

VEHICLE MAKE AND MODEL RECOGNITION BY KEYPOINT MATCHING OF PSEUDO FRONTAL VIEW

Yukiko Shinozuka, Ruiko Miyano, Takuya Minagawa and Hideo Saito

Department of Information and Computer Science, Keio University
{shinozuka, rui, takuya, saito}@hvrl.ics.keio.ac.jp

ABSTRACT

We propose a vehicle make and model recognition method for the smart phone, and implemented it onto them. Our method identifies the make and model from the variety of viewpoints while the conventional methods for VMMR work only for the frontal or rear view images. This method enables the users to take the pictures from a certain range of angles. Our method uses SIFT, that has scale and rotation invariance to solve the alignment issue. It creates the pseudo frontal view images by homography matrix and extracts the keypoints. Homography matrix is calculated with the position of the license plate. Our experimental result shows our method enables to recognize up to 60-degree angle.

Index Terms— Vehicle Make and Model Recognition, Image Retrieval, Keypoint Matching, View Invariance, SIFT

1. INTRODUCTION

In recent years, prevailing of the smart phones is increasing the demand of Web search applications. The users demand is to get information of the products by simply taking a picture and searching through image retrieval systems such as Google Goggles[1] and A9[2] etc. Image retrieval systems often recognize the object by extracting keypoints and comparing their features. It outputs the similar images. If this image retrieval system is more improved, the users can search anything around the world. It must promote the commercial industry. There is a demand for automobile industry to link to products' information, including its performance, price, users' certificates or other similar products.

On the other hand, vehicle detecting system is well-known in study of intelligence transport systems (ITS). Most of them in Computer Vision are vehicle detection, vehicle classification, license plate recognition or vehicle tracking. However, there are few papers about vehicle make and model recognition (VMMR). Difficulties in VMMR are 1. Illumination conditions highly influence its metallic surface. 2. There are only a few keypoints extracted on its body. 3. Images vary

from viewpoints since it is a 3D object. The view point invariance must be one of the most important issues in the case of smart phone usage since the users can move around the object.

In this paper, we are focusing on the second and third issues in a 2D-2D vehicle retrieval system. We conducted the pre-experiments to select the suitable feature and verify the region of interest (ROI) for VMMR. Our method uses SIFT to solve the alignment issue. To solve the second issue, the front area is defined as ROI, as the pre-experiment shows it has enough information and discriminative power to solve the second issue. To solve the third issue, our system transforms the query images to the pseudo frontal images by homography matrix and the database stores only frontal view images.

In this paper, we implemented a vehicle retrieval system onto the smart phone. The contribution of this paper is for the users to take the pictures from variety of view points. Our result shows higher performance in recognition even with the angled view images.

This paper is structured as follows. Section 2 describes the related works. Section 3 describes technical difficulties in VMMR. Section 4 describes the dataset and the proposed methods. Section 5 describes the experiments and its results. Section 6 shows our implementation of our method onto the smart phones. Section 7 mentions the conclusion.

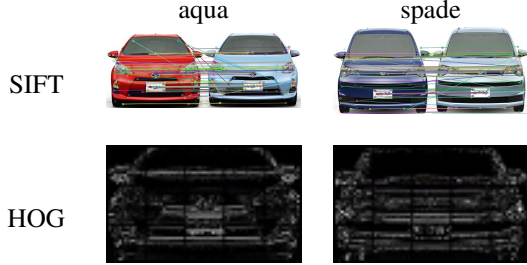
2. RELATED WORKS

Scale Invariant Feature Transform (SIFT) [3] is often used in the image retrieval system, for it is scale and rotation invariant feature detector and descriptor. This method is useful when the alignment issue occurs. Maximally-Stable Extremal Region Detector (MSER) [4], Harris-affine [5] and Hessian-affine [5] are known for the invariance to affine distortion, though there is no descriptor for each of them. Even if the keypoints are extracted, the feature description will be different from the ones extracted on the image before transformation. Therefore the keypoint matching will fail. Keypoint matching using randomized tree [6] is also useful for affine distortion, but learning process should be conducted beforehand. ASIFT [7] is the closest approach to our method. It conducts too many transformations, so we simplify the process and transform only one time.

This work was partially supported by MEXT/JSPS Grant-in-Aid for Scientific Research(S) 24220004.

Table 1. Average Ranking in SIFT and HOG

models	SIFT	HOG	models	SIFT	HOG
aqua	1.0	21.0	auris	1.0	3.1
corolla fielder	1.0	19.9	markx	1.0	10.8
porte	1.0	18.8	prius	1.0	8.4
prius alpha	2.1	13.6	spade	1.0	27.1
wish	1.0	17.3	total	1.1	15.6

**Fig. 1.** Visualization of SIFT and Difference of HOG

For VMMR to deal with the illumination issue shown in section 1, Psyllos et al. [8] method does multiple process for identification, including the measurement of the vehicle height and width using the edges. In their method, they need to have prior knowledge of the vehicle shape. To solve the keypoints issue, query images have to be the frontal ones. Therefore it is not invariant to viewpoint changes. The method of Han et al. [9] uses surveillance video to reconstruct 3D model to deal with the 3D object issue, but it has to store CAD data in database to compare. Some methods for 3D recognition store many images, which is taken from variety of angles. But in those methods, more numbers of database images are needed to raise the accuracy of recognition, which is discussed in H.Yabushita et al. work [10].

Our system can identify the make and model even with the angled view images. It uses SIFT for alignment and prior knowledge issue. To increase the invariance to the view point changes, it creates the pseudo frontal images while it only stores the frontal images in database.

3. DIFFICULTIES IN VMMR

In this section, we refer to the difficulties in VMMR. Section 3.1 shows the comparison of the features. Section 3.2 discusses the ROI for VMMR.

3.1. Comparison of the Features

We conduct the pre-experiment to select the feature for VMMR. All the images in the dataset are taken from web 3D viewer of Toyota, Mazda, Honda and Nissan cars. The detail of the dataset is described below.

Database: All the images are taken from their front. There are 30 images in total and each model has one image.

**Fig. 2.** Matching under Different Threshold**Fig. 3.** Extracted Keypoints on the Vehicle

Query: There are nine models in total, and each model have from three to ten images. They are all taken from the front to simplify the 3-D object recognition issue. The images are different color from the database images.

Conventional VMMR methods use the edges features and calculate the features from the whole images, not the local points. Our result shows keypoint-based feature is more effective to identify the models.

We compared Histogram Oriented Gradient (HOG)[11] and SIFT. Table.1 shows the average ranking of each model. Table.1 shows SIFT is more suitable for VMMR. Fig. 1 shows the visualization of each feature extraction. The images at the top row show the SIFT keypoints matching results and the ones at the bottom show the difference of HOG between the query and database image. In the difference of HOG images, alignment issue occurs as there exist the gaps between the query and database image. On the other, there are still many geometrically mismatching points in the SIFT images, though alignment issue doesn't exist since SIFT is invariant to scale and rotation. It is obvious that the conventional edge methods [8][10] will fail with the query images taken from an arbitrary angle. This result led us to select SIFT for our method.

3.2. Region of Interest for VMMR

To confirm the ROI in VMMR, we conduct the pre-experiment below. We use the images taken from the corner of the vehicle to capture the whole areas of the object. Fig. 3 shows the keypoints tend to be strongly extracted from its wheels, lights, emblems, front grill and edges of doors. Especially there exist many of them on the wheels and front. Fig. 2 is the visualization of the matching points. Matching points are found by calculating Euclidian distance of features between two points. In Fig. 2, the experiment is conducted in the situation, where the threshold th_{sift} is extremely low or high. Under high threshold condition, points are matching to geometrically incorrect points. When th_{sift} is low, the matching points remaining on the surface is mainly on the wheels and the edges of the doors.

The former result describes the area of the wheels and

front have more information than other part of the vehicle. The latter result shows there are many similar points on the surface of the vehicle and most similar points exist around the wheels.

These two results tell the wheels have most powerful features. However once the wheel is replaced, VMMR system based on wheels fails. Therefore frontal view must be one of the most informative and discriminative areas.

4. PROPOSED METHOD

In our method, we define frontal view as ROI in VMMR due to the issue in section 3.2 and store two different color images for each model due to the color variation issue. In this paper, we are focusing on viewpoint changes and improve the invariance to it. Our system creates the pseudo frontal images to compare with the database. It works when four corner of the license plate are captured in a query image.

4.1. Framework of Our Method

An outline of our proposed method is shown in Fig. 4. One of the main differences from conventional 3D object recognition methods is that our method only needs frontal view images of the vehicle in database and transforms the query image to create new appearance from a certain angle. We describe the role of each process below. More details of our contribution are described in section 4.2 and 4.3.

Input

Database stores only frontal view images and the query image can be taken from variety of angles. In database, there are two colors in each vehicle.

Query Image Transformation

The pseudo frontal images are created by transforming the query images by homography matrix, which is calculated from the corner positions of the license plates.

Find Matching Points

Keypoints are extracted from transformed query image and ones in database using SIFT. We count the number of keypoints for database and query. After finding the closest points by Brute force matching, we calculate Euclidian distance of the features between two keypoints. If the distance is under the threshold, they are defined as matched.

Reduction of the Geometric Mismatching

After finding matching points between the query and database image, We calculate homography matrix based on them by RANSAC and reproject the points, which are defined as matched in the database image, to the query image plane. Then we calculate the reprojection error and if the error by pixel is over the threshold, these points are eliminated as geometrically mismatched. Finally, we count the number of the remaining matching points.

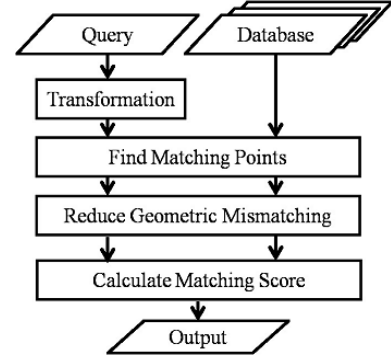


Fig. 4. Framework of Our Method

Calculate Matching Score

After finding geometrically matching points, matching score $Score$ is calculated. We use cosine similarity (eq.1) for evaluation. Higher score means better matching results.

$$Score = \frac{m}{rq} \quad (1)$$

r : #keypoints in a database image

q : #keypoints in a query image

m : #matching points

Output

Finally, this system outputs a list of database images ranking by matching scores.

4.2. Query Image Transformation

Since this system has only the frontal view images in database, the query image taken from angled view has to be transformed to the frontal view. Homography matrix is calculated to transform the query image. We set four corners of the license plate manually and transform the image to the pseudo frontal view.

4.3. Reduction of the Geometric Mismatching

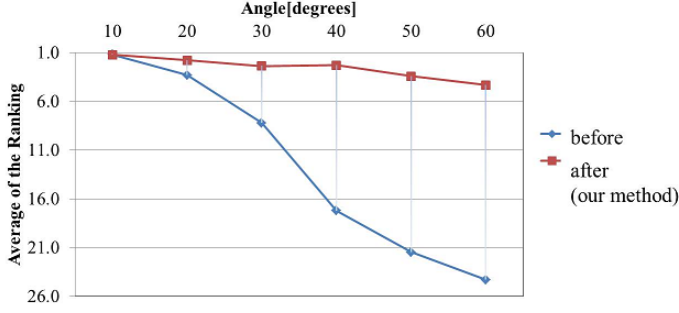
We need to keep geometric consistency between the matching points because there are geometrical mismatchings even if the features are the closest. Homography H is calculated by the positions of the matching points between the query and database image, and reproject ones on database to query image plane. If the reprojection error $Error$ is under the threshold (30 pixel), we count these points as matched.

$$Error_m = ||Hp_{mq} - p_{mdb}||_2 \quad (2)$$

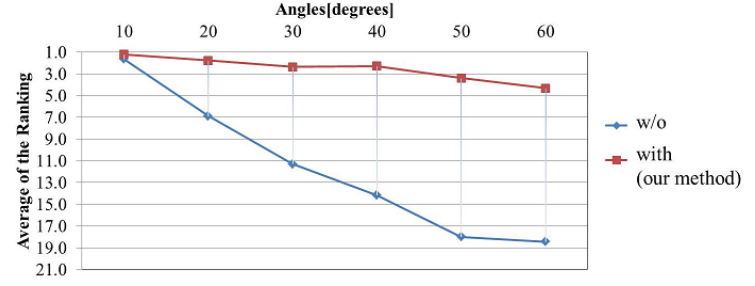
p_{mq} : the position of m th keypoint in the transformed image

p_{mdb} : the position of m th keypoint in the database image

If the number of matching points in each image is less than four, we ignore the database image in ranking due to DoF of homography matrix.



(a) Query Image Transformation



(b) Reduction of the Geometric Mismatching

Fig. 5. The Average Ranking in Each Angle

5. EXPERIMENTS

We conducted two experiments to evaluate our proposed method. Section 5.1 explains the dataset used in our experiments. Section 5.2 mentions the evaluation of our query image transformation method with comparison of before and after applying our method. Section 5.3 mentions the evaluation of our reduction of the geometric mismatching method.

5.1. Dataset

All the images are taken from web 3D viewer of Toyota, Mazda, Honda and Nissan cars to confirm the validity of our method of keypoint matching of the pseudo frontal view.

Database: All the images are taken from their front. There are 60 images in total and each model has two database images since color variation affects keypoint extraction. Hence, one is darker and the other is brighter in gray scale as shown in Fig. 5.1.

Query: There are nine models in total, and each model have three colors in every 10 degree angled view as shown in Fig. 5.1. We chose the colors randomly for the query images, so some models have the same color as the database and some have the different color. The range of the angle is 10 degree to 60 degree. All the cars are taken from left side corner since vehicles have axial symmetrical shapes from their frontal view and no need to try the images on the opposite side. We do not try over 70 degree angled-view because the occlusion occurs on the license plate and it is difficult to find the corners.

5.2. Evaluation of the Query Image Transformation

The following experiment is conducted, in order to confirm the angle limitation of SIFT and the validity of our query image transformation method.

The graph in Fig.5(a) shows the average ranking in each angle to compare the result before and after our transformation method. It shows the angle limitation of SIFT is 20 degrees, and that the ranking result of our method gets better as

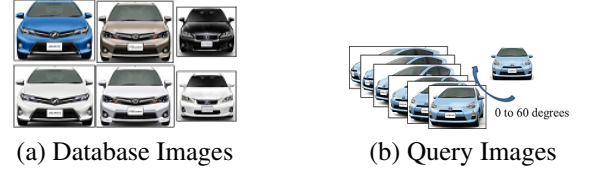


Fig. 6. Dataset

the angle increases. The ranking in our method stays less than five in every angles.

Table 2 is the comparison of the average ranking result in each models. It shows the maximum angle is 30 degree at most successful condition before our transformation method. On the other hand, our method improves the result at any models and angles.

Fig.8 shows the visualization of the matching between non-transformed query and database images. Many keypoints on the side and wheel are extracted in angled-view images and these points match to geometrically incorrect position. The top row in Fig.9 shows the visualization of matching points after making pseudo frontal view. Even the transformed images are more distorted as the angle increases, SIFT can extract the same keypoints on the front grill around the license plate.

This result shows the pseudo frontal view images are important to extract the same keypoints as in the real front view images so that SIFT is not invariant to viewpoint changes.

5.3. Evaluation of Reduction of Geometric Mismatching

The following experiment is conducted to confirm the validity of our reduction of the geometric mismatching method.

The graph in Fig.5(b) shows the average ranking in each angle to compare the result with and without our reduction of the geometric mismatching method. Even after applying our query transformation method, the result of ranking without the reduction method gets worse as the angle increases. Our method improves the result up to 60 degree angled-view images.

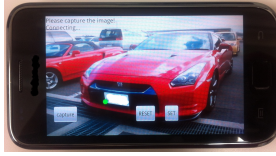


Fig. 7. Screen Shot of Our System

Table 3 shows the comparison of the average ranking result in each models. It shows that our method improves the result at any models and angles.

Fig.9 shows the visualization of the matching with and without our reduction method. Before applying our method, many keypoints are matched to their geometrically incorrect positions. It shows our method can exclude them and remain only reasonable matching points.

This experiment shows reduction of the geometric mismatching is also one of the important process to improve the accuracy. That is because there are many similar keypoints extracted on the surface of the vehicle. It means that even the closest features are not always the geometrically same points.

6. IMPLEMENTATION ONTO THE SMART PHONES

We implemented our method onto the smart phones. The smart phone sends the image of the car and the positions of the license plate. The server receives the data from it, figures out the result sends it back. Fig.7 shows the screen using our system. The users press the CAPTURE button to take a picture and lock the image. Then, they input the position on the touch screen and press SET button to set four corners of the license plate. To restart the system, press RESTART button. The result is shown in the left top corner of the screen.

7. CONCLUSION

We proposed and implemented a novel vehicle make and model recognition method for the smart phone applications. Our method showed the efficiency of creating the pseudo frontal images. It is necessary to focus on the ROI especially for the object with a few keypoints. Our method has view invariance by transforming query images with homography matrix and higher accuracy of the recognition due to reduction of the mismatching.

In the experiments, creating the pseudo frontal images by homography matrix is significant to do keypoints matching by SIFT more effectively. The raw query images output worse results because the same keypoints are not extracted in angled view as in frontal view and it has more mismatching points. Our reduction of the geometric mismatching method reduces the mismatching points and improves the accuracy of the result. That is because there are many similar keypoints extracted on the surface of the vehicle.

In future work, for the automatic identification, the conventional license plate detection method can be applied. To raise the accuracy of the keypoints extraction and deal with absence of the license plate, more efficient transforming

methods (e.g. [12]) can be applied. We are planning to deal with the car region segmentation, reduction of the light conditions and blurring issue.

8. REFERENCES

- [1] "Google goggles," <http://www.google.com/mobile/goggles>.
- [2] "A9," <http://www.engadget.com/>.
- [3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [4] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of British Machine Vision Conference*, 2002, vol. 1, pp. 384–393.
- [5] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, Oct. 2004.
- [6] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 9, pp. 1465–1479, 2006.
- [7] J.M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, Apr. 2009.
- [8] A. Psyllos, C.N. Anagnostopoulos, and E. Kayafas, "Vehicle model recognition from frontal view image measurements," *Computer Standards & Interfaces*, vol. 33, no. 2, pp. 142–151, Feb. 2011.
- [9] D. Han, M. J. Leotta, D. B. Cooper, and J. L. Mundy, "Vehicle class recognition from video-based on 3d curve probes," in *Proceedings of the 14th International Conference on Computer Communications and Networks*, 2005, ICCCN '05, pp. 285–292.
- [10] H. Yabushita, J. Shimamura, and M. Morimoto, "A framework of three-dimensional object recognition which needs only a few reference images," in *21st International Conference on Pattern Recognition 2012*, 2012, pp. 1375–1378.
- [11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, CVPR '05, pp. 886–893.
- [12] S.M. Seitz and C.R. Dyer, "View morphing," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, SIGGRAPH '96, pp. 21–30.

Table 2. Comparison of Our Query Image Transformation Method in the Average Ranking

	10 degrees		20 degrees		30 degrees		40 degrees		50 degrees		60 degrees	
	before	after	before	after	before	after	before	after	before	after	before	after
aqua	1.0	1.0	1.3	1.0	1.3	1.0	2.3	1.0	7.7	1.0	16.0	1.0
auris	1.0	1.0	1.0	1.0	3.3	1.0	25.0	1.0	27.7	1.0	35.7	4.7
corolla fielder	1.0	1.0	1.0	1.0	2.3	1.0	3.0	1.0	17.0	1.3	25.7	1.3
markx	1.0	1.0	3.3	1.3	8.3	1.3	23.0	2.0	15.7	4.7	20.0	7.0
porte	1.0	1.0	1.7	1.0	6.0	1.0	5.3	1.0	6.7	1.0	8.7	1.0
prius	1.0	1.0	1.7	1.0	4.0	1.0	9.7	1.0	30.7	1.0	12.7	3.3
prius alpha	1.7	3.0	9.3	7.7	21.0	12.3	23.3	8.3	22.3	12.0	32.3	13.3
spade	1.0	1.0	1.7	1.0	13.3	1.0	29.3	1.0	29.7	1.3	31.0	1.0
wish	2.0	1.0	8.7	1.0	14.3	1.7	34.0	4.3	36.0	7.3	36.7	6.3



10 degrees



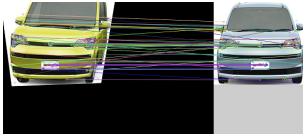
30 degrees



60 degrees

Fig. 8. Matching between the Raw Query and Database Images(spade)**Table 3.** Comparison of Our Reduction of Geometric Mismatching Method in the Average Ranking

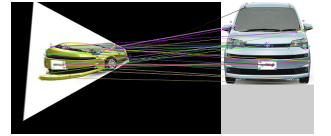
	10 degrees		20 degrees		30 degrees		40 degrees		50 degrees		60 degrees	
	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with
aqua	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	3.3	1.0	7.0	1.0
auris	1.0	1.0	2.3	1.0	5.7	1.0	18.0	1.0	16.0	1.0	17.7	4.7
corolla fielder	1.0	1.0	1.3	1.0	7.3	1.0	4.7	1.0	10.0	1.3	20.3	1.3
markx	1.0	1.0	5.3	1.3	10.3	1.3	19.7	2.0	28.3	4.7	18.7	7.0
porte	1.0	1.0	2.3	1.0	3.0	1.0	3.7	1.0	5.3	1.0	5.7	1.0
prius	1.0	1.0	3.3	1.0	2.7	1.0	13.0	1.0	7.7	1.0	11.0	3.3
prius alpha	3.3	3.0	20.3	7.7	26.7	12.3	23.3	8.3	29.7	12.0	30.7	13.3
spade	2.0	1.0	4.3	1.0	15.7	1.0	20.0	1.0	28.0	1.3	31.7	1.0
wish	3.7	1.0	21.7	1.0	28.3	1.7	24.3	4.3	33.7	7.3	23.3	6.3



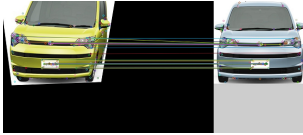
W/O 10 degrees



W/O 30 degrees



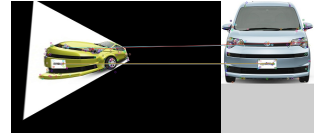
W/O 60 degrees



W/ 10 degrees



W/ 30 degrees



W/ 60 degrees

Fig. 9. Matching between the Transformed Query and Database Images(spade) (Top row shows before applying our reduction method. Bottom row shows after applying it.)