this document, which devotes more weight. Inverse document frequency represents the frequency that some term exists in other document, and the bigger \( idf \) value shows that this term is not that unique, which devotes less weight in this document. Combining \( tf \) and \( idf \), we can weigh the features. \( tf-idf \) is calculated as follow:

1) Calculate \( tf \) value:

\[
tf_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{k} n_{i,j}}
\]  

\( n_{i,j} \) indicates the number of occurrence of \( j \)th word in the \( i \)th image, and \( k \) is the number of words in \( i \)th image. \( \sum_{j=1}^{k} n_{i,j} \) is the total number of occurrence of all words in \( i \)th image.

2) Calculate \( idf \) value:

\[
idf_{i,j} = \log \frac{N}{n_j}
\]

\( N \) is the total number of images, and \( n_j \) is the number of images which contain \( j \)th word.

3) Calculate \( tf-idf \):

\[
tf - idf_{i,j} = tf_{i,j} \times idf_{i,j}
\]

Finally we utilize \( tf-idf \) as the weighted descriptor of image.

2.4 Vehicle recognition

Cosine of angle is used to measure the distance between the query image vector and all document vectors in the vehicle database. Cosine similarity is calculated as follow:

\[
\cos \theta = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} (A_i \times B_i)}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \sqrt{\sum_{i=1}^{n} (B_i)^2}}
\]

If result of formula (4) is close to 1, it shows that there is a high similarity between \( A \) and \( B \); when it is close to 0, it shows that the distance between \( A \) and \( B \) is far. Compute the similarity between query images and all images in the database, rank all the distances, and get the best matched vehicle from the database.

3 EXPERIMENT AND DISCUSSION

Face pictures of vehicle are chosen in this experiment. We collect vehicle images in various weather and lighting condition, including cloudy day, sunny day, noonday and dusk.

We have 22 types of vehicles in our database, including sedan, bus, SUV, and midsized car (see Figure 6). We labeled these images with their type names, so that we can get the result by comparing the labels of retrieved image and query image. This database consists of 10 images of each type of vehicle, and a total of 220 images in it. In query set, we have 300 vehicle images in it, and all vehicle types in query set are included in database.
Even though we shot the photos in the similar view angle, the pose of vehicle is not exactly identical to each other. Considering we can’t control the pose of vehicle be shot in actual scenario, we collect a variety of pose into dataset as much as possible, to ensure that we always can get the most similar image which have same type from dataset.

The resolution of each vehicle image (query set and vehicle database) is not the same.

The following experiments measure the accuracy and recognition speed of our method.

When we input a vehicle image, the system will show us the top 5 retrieved images (see Figure 7, the first image of each type is query image, and the rest are retrieved images, the number below the image is the cosine similarity coefficient between query image and retrieved image). According to these results, we can find that our method is robust to lighting change, occlusion, rotation and perspective change.

When we build the vocabulary, the number of words we defined will have an important effect on the retrieval result.

Figure 8 shows the recognition accuracy when we use different vocabulary.

We can conclude that in a certain words number range, a bigger size vocabulary can bring a better accuracy, but when the size exceed a peak, the accuracy will not heighten anymore, this result caused by over-fitting. If we define an oversize vocabulary, the words in the vocabulary will lost the uniqueness, and words will be sensitive to noise. In addition, it leads to a large amount of calculation when we cluster the center of features, and when we count the histogram of words, the dimensional of the histogram will increase, which will cause the waste of storage space and slow...
down the speed of recognition. However, if we define an undersized vocabulary, the words will be incapable to represent the features of image. So, it is important to decide the number of words.

3.1 Comparison of accuracy between SURF and SIFT feature

Some article utilize SIFT to recognize the vehicle type [3], this paper reappeared this method. The comparison of accuracy is shown in Figure 9.

From the figure we can know that the accuracy of system which utilizes SIFT is lower than that utilizes SURF. But in our previous experiments, we used our scheme on blurred images, in that experiment, the accuracy of SIFT features is much better than SURF features. As paper [10] concluded, SIFT is not good at illumination changes, while it is invariant to rotation, scale changes and affine transformations. SURF has good performance as the same as SIFT, and SURF have best performance in illumination changes. SURF is faster than SIFT. Considering the actual road condition and the requirements for real-time recognition, we think SURF is a better descriptor in vehicle recognition.

3.2 Comparison of accuracy between histogram of BOF and tf-idf

We have two ways to compute the similarity between images each other. One way is to calculate the distance between image histogram, and another way is to compute the similarity between image tf-idf. Histogram is a statistic of frequency of visual words occurrence. tf-idf is an data which be obtained after weight the features of image. tf-idf can reduce the influence of useless features, it is used to replace histogram. The experiment result is shown in Figure 10.

According to the figure above, we can draw a conclusion that after weight features by tf-idf, the accuracy of recognition is improved, and we can get a better match between vehicle images. The retrieval result and similarity coefficient between query image and retrieved image is shown as Figure 11.

4 CONCLUSION

In this paper we achieved vehicle recognition using image local feature named SURF feature. To quantize features, we used Bag-of-features to transform the high dimensional feature to a vector of histogram. To weaken unnecessary features, we substitute tf-idf for statistical histogram. We can conclude that this way can recognize vehicle type effectively, and the recognition rate is above 90% when we defined an appropriate vocabulary size.

But there is still some restriction in this experiment. Through analysis the result, we found some recognition errors is due to complex background, most of the misidentifications is caused by blur, and the performance of SURF on blur is inferior to SIFT. If the clarity of query image is different from images in database, it might result in wrong recognition.

In our experiment, we collected 22 types of vehicle, and each type contains only 10 images in the database. Our experiments demonstrate that to some extent, more sample images in database can improve the
recognition accuracy. The chosen of database images can directly affect the recognition rate, so it has some uncertainty in the experiment. If we collect more vehicle type, the recognition accuracy might decrease, and a larger vocabulary might be required.

In the future work, we could attempt to increase the vehicle type, and try to find an appropriate vocabulary size in practical application scenario.

REFERENCES


