The Journal of Engineering

Research Article

Colour fast-match for precise vehicle retrieval

eISSN 2051-3305
Received on 12th June 2019
Revised 16th October 2019
Accepted on 21st October 2019
E-First on 2nd March 2020
doi: 10.1049/joe.2019.0882
www.ietdl.org

Jian Wei[1,2], Yue Wang[2], Feng Liu[1,2] ✉, Qiuli Lin[2], Ning Wang[2]

[1]*College of Education Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China*

[2]*Jiangsu Province Key Lab on Image Processing and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China*

✉ *E-mail: liuf@njupt.edu.cn*

Abstract: The explosive growth of vehicles has increased the importance of intelligent traffic system. However, compared with face recognition, vehicle retrieval has not attracted the attention of researchers in vision community. Precise vehicle retrieval has always been a challenging task because it requires the retrieval of all vehicles with the same visual attributes from a large number of vehicles with subtle visual differences. To handle this, the authors propose to implement precise vehicle retrieval using an improved fast affine matching colour image retrieval method based on the features of annual inspection label area. Moreover, regional colour constant and hue and saturation feature are introduced to the proposed method so as to settle the illumination change problem in the real surveillance scene. To fully evaluate the proposed algorithm, they perform experiments on the ReIDcar and VehicleID datasets, which differ in data scale and image quality. The experimental results show that the presented algorithm outperforms traditional methods in vehicle retrieval. It is verified that the extracted features and the feature matching method can distinguish subtle differences between vehicles.

1 Introduction

Nowadays, License plate has been one of the core research objects in the area of intelligent traffic systems for a long period of time [1]. However, license plates on vehicles are not always fully visible and not easy to recognise under certain situations. First of all, some surveillance cameras are not designed for license plate capturing, thus, plate recognition performance drops dramatically on images or video data captured by these cameras. Moreover, license plates are often easily occluded, removed or even faked in a large number of previous security events [2]. Therefore, vision-based vehicle retrieval has a great practical value in real-world surveillance applications. Specifically, vehicle retrieval is the problem of identifying the same vehicle across different surveillance camera views.

As an important part of intelligent vehicle information management system, vehicle retrieval has great prospects in many applications, such as intelligent parking lots management system, highway automatic charge, road monitoring, and parking timeout detection. However, retrieving vehicles in uncontrolled environments is a challenging task. From our knowledge, the difficulties mainly stem from two aspects: first, due to of the lack of high-quality and large-scale vehicle retrieval datasets [3], there is no previous attempt on the user level to retrieve the vehicle accurately purely in terms of visual appearance of the vehicle; second, vehicle images should be collected from multiple surveillance cameras at different time intervals and observation angles, which makes images inevitably affected by various interference factors, for instance of which includes illumination variation, haze, rain, and snow.

In spite of having been investigating the vehicle retrieval and re-identification in decades, in fact, most algorithms, having been announced, can only be applied in the field of vehicle retrieval by sharing the same coarse-level properties such as colour, size, model as well as brand, instead of the one exactly in the query image. By observing the visual characteristics of a large number of vehicle samples, it is generally found that there is a difference in the pasting position of the annual inspection label of each vehicle. In other words, the annual inspection label area is a unique feature of each vehicle, just like the vehicle number plate. To this end, we propose a new method for fast affine template matching of colour images to apply this special feature to fine-grained vehicle retrieval. The flow chart of the method is shown in Fig. 1, and an example of a special mark is shown in Fig. 2. The proposed method is extended based on the previous version [4] of the conference, which differs from the conference version as follows: first, we implemented the proposed algorithm on the Tensorflow platform using the Python language instead of the previous version of Matlab platform; second, we add an overview of the proposed retrieval algorithm to demonstrate the computational process more intuitively; third, we use four deep learning frameworks: Googlenet [3], OverFeat [5], VGG_CNN_M_1024 [2], and Resnet50 [6] to train different inspection networks during the vehicle detection phase, and divide the dataset into three grades according to the degree of difficulty and carry out experiments in four detection networks, respectively; fourth, we add three fine-grained feature

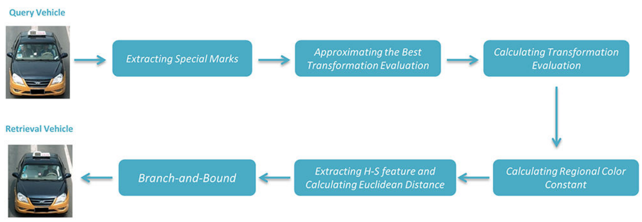

Fig. 1 *Flowchart of the proposed method*

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

Fig. 2 *Special marks for vehicle retrieval*

Fig. 3 *Different types of vehicle annual inspection labels*

extraction algorithms for vehicle retrieval; finally, the vehicle retrieval results of the proposed algorithm are more intuitively displayed. Compared with the traditional method, it has three advantages as follows:

- Although the original fast-match algorithm [7] has the advantages of being fast and efficient, there is still a problem with the loss of colour information. The main reason for this problem is that colour images must be converted to greyscale images to complete the template affine transformation and matching. Accordingly, we solve the problem of colour information loss in the original algorithm [7] by constructing colours-sum-of-absolute-difference of RGB three channels.
- As the time and place change, although the pixel colour of the vehicle annual inspection label is often affected by many interference factors, the colour change rate of different areas is constant. Therefore, colour constant information is applied to the proposed method, that is, the Laplacian formula is used to get the colour change rate of each colour region, thereby suppressing the adverse effect of the change of illumination conditions on the retrieval results.
- Based on conditions with constant recall rate, image accuracy can be improved in image retrieval algorithms by using hue and saturation (H–S) colour histograms. Comprehensive experimental results illustrate that the introduction of colour information contribute much to the improvement of vehicle retrieval accuracy.

For vehicles that are similar in their entirety, subtle special signs are essential to improve the accuracy of vehicle retrieval [8]. The vehicle inspection label has a variety of different colours and shapes, such as yellow oval, blue rectangle, and so on (as illustrated in Fig. 3). As we all know, the annual inspection of each vehicle is different and random, and the relationship between the labels is independent of each other; at the same time, the order and combination of the vehicle inspection labels will not change in the short term. Therefore, the feature of vehicle annual inspection is introduced in model matching, which makes the algorithm more accurate and effective.

2 Related work

In general, two core elements have influence on the accuracy of the vehicle retrieval problem, such as: how to extract features more effectively and to measure the distance between different image features more accurately. The method of image retrieval is

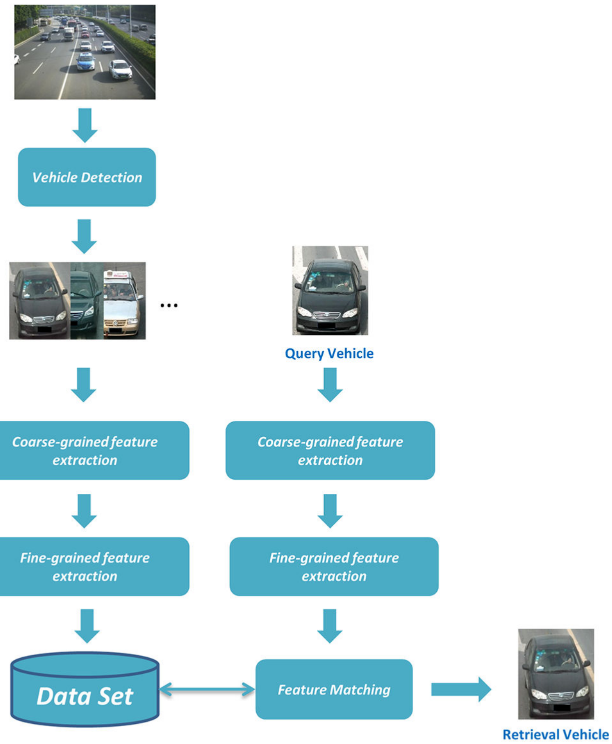

Fig. 4 *Flowchart of the vehicle retrieval system*

generally divided into direct and feature-based types [9]. For the first one, the excellent work by Baker and Matthews [10], in which the minimisation of the sum of the squared differences between two images is obtained by seeking the parameter optical flow mapping between the images. ASIFT [11] is a good example of a feature-based approach, which is designed to be affine invariant. However, these methods do not perform well on fine-grained processing, such as accurate retrieval which requires information other than attribute tags. Fast-match [7] handles the explosion by properly discretising the 2D affine transformations space. Its assumption based on image smoothness is the most important improvement, that is, the number of potential transformations evaluated can be bounded.

In order to deal with these challenges, there are a large number of relevant research results. Liu *et al.* [2] present a deep relative distance learning (DRDL) method, which introduces a bifurcated deep convolution network to map initial vehicle images to Euclidean space. We can measure the similarity between vehicles directly through the Euclidean space distance. Aiming at the problems of low accuracy and low recognition rate in the traditional method, an improved SURF method is proposed for vehicle video detection in [12]. In addition, Tsai [13] presents a two-level vehicle retrieval method based on depth feature coding. They construct and improve the deep convolutional network for extracting vehicle image features, and make use of the two-level retrieval strategy and similarity measure function to complete the paired retrieval of vehicle and vehicle brands. In 2016, Liang *et al.* [14] present a vehicle retrieval method based on multi-feature fusion such as the histograms of hue, saturation, and grey images, colour layout descriptor, perceptual hash hamming distance, and SIFT key point matching. In 2017, Liang *et al.* [15] propose a novel supervised deep hashing method to deal with large-scale instance-level vehicle searches. In 2018, Cheaon *et al.* [16] introduce spatio-temporal cubes into the smaller search chunks to solve the explosion growth of retrieval time.

Our work mainly constructs a new vehicle retrieval system shown in Fig. 4, which is different from the above mentioned algorithms. Compared with above methods, the improvements are reflected by the following three aspects. Firstly, we train a deep network based on Faster-RCNN to make sure the vehicle retrieval system can detect vehicles from a particular surveillance video. Secondly, vehicle colour recognition [17] is taken as the coarse-grained feature extraction part in the proposed method. Last but not

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

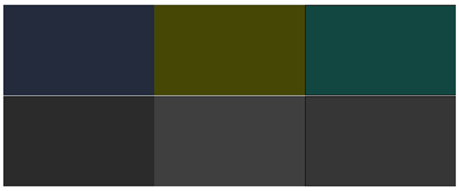

Fig. 5 *Converting a colour image directly to greyscale images will result in the loss of colour information and so reduce the accuracy of matching*

least, the most important improvement is that we are able to search the vehicle well at a fine-grained level, using a fast affine model matching method based on colour images, and combining the colour constants of special markers.

3 Proposed method

In this work, we propose a fast colour matching method for accurate vehicle retrieval. The overall framework of the paper is shown in Fig. 4. Our algorithm can be mainly divided into five steps. Firstly, we extracted the annual inspection labels from the target vehicle image. Secondly, the labels are used to approximate the best transformation evaluation and the transformation evaluation is calculated then. Thirdly, we calculated the regional colour constant. Fourthly, we extracted the H–S feature from the labels. Last but not least, the branch-and-bound method is applied for the target vehicle retrieval.

3.1 Preliminaries and the best transformation evaluation

The annual inspection label I_1 is extracted from the query vehicle and regarded as the template while another annual inspection label I_2 is obtained from image in the dataset. As explained in [7], the method to measure similarity distance between I_1 and I_2 is as follows: $\Delta_T(I_1, I_2)$ is the (normalised) sum-of-absolute-difference distance between template I_1 and image I_2, including a transformation T that maps pixels $p \in I_1$ to pixels in I_2 [18]. Mathematical formulas can be described as

$$\Delta_T(I_1, I_2) = \frac{1}{n_1^2} \sum_{p \in I_1} \left\| I_1(p) - I_2(T(p)) \right\|, \tag{1}$$

furthermore, when the input images I_1 and I_2 are coloured, they need to be converted to greyscale images. Therefore, formula (1) can be modified as

$$\Delta_T(I_1, I_2) = \frac{1}{n_1^2} \sum_{p \in I_1} \left\| \begin{array}{c} (I_1^R(p) - I_2^R(T(p))) \times 0.299 + \\ (I_1^G(p) - I_2^G(T(p))) \times 0.587 + \\ (I_1^B(p) - I_2^B(T(p))) \times 0.114 \end{array} \right\|. \tag{2}$$

Obviously, with the above conversion method, matching errors will occur in areas with similar but distinctive colours. For Fig. 5, the first row includes three colour patches each having an RGB value of (35, 43, 61), (71, 71, 5), (18, 71, 65), and the second row is the corresponding results of the greyscale image.

After greyscale image conversion, the red, green, and blue regions have the same intensity value 23, which results in being undistinguishable from each other in the greyscale image [18]. To solve the problem in Fig. 5, we calculate the absolute value of the colour difference between the image and the template in each channel, so formula (2) is modified as

Input: Color image I ;
Output: Score coefficient ΔG ;
Run the algorithm of density clustering DBSCAN [19] to get class type of each point;
Calculate class size n and class type number C_n;
Cumulate each class RGB value and number;
for $i = 1$ to n **do**
$sum(class(i), :) = sum(class(i), :) + I(i, :)$;
$num(class(i)) = num(class(i)) + 1$;
end for
Compute ΔG and return ΔG;
for $i = 1$ to C_n **do**
$C(i, :) = sum(i, :)/num(i)$;
$\Delta G = 1/num(i)$;
end for

Fig. 6 *Algorithm 1: Calculating score coefficient* ΔG

$$\Delta_T(I_1, I_2) = \frac{1}{n_1^2} \sum_{p \in I_1} W(I_1(p), I_2(T(p))), \tag{3}$$

$$W(p) = \left(\begin{array}{c} \left\| I_1^R(p) - I_2^R(T(p)) \right\| + \\ \left\| I_1^G(p) - I_2^G(T(p)) \right\| + \\ \left\| I_1^B(p) - I_2^B(T(p)) \right\| \end{array} \right) \times \Delta G(p), \tag{4}$$

where $\Delta G(p)$ is the difference coefficient between p and $T(p)$. In addition, $\Delta G(p)$ can be calculated by Algorithm 1 (see Fig. 6).

The affine transformation will not change the colour distribution of point. In other words, the corresponding point $T(p)$ will have similar colour distribution as p. Obviously, more colour information in the corresponding areas of I_1 means more precise in the similarity measures. If p is mapped out of the area I_2 and then $W(I_1(p), I_2(T(p)))$ is taken to be 1. Hence, formula (4) is modified as (see (5)) . Actually, we try to seek a transformation T which minimises $\Delta_T(I_1, I_2)$ as far as possible. Therefore, the net of transformations is a crucial part in fast-match algorithm. A small set of any affine transformations compose this net. Namely, L is the Euclidean distance between transforms T and T', which can quantise the mapping distance of any point p in I_1 according to T [7]. The mathematical model is as follows:

$$L(T, T') = \max_{p \in I_1} \left\| T(p) - T'(p) \right\|_2. \tag{6}$$

In the target image plane, $L(T, T')$ is the Euclidean distance between $T(p)$ and $T'(p)$. Particularly, this definition depends on the mappings T, T' and the dimension of the source image I_1 rather than the pixel values of the images. Furthermore, we can bound the differences between $\Delta_T(I_1, I_2)$ and $\Delta_{T'}(I_1, I_2)$ in terms of $L(T, T')$. It means that finite transform set is more efficient than the complete one [7]. In the proposed algorithm, we apply the δn_1-cover affine transformations, where $\delta \in (0, 1]$ is used as an accuracy parameter for the algorithm input. A fast randomised method is provided in

$$W(p, T(p)) = \begin{cases} \left(\begin{array}{c} \left\| I_1^R(p) - I_2^R(T(p)) \right\| + \\ \left\| I_1^G(p) - I_2^G(T(p)) \right\| + \\ \left\| I_1^B(p) - I_2^B(T(p)) \right\| \end{array} \right) \times \Delta G(p), & p \in I_2 \\ 1, & p \notin I_2 \end{cases}. \tag{5}$$

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

Input: Color images I_1,I_2 and a precision parameter δ;
Output: A transformation T;
1: Create a net $N_{\delta/2}$ that is a $\delta_{n_1}/2$-cover of the set of affine transformations;
2: For each $T \in N_{\delta/2}$ approximate $\Delta_T(I_1, I_2)$ to within precision of $\delta/2$.Denote the resulting value d_T;
3: Return the transformation T with the minimum value d_T.

Fig. 7 *Algorithm 2: Approximating the best transformation*

Input: Color images I_1 and I_2, a precision parameter l and a transformation T;
Output: An estimation of the distance $\Delta_T(I_1, I_2)$;
1: Sample $m = \Theta(1/l^2)$ [7] denotes values of pixels $p_1 \ldots p_m \in I_1$;
2: Return $\Delta_T(I_1, I_2) = \frac{1}{n_1^2} \sum_{p \in I_1} W\left(I_1(p), I_2\left(T(p)\right)\right)$.

Fig. 8 *Algorithm 3: Transformation evaluation*

the fast-match algorithm to obtain a transformation T with high probability. More importantly, the fast-match algorithm also examines the transformations in the net N_δ [7]. Meanwhile, a function of the net parameter δ and the total variation V are given to guarantee the quality of approximation.

As mentioned above, the best transformation T can be approximated by the net N, see Algorithm 2 (see Fig. 7) for details. Furthermore, we proceed to create a boundary of the difference between the quality of the algorithm results and that of the optimal transformation based on parameters V and δ. Here, δ also contributes to the control of the net size so as to determine the running time. After obtaining the transformation T, the distance metric $\Delta_T(I_1, I_2)$ is calculated, which is to measure the feature difference on different label images according to formula (4). To reduce the time complexity of Algorithm 2, we design a sub-linear algorithm, that is Algorithm 3 (see Fig. 8). It estimates the distance by a small fraction of the pixels in image, where the number of sample image pixels with regard to the precision parameter l, but not related to image size.

3.2 Regional colour constant and the H–S colour histogram

The improved fast-match algorithm contributes a lot in raising the recall rate of precise vehicle retrieval. In particular, it has a significant effect on matching accuracy of images with distinctive colour difference regions. However, for the vehicle retrieval in real scenarios, changes in illumination can cause colour shifts in annual inspection labels of vehicles. Therefore, regional colour constants and H–S colour histograms are introduced to improve the accuracy.

3.2.1 Regional colour constant: Annual inspection labels I_1 and I_2 are extracted from the query vehicle and the dataset, respectively. At first, the logarithm of the colour values in I_1 and I_2 are calculated. Then their derivatives are obtained by the Laplace formula so as to produce a new three tuple. The mathematical description is as follows:

$$J_k(x, y) = \ln(I_k(x, y)), \tag{7}$$

where $J_k(x, y)$ is the logarithm, $k = 1, 2, 3$, and $I_k(x, y)$ represents the value of R, G, B in (x, y)

$$D_{m,k}(x, y) = \nabla_m J_k(x, y), \tag{8}$$

where $D_{m,k}(x, y)$ is the derivative operator, $m = 1, 2, 3, 4$ represents four directions.

The new three tuple $H(d_{m1}, d_{m2}, d_{m3})$ can be constituted by formulas (7) and (8), representing the colour change rate in each colour space. Intuitively, even if the illumination changes result in the change of the image RGB values, the colour change rate in each colour space does not change significantly [20, 21]. The negative influence on the retrieval results caused by illumination change will be conquered.

Fig. 9 *Two annual inspection label examples extracted from the same vehicle but different images with illumination and viewpoint changes. The first column is the origin annual label images. The next three columns are the corresponding H, S, and V channel feature images*

3.2.2 H–S colour histogram: We use the features of the H–S channels in the HSV colour space to characterise the annual inspection labels of vehicles, because H and S channel features have been proven to have good illumination invariance. However, the V channel feature does not have this property. The results shown in Fig. 9 that it is effective to extract feature only in the H and S channels without S channel. We show such an influence more vividly from the original vehicle annual inspection label images and their feature images in the HSV three-channel, respectively.

3.3 Recognition and retrieval

In the latter recognition phase of the retrieval system [22, 23], we measured the distance between the fusion features obtained from the fast match, region colour constant, and H–S colour histogram. It is easily conceivable that if the extracted feature x is closest to another vehicle feature y, such a vehicle is considered to be the search result of the target vehicle [24, 25]. According to the shortest distance principle, the closest vehicle for the retrieval results can be obtained. The distance matric is described as follows:

$$d_{\text{hist}}^2(x, y) = (x - y)^{\mathrm{T}}(x - y), \tag{9}$$

where $d_{\text{hist}}^2(x, y)$ represents the distance between the query vehicle and the object vehicle. As mentioned earlier, vehicle colour recognition [17] is used as the coarse-grained feature portion in the proposed vehicle retrieval system, which reduces the range of fine-grained retrieval to a certain extent, thereby reducing the time loss.

3.4 Proposed retrieval algorithm

An overview of the proposed retrieval algorithm is listed in Algorithm 4 (see Fig. 10).

4 Experimental results

4.1 Datasets and experimental settings

The whole system is implemented on the Tensorflow platform. In our experiments, the vehicleID vehicle dataset is applied. It contains vehicle images captured by multiple surveillance cameras in real scene. More importantly, the properties of each vehicle are almost annotated well, such as the bounding box, model, colour, and license plate number. Each vehicle image in the dataset is assigned an ID based on its license plate number; a total of 237,393 images of 38,646 vehicles. For the reason that vehicle license plate is usually not a key feature in vehicle retrieval, the plates of all vehicles in the vehicleID are covered by a mask. In our experiments, we use vehicle colour and model features as coarse-grained attributes, with the colour divided into eight classes and the model divided into 250 classes. More importantly, the vehicle in the vehicleID dataset has a high recurrence rate, that is, the same vehicle has multiple images. Therefore, such a dataset is suitable for vehicle retrieval study. In addition, another smaller dataset, ReIDcar, is used in our experiments. It includes 9848 low-quality

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

Input: Color vehicle image I_1 and I_2;
Output: The distance d_{hist}^2 between I_1 and I_2;
Extract special marks $i_1 \in I_1$ and $i_2 \in I_2$;
for $i_1 \in I_1$ **do**
Calculate score coefficient ΔG using Algorithm 1;
    for $i_2 \in I_2$ **do**
Calculate the SAD distance between i_1 and i_2 using Equation (3) and (5);
Approximate the Best Transformation Evaluation using Equation (6);
Return the transformation T with the minimum value d_T using Algorithm 2;
Calculate Transformation Evaluation $\Delta_T(i_1, i_2)$ using Algorithm 3;
    end for
Calculate Regional color constant $H(d_{m1}, d_{m2}, d_{m3})$ using Equation (7) and (8);
Extract H-S feature and Calculate the distance d_{hist}^2 using Equation (9);
Return the minimum d_{hist}^2 as the Euclidean distance between I_1 and I_2;
end for

Fig. 10 *Algorithm 4: The proposed retrieval algorithm*

Fig. 11 *Eight colour vehicle images are divided from VehicleID and ReIDcar datasets. From top to bottom, the first row is black, blue, cyan, and green from left to right, respectively; the second row is grey, red, white, and yellow from left to right, respectively*

vehicle images affected by various disturbance factors, including shooting equipment, weather conditions, and so on. As is shown in Fig. 11, the two datasets are divided into eight classes for our experiments, which ensures the fairness of the experiments. It is advantageous to use the control variable method to verify each stage of the system.

4.2 Comparative results

4.2.1 Vehicle detection: The vehicle detection is an important part of the vehicle retrieval application system. The detection of a complete and clear vehicle is very important for extracting colour information and vehicle inspection mark information. Meanwhile, the time efficiency of vehicle detection has a significant impact on the effectiveness of the whole vehicle retrieval system. There are two main reasons as follows: first, vehicle detection is the most important part of the time consumption through experimental observation; second, vehicle detection process requires a good hardware environment. In summary, vehicle detection plays an important role in the whole vehicle retrieval system. The purpose of vehicle detection is to solve the problem of vehicle and background classification in frame image of road surveillance video. With regard to this, the latest research shows that the most effective method for vehicle detection is deep learning framework. Therefore, the deep learning framework is applied in the proposed algorithm to improve the accuracy of vehicle retrieval.

During the vehicle detection stage, we use the Resnet50 model [6] as the feature extractor to train the deep network based on Faster-RCNN, which can get a great capability to achieve high detection rate. We introduce three state-of-the-art models to perform the comparison. The first model GoogleNet is referred from paper [3]. In their paper, they utilise the GoogleNet to train a vehicle model classification model. From [5], the second model OverFeat is used to train a detection model. The VGG_CNN_M_1024 model [2] is used as the feature extractor to train the retrieval model in paper. In order to ensure the reliability of these experiments, we use the same training dataset including 200,000 vehicle images, which collected from the real traffic surveillance. The testing datasets contain three levels, as easy, medium, and hard. The confidence threshold is set to 0.8, and the threshold of non-maximum suppression is set to 0.2. Table 1 presents the final detection accuracy of the above methods and the best results are highlighted in bold. The Resnet50 model outperforms than other models in medium and hard datasets and the VGG_CNN_M_1024 model achieves the best performance in easy dataset. Moreover, compared with other three models, the Resnet50 model has lower computational complexity. Overall, we chose the Resnet50 model as the feature extractor to train the deep network. In the future work, we will try to use the deterministic annealing neural network algorithm [26] to achieve vehicle detection. The algorithm can try to find the global optimal solution strategy to obtain high-precision vehicle detection results under the condition of poor vehicle image quality.

4.2.2 Vehicle fine-grained retrieval: We conducted a large number of experimental comparisons in terms of colour fast matching, regional colour constants, and H–S colour feature. In order to verify the performance of colour fast match, we experiment with many existing methods, including the original fast match, OpenCV-based template matching, and the matching method presented in [17]. Meanwhile, we conduct experiments in these two datasets and analyse the performance of the proposed algorithm for different qualities of images. The effect of image quality on vehicle retrieval is validated in the experimental part of the paper. We divide the database into three levels: hard, medium, and easy according to image quality. However, in the practical application process, more detailed image quality assessment [27, 28] is another important and challenging research. The results of these experiments are shown in Table 2. The optimal results are highlighted in bold. As illustrated in Table 2, the OpenCV-based template matching method always performs the worst because it is not robust to illumination change. Experimental results confirm that our method outperforms the original fast-match method, which means that colour information is very important in the field of vehicle retrieval. In addition, the performance of the same method on the two datasets varies greatly due to the uneven image quality. In general, the quality of images in ReIDcar is relatively poor. Therefore, the presented method performs well in vehicle retrieval by using colour information.

We conduct three experiments to measure performance of regional colour constant and H–S colour feature. In the first experiment, without these two features, we extract templates from vehicle images and match them back to the datasets directly. In the second experiment, in order to verify the validity of the region colour constant, we extract templates from vehicle images with region colour constant and match them to the images to be retrieved. In the last experiment, we extract the H–S colour feature from the annual inspection label images and map them to the images in the two datasets. In addition, although our proposed vehicle retrieval system includes two main steps of coarse-grained retrieval and fine-grained retrieval [29, 30], here we mainly study the performance of fine-grained retrieval. That is to say, our fine-grained retrieval is based on the condition that the coarse-grained retrieval results are known. We design four fine-grained retrieval experiments to evaluate the proposed algorithm [31, 32]. All experiments are carried out under known conditions in which the vehicles are divided into eight colours. The experimental results are presented in Fig. 12.

As can be seen from the accuracy of Fig. 12, our method outperforms other methods for vehicles of all colours. In addition, the method of missing regional colour constant and HS colour feature is significantly less accurate, which indicates that regional colour constant and HS colour features helps to heighten the accuracy of fine-grained vehicle retrieval.

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

Table 1 Vehicle detection accuracy of several models on VehicleID

Accuracy	GoogleNet	OverFeat	VGG_CNN_M_1024	Resnet50
easy	0.920	0.891	**0.945**	0.921
medium	0.834	0.823	0.929	**0.942**
hard	0.796	0.745	0.842	**0.896**
average time	171	181	158	**143**

The optimal results are highlighted in bold. The average time is in milliseconds per frame.

Table 2 Vehicle retrieval accuracy of different match methods on VehicleID and ReIDcar

Accuracy	OpenCV	Original fast-match [7]	Method by [17]	Our method
vehicleID	0.603	0.746	0.744	**0.822**
reIDcar	0.655	0.739	0.716	**0.812**
average Time	120	64	52	**43**

The optimal results are highlighted in bold. The average time is in milliseconds per frame.

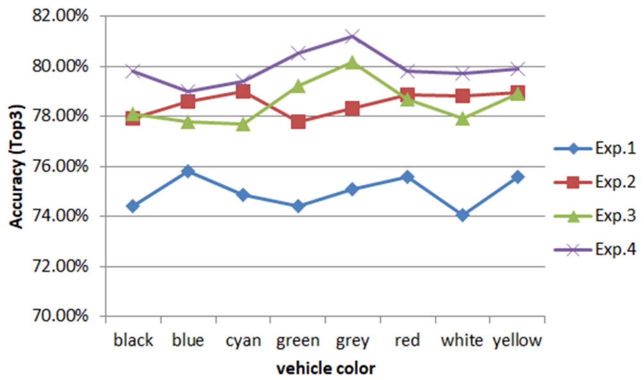

Fig. 12 *Ablation comparison of using four different fine-grained retrieval methods on eight colour vehicle subsets. Exp. 1 to 4 correspond to four methods, namely without region colour constant and H–S colour feature, with region colour constant, with H–S colour feature, and with both two features above*

Table 3 Comparison of our method with other six state-of-the-art methods on VehicleID and ReIDcar datasets in terms of vehicle retrieval accuracy

Accuracy	Deep hashing [15]	Method by [13]	DRDL [2]	Method by [33]	Method by [34]	Method by [21]	Our method
black	0.745	0.779	0.781	0.623	0.687	0.765	**0.796**
blue	0.756	0.785	**0.789**	0.650	0.660	0.719	0.776
cyan	0.749	0.789	0.775	0.698	0.682	0.673	**0.793**
green	0.746	0.776	0.805	0.690	0.702	0.721	**0.792**
grey	0.751	0.783	0.802	0.702	0.680	0.766	**0.812**
red	0.754	0.789	**0.796**	0.671	0.704	0.677	0.787
white	0.741	0.787	0.778	0.682	0.659	**0.799**	0.794
yellow	0.753	0.790	0.789	0.686	0.632	0.707	**0.798**

The optimal results are highlighted in bold.

4.2.3 Vehicle retrieval algorithm comparison: In order to evaluate our proposed method more comprehensively, other six state-of-the-art vehicle retrieval algorithms are carried out as comparative experiments. The comparison results are shown in Table 3.

The experimental results in Table 3 show that the DRDL framework achieves the best performance on the blue and red subsets, while the proposed method is best in the other five subsets [35, 36]. Although our method is far from perfect, it gets the first place in black, cyan, green, grey, and yellow datasets. We designed these experiments to measure how much improvement the full use of vehicle annual inspection labels in our framework brings. Our method takes advantage of the vehicle annual inspection labels to get fine-grained features. Compared with other three retrieval methods, the improvement of our method is that the colour information are both extracted from vehicle annual inspection labels and the whole vehicles rather than simply gotten from coarse-grained features like three other methods [21, 33, 34].

4.3 Experimental analysis

In order to be more intuitively, Fig. 13 shows three query vehicle images and their retrieval results including the top three vehicles with the highest similarity. As can be seen, it is robust to interference factors, such as vehicle colour and vehicle type. However, when faced with complex environmental conditions, our proposed vehicle retrieval system did not get the best performance. First of all, when vehicle pictures are too vague to identify vehicle annual inspection labels, our vehicle retrieval framework cannot extract the fine-grained features effectively. This will lead to the decrease of accuracy in template matching phase. In the next place, we cannot guarantee that there are no similar vehicle inspection labels for two cars of the same type and colour. When the vehicles are all of the same colour, after calculating the colour similarity of the vehicles, the proposed algorithm classifies these vehicles according to the fine-grained characteristics of the vehicles inspection signs. Vehicle inspection signs are the decisive factor in classifying vehicles of the same colour. Due to the randomness of the vehicle inspection signs, each vehicle can be distinguished according to the matching rate of the vehicle inspection signs. The identical coarse-grained features and the analogous fine-grained

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

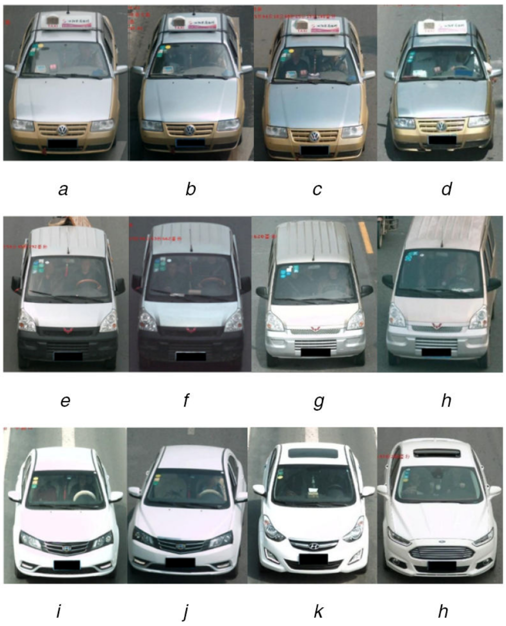

Fig. 13 *Vehicle retrieval examples by the proposed algorithm. These examples are the retrieval results of the vehicle retrieval system. (a), (e), and (i) as query vehicles are extracted from particular surveillance videos. Each query vehicle will get the top three vehicles with the highest similarity in our datasets*

features will have a negative effect on the retrieval results. For the sake of the above problems, our framework should improve the fineness of approximating and calculating the best transformation evaluation in the future work. Meanwhile, introducing the triplet loss in the step of calculating distance metric is an effective method to compare the identical coarse-grained features and the analogous fine-grained features.

4.4 Limitation analysis

The proposed algorithm has some limitations. First, for bad weather and poor monitoring equipment, the quality of the obtained vehicle image is not ideal. That is to say, the success rate of vehicle retrieval is lower in this case, which gives our work brings challenges and is the focus of future work. Second, due to the different traffic regulations between the regions, the vehicle annual inspection sign used in the vehicle retrieval are different, that is to say, the fine-grained features extracted are different, the proposed algorithm cannot be used to detect vehicles in all regions, which is only suitable for mainland China. The vehicle annual inspection sign is an important method of traffic control in China. The fine-grained features used in the proposed algorithm mainly depend on vehicle annual inspection sign.

5 Conclusions

To solve the vehicle fine-grained retrieval problem, we present an improved fast affine matching method combining the regional colour constant and H–S colour feature of the vehicle annual inspection label for vehicle fine-grained retrieval. We perform effective and extensive experiments on two datasets with up to one million vehicles. Compared with several advanced current methods, the application of vehicle annual inspection labels contributes to the higher predict accuracy. Comprehensive experiments illustrate that the presented method outperforms other traditional methods. From the above, we believe that the proposed method can be widely implemented in a vehicle retrieval system.

6 Acknowledgments

This work was supported in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province KYCX19_0886, in part by the Peak of Six Talents in Jiangsu Province under grant RLD201402, in part by the 1311 Talent Program of NJUPT.

6 References

[1] Bulan, O., Kozitsky, V., Ramesh, P.*, et al.*: 'Segmentation-and annotation-free license plate recognition with deep localization and failure identification', *Trans. Intell. Transp. Syst.*, 2017, **18**, (9), pp. 2351–2363

[2] Liu, H., Tian, Y., Wang, Y.*, et al.*: 'Deep relative distance learning: tell the difference between similar vehicles'. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 2167–2175

[3] Yang, L., Luo, P., Loy, C.C.*, et al.*: 'A large-scale car dataset for fine-grained categorization and verification'. Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015, pp. 3973–3981

[4] Liu, F., Wang, Y., Wei, J.*, et al.*: 'Vehicle precise retrieval via color image retrieval method based on improved fast-match'. 2018 10th Int. Conf. on Wireless Communications and Signal Processing (WCSP), Hangzhou, China, October 2018, pp. 1–7

[5] Sermanet, P., Eigen, D., Zhang, X.: 'Overfeat:integrated recognition, localization and detection using convolutional networks', arXiv, 2013

[6] Jung, H., Choi, M.K., Jung, J.: 'Resnet-based vehicle classification and localization in traffic surveillance systems'. Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, July 2017, pp. 934–940

[7] Korman, S., Reichman, D., Tsur, G.*, et al.*: 'Fast-match: fast Affine template matching'. Proc. of the 2013 IEEE Conf. on Computer Vision and Pattern Recognition, Portland, OR, USA, June 2013, pp. 2331–2338

[8] Du, S., Ibrahim, M., Shehata, M.: 'Automatic license plate recognition (ALPR): A state-of-the-art review', *IEEE Trans. Circuits Syst. Video Technol.*, 2013, **23**, (2), pp. 311–325

[9] Datta, R., Joshi, D., Li, J.*, et al.*: 'Image retrieval: ideas, influences, and trends of the new age', *ACM Comput. Surv. (Csur)*, 2008, **40**, (2), p. 5

[10] Baker, S., Matthews, I.: 'Lucas-Kanade 20 years on: a unifying framework', *Int. J. Comput. Vis.*, 2004, **56**, (3), pp. 221–255

[11] Morel, J.M., Yu, G.: 'ASIFT: a new framework for fully affine invariant image comparison', *Siam J. Imaging Sci.*, 2009, **2**, (2), pp. 438–469

[12] Zhang, Z., Jing, X., Qiao, H.*, et al.*: 'The vehicle retrieval methods of traffic video based on improved SURF algorithm', *J. Northwestern Polytechnical Univ.*, 2014, **32**, (2), pp. 297–301

[13] Tsai, T.-H., Chang, W.-C.: 'Two-stage method for specific audio retrieval based on MP3 compression domain'. Proc. of the 2009 IEEE Int. Symp. on Circuits and Systems, Taipei, Taiwan, May 2009, pp. 713–716

[14] Liang, Z., Guo, Q., Hu, J.: 'A vehicle retrieval method based on multi-feature fusion', *Informatization Res.*, 2016, **42**, (4), pp. 51–58

[15] Liang, D., Yan, K., Wang, Y.*, et al.*: 'Deep hashing with multi-task learning for large-scale instance-level vehicle search'. Proc. of the 2017 IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW), Hong Kong, China, July 2017, pp. 192–197

[16] Cheong, C.W., Lim, R.W.S., See, J.: 'Vehicle semantics extraction and retrieval for long-term carpark video surveillance'. Proc. of the 2018 24th Int. Conf. on Multimedia Modeling, Bangkok, Thailand, February 2018, pp. 315–326

[17] Hu, C., Bai, X., Qi, L.*, et al.*: 'Vehicle color recognition with spatial pyramid deep learning', *IEEE Trans. Intell. Transp. Syst.*, 2015, **16**, (5), pp. 2925–2934

[18] Jia, D., Cao, J., Song, W.D.: 'Colour FAST (CFAST) match: fast affine template matching for colour images', *Electron. Lett.*, 2016, **52**, (14), pp. 1220–1221

[19] Yang, Y., Lu, Z., Sundaramoorthi, G.: 'Coarse-to-fine region selection and matching'. Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015, pp. 5051–5059

[20] Yue, J., Li, Z., Liu, L.: 'Content-based image retrieval using color and texture fused features', *Math. Comput. Model.*, 2011, **54**, (3–4), pp. 1121–1127

[21] Liu, G.H., Yang, J.Y.: 'Content-based image retrieval using color difference histogram', *Pattern Recognit.*, 2013, **46**, (1), pp. 188–198

[22] Momin, B.F., Mujawar, T.M.: 'Vehicle detection and attribute based search of vehicles in video surveillance system'. Proc. of the 2015 Int. Conf. on Circuits, Power and Computing Technologies, Nagercoil, India, March 2015, pp. 1–4

[23] Karaimer, H.C., Cinaroglu, I., Bastanlar, Y.: 'Combining shape-based and gradient-based classifiers for vehicle classification'. Proc. of the 2015 IEEE 18th Int. Conf. on Intelligent Transportation Systems, Las Palmas, Spain, September 2015, pp. 800–805

[24] Zhao, R., Ouyang, W., Wang, X.: 'Learning mid-level filters for person re-identification'. Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, OH, USA, June 2014, pp. 144–151

[25] Krause, J., Stark, M., Deng, J.*, et al.*: '3d object representations for fine-grained categorization'. Proc. of the 2013 IEEE Int. Conf. on Computer Vision Workshops, Sydney, NSW, Australia, December 2013, pp. 554–561

[26] Wu, Z., Karimi, H.R., Dang, C.: 'An approximation algorithm for graph partitioning via deterministic annealing neural network', *Neural Netw.*, 2019, **117**, pp. 191–200

[27] Min, X., Gu, K., Zhai, G.: 'Blind quality assessment based on pseudo-reference image', *IEEE Trans. Multimed.*, 2017, **20**, (8), pp. 2049–2062

[28] Min, X., Zhai, G., Gu, K.: 'Blind image quality estimation via distortion aggravation', *IEEE Trans. Broadcast.*, 2018, **64**, (2), pp. 508–517

[29] Chai, Y., Lempitsky, V., Zisserman, A.: 'Symbiotic segmentation and part localization for fine-grained categorization'. Proc. of the 2013 IEEE Int. Conf. on Computer Vision, Sydney, NSW, Australia, December 2013, pp. 321–328

[30] Zhang, N., Farrell, R., Iandola, F.*, et al.*: 'Deformable part descriptors for fine-grained recognition and attribute prediction'. Proc. of the 2013 IEEE Int. Conf. on Computer Vision, Sydney, NSW, Australia, December 2013, pp. 729–736

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

[31] Qian, Q., Jin, R., Zhu, S., *et al.*: 'Fine-grained visual categorization via multi-stage metric learning'. Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015, pp. 3716–3724
[32] Zhang, X., Zhou, F., Lin, Y., *et al.*: 'Embedding label structures for fine-grained feature representation'. Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 1114–1123
[33] Zhang, C., Wang, X., Feng, J.: 'A car-face region-based image retrieval method with attention of SIFT features', *Multimedia Tools Appl.*, 2017, **76**, (8), pp. 1–20
[34] Hu, M., Zhang, M., Lou, Y.: 'Retrieval of vehicle images based on color space fuzzy quantification in criminal investigation', *Int. J. Performability Eng.*, 2017, **13**, (6), pp. 823–831
[35] Russakovsky, O., Deng, J., Su, H.: 'Imagenet large scale visual recognition challenge', *Int. J. Comput. Vis. (IJCV)*, 2015, **115**, (3), pp. 211–252
[36] Zhang, Z., Tan, T., Huang, K., *et al.*: 'Three dimensional deformable-model-based localization and recognition of road vehicles', *IEEE Trans. Image Process.*, 2012, **21**, (1), pp. 1–13

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)