Car manufacturer and model recognition based on scale invariant feature transform

Yongbin Gao
Division of Computer Science and Engineering,
Chonbuk National University,
567 Baekje-Daero, Deokjin-Gu,
Jeonju 54596, Republic of Korea
Email: gaoyongbin.sam@gmail.com

Hyo Jong Lee*
Division of Computer Science and Engineering,
Center for Advanced Image and Information Technology,
Chonbuk National University,
567 Baekje-Daero, Deokjin-Gu,
Jeonju 54596, Republic of Korea
Email: hlee@chonbuk.ac.kr
*Corresponding author

Abstract: Vehicle analysis involves licence plate recognition, vehicle type recognition, and car manufacturer and model recognition. Car manufacturer and model recognition plays an important role in providing supplementary information to licence plate recognition for the unique identification of a car. In this paper, we propose a framework to recognition car manufacturer and its model based on scale invariant feature transform (SIFT). We first detect a moving car using frame differences; the resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is then used to identify a car based on SIFT algorithm. Experimental results show that our proposed framework achieves favourable recognition accuracy.

Keywords: moving car detection; car model recognition; scale invariant feature transform; SIFT.


Biographical notes: Yongbin Gao received his BS and MS degrees in School of Information Technology from Jiangxi University of Finance and Economics, China in 2010 and 2013, respectively. Since September 2013, he is with the School of Computer Science and Engineering from Chonbuk National University as a PhD candidate.

Hyo Jong Lee received his MS and PhD degrees in Computer Science from the University of Utah, specialised in computer graphics and parallel processing. He is currently a Professor in Division of Computer Science and Engineering and Director of the Center for Advanced Image and Information Technology, Chonbuk National University. His research interests are image processing, medical imaging, visualisation, and parallel programming.

Copyright © 2018 Inderscience Enterprises Ltd.
This paper is a revised and expanded version of a paper entitled ‘Moving car detection and model recognition based on scale invariant feature transform’ presented at International Conference on Green and Human Information Technology (ICGHT), Da Nang, Vietnam, 4–6 February 2015.

1 Introduction

Vehicle analysis is an important task for various intelligent applications, such as automatic toll collection, driver assistance systems, intelligent parking, vehicle self-guidance, and traffic monitoring (i.e., vehicle speed, flow and count). Specially, an electronic toll collection system can automatically collect tolls by identifying vehicle models. The identification of vehicle models can also provide valuable information to law enforcement agencies searching for suspected vehicles. The visual design of a vehicle will change under varying environmental conditions and market requirements. The shapes of vehicles between auto manufacturers and models are very similar, which is confusing for vehicle model recognition. This makes the vehicle model recognition a challenging task.

The detection and matching of interest points serve as the basis for many computer vision applications including image/video retrieval, object categorisation and recognition, and 3-D scene reconstruction directions. Two of the most widely used detectors are the Harris corner detector (widely used in Europe) and the Kanade-Lucas-Tomasi corner detector (used in the USA). However, these corner detectors are not invariant to scale and affine transformations. To address this problem, Lowe (2004) approximated the Laplacian matrix of Gaussians with a difference of Gaussian (DOG) filter to propose a rotation-invariant descriptor called SIFT, which computes a histogram of locally oriented gradients around the interest points and stores into bins of a 128-D vector.

Vehicle detection is a prerequisite of vehicle analysis. There are many methods proposed for vehicle detection. Most of these proposed approaches (Faro et al., 2011; Unno et al., 2007; Jazayeri et al., 2011; Foresti et al., 1999) use background subtraction to extract motion features to detect moving vehicles from videos. However, this motion feature is no longer usable and is not available in still images. To address the same issues for still images, Wu et al. (2001) used wavelet transformation to extract texture features to locate possible vehicle candidates from the road. Each vehicle candidate was then verified using a PCA classifier. Tzomacas and Seelen (1998) found that the shadow underneath a vehicle is a good cue for detecting vehicles. Ratan et al. (2000) developed a scheme for detecting vehicle wheels to locate possible vehicles and then used a diverse density method to verify each vehicle candidate.

There are various applications of vehicle detection. Chen et al. (2011) proposed a model-based approach to classify vehicles on the road into four classes – car, van, bus, and bicycle/motorcycle – using SVM and random forests. Ma and Grimson (2005) used edge points and modified SIFT descriptors to represent vehicles and then modified a constellation model to classify them into two classes, namely, cars versus minivans, and sedans versus taxis. In these applications, variations between the classes are quite wide, while for car model recognition, the variation in appearance between different models is very low. AbdelMaseeh et al. (2012) proposed a hybrid method of global and local cues
to recognise a car’s make and model. Hsieh et al. (2014) proposed symmetrical SURF for both vehicle detection and model recognition. Gao and Lee (2015a) proposed to use deep learning for identify the vehicle model and achieved good performance.

As a feature extraction method, scale invariant feature transform (SIFT) achieves better performance for various applications (Gao and Lee, 2015b), which have properties of scale and rotation invariance. While it is not invariant towards the affine transformation, Affine SIFT was proposed to provide the affine invariance (Morel and Yu, 2009). The SIFT algorithm has been combined with principal component analysis (PCA) to improve accuracy. That is, the PCA is used to alter the histogram to represent the feature descriptor, which shows superior performance and robustness (Yan and Sukthankar, 2004). Another variation uses the speeded up robust features (SURF) algorithm: the SURF reduces the computational time of SIFT to make it more practical in real life applications (Zurich and Leuven, 2008).

In this paper, we propose a framework to detect moving cars and their models based on SIFT. We first detect the moving car using frame differences; the resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is then used to identify a car based on SIFT algorithm. The framework of the proposed system is presented in Figure 1. It is noted that frontal view suffers less variance than a view of an entire car. Therefore, in this paper, we prefer to use the frontal view instead of a view of an entire car.

**Figure 1** Framework of proposed car detection and model recognition system based on SIFT (see online version for colours)

![Framework of proposed car detection and model recognition system based on SIFT](image)

Note: A moving car is detected by frame difference; the resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is then used to identify a car based on SIFT algorithm.

The remainder of this paper is organised as follows. Section 2 describes the moving car detection and frontal view extraction method. We then introduce the car model recognition based on SIFT in Section 3. Section 4 applies the above algorithm to our car...
database, and presents the experimental results. Finally, we conclude this paper in 
Section 5.

2 Moving car detection and frontal view extraction

Detecting a car in motion is the baseline of car analysis. Frame difference is an effective 
and simple algorithm for moving car detection, which is used in our car model 
recognition system due to the fact that our camera is fixed on the street with a simple 
background. In this case, frame difference is sufficient for moving car detection. It 
requires short computational time which enables a real time system.

3D object recognition is a difficult task because of its large variance due to angle 
changes within one class. In this paper, we use the frontal view of a car to recognise the 
car model base on two considerations:

1 The frontal view of a car is planar, which is easier to recognise than a 3D object. 
   Also, many papers exhibit considerable performance by using the rear or frontal 
   view of a car for car model recognition 

2 The frontal view of a car reduces the computational time as the size of a frontal view 
   is much smaller than an image of an entire car.

Considering the symmetrical property of a frontal view of a car image, we use a 
symmetrical filter to extract the symmetrical region from the binary image of a car, and 
consider the symmetrical region as the frontal view of a car. The symmetrical filter sums 
up the values of pixels in the same column, then the car image results in a vector of the 
same length as the width of the car image. After that, the symmetrical filter is performed 
on the vector with a window the size of a car’s approximate width, which calculates the 
difference between the left and right portions of the window. The window with the 
minimal difference is regarded as the frontal view of a car.

3 Scale invariant feature transform

SIFT is a local descriptor, which is scale and rotation invariant. Images are transformed 
into scale-invariant space and localised keypoints. Each of the keypoints is assigned to 
one or multiple descriptors. SIFT algorithm comprises of the following four steps (Lowe, 
2004):

1 Scale-space extrema detection: various scales and sizes are generated from the input 
   image. Extrema are first searched by seeking the maxima and minima over all scales 
   and sizes based on the DOG that is scale and orientation invariant.

\[
D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)
\]  

where \(D(x, y, \sigma)\) is the DOG function, \(G\) is the Gaussian function and \(k\) is the factor 
of nearby scales. \(I\) is the input image. Extrema are localised by comparing the value 
of a pixel with its 26 neighbours at the current scale and two adjacent scales, which 
are 8, 9, 9 pixels for each scale as shown in Figure 2. These extrema serve as the 
candidate keypoints.
Keypoint localisation: to increase the stability of keypoint localisation, the extrema are refined by excluding low contrast and poor localised pixels. This is implemented by examining the refined location and scale as well as the ratio of principal curvatures.

Orientation assignment: one or more orientations are assigned to each keypoint based on the histogram of the local image gradient. Considering the scale and rotation invariance, the local image data are transformed to the assigned orientation and scale for further calculation of keypoint descriptor.

Keypoint descriptor: keypoint descriptor is calculated in a local spatial region centred at the keypoint based on the histogram of the gradients. The descriptor is characterised to allow significant local shape distortion and variance in illumination. As detailed, the keypoint descriptor is calculated by the gradient and orientation in a region around the keypoint that is weighted by a Gaussian window. The region composed of $4 \times 4$ subregions, and for each subregion, a histogram of orientation with eight bins is calculated. Each bin of the histogram corresponds to the summation of the gradient magnitudes that is close to the orientation of the bin.

Once the keypoints are calculated for two images, we are able to calculate the number of matching keypoints between the two similar images. There are many researches on image matching and recognition for SIFT algorithm, such as BBF (Beis and Lowe, 1997), and Hough transform (Hough, 1962). However, the nearest neighbour is the primitive algorithm for matching, and it is simple and effective for real time application.

Figure 2  Illustration of DOG, octaves from bottom to top are generated by down-sampling

Note: The initial image is convolved with Gaussian filter using different scales for each octave. DOG images is generated from this Gaussian filtered image. Extrema are localised by finding the maxima and minima compared to neighbouring pixels in the current scale and adjacent scales as shown on the right.
As for car model recognition, 30 images were collected for each car model. The images of the cars are divided into gallery images and probe images. We first calculate the keypoints for all the images in the gallery for each car model. The keypoints of a probe image are compared with the keypoints of each car model as shown in Figure 3. The nearest neighbour and the second nearest neighbour are ranked according to the Euclidean distance. The ratio of these two distances is compared with a threshold that is defined to determine whether two keypoints are matching. A ratio that is smaller than the threshold is regarded as matching keypoints, and the similarity of two car model images is determined by the number of matching keypoints. The car model with the maximum number of matching keypoints is considered the recognised car model.

Figure 3  Car model recognition based on SIFT (see online version for colours)

![Figure 3](image)

4 Results

A car model dataset is built to evaluate the performance of the proposed system. The dataset comprised of 3,210 car images, which includes 107 different car models and 30 samples for each model. Images are easier for measuring accuracy than videos, while frame difference is used in our framework. Thus, ten pixels are shifted for each car image to virtualise a neighbouring image. The difference image is achieved over the image and corresponding shifted image.

Figure 4 shows some sample results of frontal view detection, which involves five car models from four manufacturers. The left column shows the car images with detected frontal views of the cars marked with a red rectangle, while the right ones show the binary frontal views of cars. Experimental results show that our framework is capable of detecting frontal views of car images perfectly in cases where the camera is static on the street without a complex background. The performance in terms of detection accuracy is 100% over the 3,210 images in our dataset. Thus, utilising the frame difference followed by a symmetrical filter is an effective and fast method for extracting frontal views of cars.

As the last step of our framework, car model recognition is performed by comparing the frontal view of a probe image to car model databases with labelled models using SIFT algorithm. As we know, there are 30 images for each car model, and we divided them
into a gallery set and probe set. A probe image is compared with all the images in the gallery set to get a maximum number of matching keypoints as shown in Figure 3. By varying the number of images in the gallery set, we achieve the performance of car model recognition as shown in Figure 5. The accuracy of car model recognition increases with the number of compared images. We also compared our algorithm with widely used local descriptors as shown in Table 1, which includes local binary pattern (LBP) (Ahonen et al., 2006) and local Gabor binary patterns (LGBP) (Zhang et al., 2005). The experimental results demonstrate that SIFT is better than the other algorithms.

Figure 4  Results of frontal view extraction from sample images with four manufacturer and five models (see online version for colours)
5 Conclusions

In this paper, we propose a framework to detect vehicle makes and models based on SIFT. We first detect a moving car using frame difference. The resultant binary image is used to detect the frontal view of a car by a symmetry filter. The detected frontal view is then used to identify a car based on SIFT algorithm. Experimental results show that our proposed framework achieves higher recognition accuracy. The proposed framework has proven efficiency in cases where the camera is fixed on the street. The frame difference and symmetrical filter works efficiently and in real-time to detect moving vehicles. The accuracy of detecting moving vehicles was 100% in the actual data collected on real streets. The accuracy of the proposed model recognition algorithm is 25% higher than the LBP algorithm, although it shows slightly improved accuracy compared to the LGBP method.
Acknowledgements

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea. This work was also supported by the Business for Academic-Industrial Cooperative establishments that were funded by the Korea Small and Medium Business Administration in 2015 (Grants No. C0221114). This research was also supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R0992-16-1023) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References


