View Independent Recognition of Vehicle Make and Model from View Morphed Frontal Image

Yukiko Shinozuka, Takuya Minagawa and Hideo Saito
Graduate School of
Science and Technology
Keio University
Yokohama, Japan 223–8522
Email: shinozuka@hvrl.ics.keio.ac.jp

***Abstract*—This paper proposes a novel view independent vehicle make and model recognition method(VMMR). Our system identifies the make and model from the variety of viewpoints while the conventional methods for VMMR work only for the fixed frontal or rear images. In addition, it needs only the 2D images not CAD data for database. To solve the alignment issue, our method uses SIFT, that has scale and rotation invariance. For the view independent recognition, it creates the more realistic and less distorted frontal view images by view morphing and extracts the keypoints from them. Our method enables to recognize up to 40-degree angle with high accuracy due to the less distorted morphed images. Our method can be extended to the other product model recognition.**

Keywords*—*Vehicle Make and Model Recognition, SIFT, View Independent, View Morphing

I. Introduction

Prevailing of the smart phones is increasing the demand of Web search applications. The users simply take a picture of the unknown product and search through the 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 the commercial industry to link to products' information, including its performance, price, users' certificates or other similar products. Towards this coming future, more accurate and practical retrieval system is required.

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). Difficult factors in VMMR are 1. Specular reflection due to illumination. 2. A few keypoints due to less-textured surface. 3. Appearance variance due to 3D object. While most of the method deal with the first and second issue, the third issue remains to be solved. The view independent recognition must be the next step for VMMR, 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 virtual frontal images by view morphing and the database stores only frontal view images.

This paper proposes a novel view independent VMMR method. The contribution of this paper is to show the importance of creating the virtual less distorted frontal view for VMMR and to identify the model from a certain range of angles only with the 2D images. Our result explicitly shows the importance of focusing on the region of the interest for the identification. Our result shows higher performance in identification even with the angled view images.

This paper is structured as follows. Section II describes the related works. Section III describes technical difficulties in VMMR. Section IV describes the proposed methods. Section V describes the experiment and its result.

II. Related Works

Scale Invariant Feature Transform (SIFT) [3] is often used in the image retrieval system, for it has both detector and descriptor with high repeatability. It is robust to scale and rotation and can solve the alignment issue. Maximally-Stable Extremal Region Detector [4],Harris-Affine [5] and Hessian-Affine [5] are known for the affine-transform invariant detector, though there is no descriptor for each of them. Even if it is possible to extract the keypoints on the distorted plane, other descriptors cannot describe the feature enough on the point because they are not designed as the affine-transform invariant descriptor. This kind of description will lead the keypoint matching to fail. Keypoint matching using randomized tree [6] is also useful for affine transform, 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.

For VMMR to deal with the illumination issue shown in section I, 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

TABLE I. 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

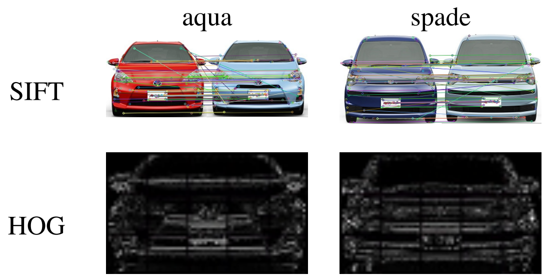

Fig. 1. Visualization of SIFT and Difference of HOG

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]. Shinozuka *et al*. [14] method tries to solve the 3D object issue. Their method create the pseudo frontal image by the homography matrix. This homography is calculated for the licence plate to face the front. However, this method only creates the distorted frontal images.

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 less distorted frontal images by view morphing and only stores the frontal images in database.

III. DIFFICULTIES IN VMMR

In this section, we refer to difficulties in VMMR. Section III-A shows the comparison of the features. Section III-B discusses the ROI for VMMR.

A. *Comparison of the Features*

We conduct the pre-experiment to select the suitable 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.
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 extract 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.I shows the average ranking of each model. Table.I shows SIFT is more suitable for VMMR. Fig. 1 shows the

Fig. 2. Extracted Keypoints on the Vehicle

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 similarity transform(i.e. scale and rotation). The calculation of the reprojection error between the matching points can reduce these mismatchings, so mismatching is not a big issue if the objects are facing the same sides. It is obvious that the conventional edge methods [8][10] will fail with the query images taken from arbitrary angle. This result led us to select SIFT for our method because of its repeatability.

B. *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 whole areas of the object. Fig. 2 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.

This result shows the wheels and front area have more keypoins than other parts of the vehicle. However once the wheel is replaced, VMMR system based on wheels fails. If the keypoints are extracted and described in the comparable way, the keypoints on the front have repeatability and discriminative power for VMMR.

IV. PROPOSED METHOD

In our method, we define frontal view as ROI in VMMR due to the issue in section III-B. In this paper, we are focusing on viewpoint changes and improve the invariance to it. Our system creates the less distorted virtual frontal images to compare with database. It works upto 40 degrees because view morphing often fails when the epipole is within the original image. That means the certain amount of area has to be seen from both sides–left and right.

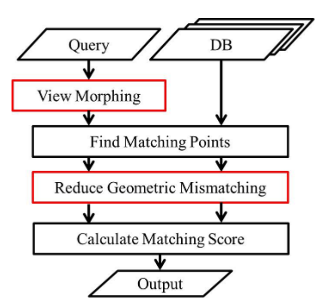

Fig. 3. Framework of Our Method

A. Framework of Our Method

An outline of our proposed method is shown in Fig. 3. One of the main differences from conventional 3D object recognition methods is that our method only use the images for dataset. It 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 IV-B and IV-C.

Input

Database stores only frontal view images and the query image can be taken from arbitrary angle.

View Morphing

The virtual frontal images are created by view morphing.

Find Matching Points

Keypoints are extracted from the morphed image and database by SIFT. We count the number of keypoints on both images. After finding the matching points by bruteforce matching, we calculate Euclidian distance of the features between two keypoints. If the distance is over the threshold, these points are eliminated as mismatching.

Reduction of the Geometric Mismatching

Homography matrix for reprojection is calculated from the positions of the matching keypoints on the morphed and database images. Then the points on the database are reprojected onto the morphed image plane. If the reprojection error is over the threshold, these points are eliminated as geometrically mismatched. Finally, we count the number of the remaining matching points.

Calculate Matching Score

Matching score $Score$ is calculated to compare the similarity. We use cosine similarity(eq.1) for evaluation. The range is $[0, 1]$ Higher score means better matching results.

$$Score = \frac{m}{rq} \tag{1}$$

r : #keypoints in a database image
q : #keypoints in a morphed image
m: #final matching points

Output

This system outputs a list of database images ranking by matching scores.

B. View Morphing

This system has only the frontal view images in database, so the query image taken from angled view has to be transformed to the compatible image. Our method conducts view morphing to create the less distorted frontal view. The basic process of the view morphing is shown in Fig.4 and described below.

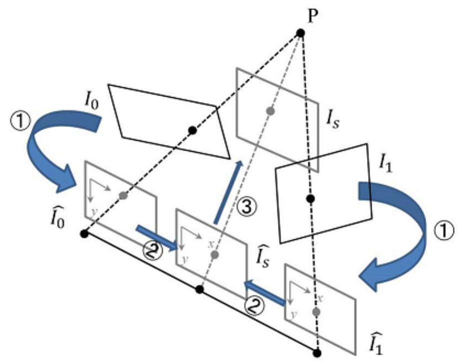

Fig. 4. View Morphing

1) Take the matching points and calculate the fundamental matrix between two imagesI_0,I_1. Prewarp and rectify the original images I_0,I_1 to $\hat{I}_0$,$\hat{I}_1$.
2) Morph: create a mesh based on the points, linearly interpolate positions and intensities of matching pixels in $\hat{I}_0$ and $\hat{I}_1$ to form $\hat{I}_s$.
3) Postwarp the image $\hat{I}_s$ to I_s.

There are three differences from the basic method. First of all, the number of the input image for the basic method are two(i.e. left and right), but our method only needs one image. The shape of the vehicle from the frontal view should be symmetric, so I_1 can be a mirror image of I_0.

Second point is rectification. Hartley's method[13] is applied since the camera parameters are unknown. Either of the homography matrices to postwarp sometimes creates the distorted postwarped image. It makes it difficult to restore the images to the original states. The upper limit of the angle is between 40 degrees and 50 degrees from our experiments. That is because rectification is the process to force the epipolar lines on both images to be paralleled.

Third point is interpolation. Basic method blends two images to create free-view point images, though our method does not blend them. It simply stitches the left side from the left morphed image and right side from the right. That is because the basic interpolation technique blends each mesh and occurs the blur. The mesh has to be a plane, but when we take the points manually it is difficult to extract the points on every edges and corners.

In the case of applying our method to other products, left and right images are required to create the morphed view.

C. Reduction of the Geometric Mismatching

As shown in section III-B, geometrical mismatchings exist even if the points matched as the closest set. To keep the geometrical consistency, homography H is calculated by the positions of the matching points between the morphed and database image, and reproject ones on database to morphed image plane. If the reprojection error $Error$ is under the threshold (30 pixel), we count these points as matched.

$$Error_m = ||Hp_{mq} - p_{mdb}||_2 \tag{2}$$

p_{mq} : the position of mth keypoint in the morphed image
p_{mdb} : the position of mth 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.

V. Experiments

We conducted two experiments to evaluate our proposed method. Section V-A explains the dataset in our experiments. Section V-B mentions the evaluation of our method.

A. 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 morphed frontal view.

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

Query: There are nine models in total, and each model have three colors in every 10 degree angled view as shown in Fig. 5(b). 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 50 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.

(a)Database (b)Query Images

Fig. 5. Dataset

B. Evaluation

The graph in Fig.6 shows the average ranking in each angle. It compares without transformation(normal), Shinozuka *et al*.[14], without reduction of the geometric mismatching and our proposed method. It shows the angle limitation of SIFT is 20 degrees, and that the ranking result of our method gets better as the angle increases except for 50 degrees. The ranking in our method keeps at first place upto 40 degrees. The reduction method efficiently works from 30 degrees. This result shows our method is valiant upto 40 degrees in VMMR.

Fig.7 shows the visualization of matching. Many keypoints on the side and wheel are extracted in angled-view images and these points match to geometrically incorrect position. While the frontal images are more distorted in Shinozuka *et al*. method, our proposed method can create less distorted frontal view. Our method have the matching points on the light, though Shinozuka *et al*. does not have less matching points on the light. That is because when calculating reprojection error, the points on the distorted area are deleted. Bottom two rows show our reprojection process is necessary to eliminate the geometrically incorrect matchings. Without this process, there are many similar points extracted on the surface of the vehicle, so even if the features are the closest, it can easily matches to the incorrect point.

Fig.8 shows the morphed images on AQUA.From 10 to 30 degrees, it succeed to create the undistorted frontal view. At

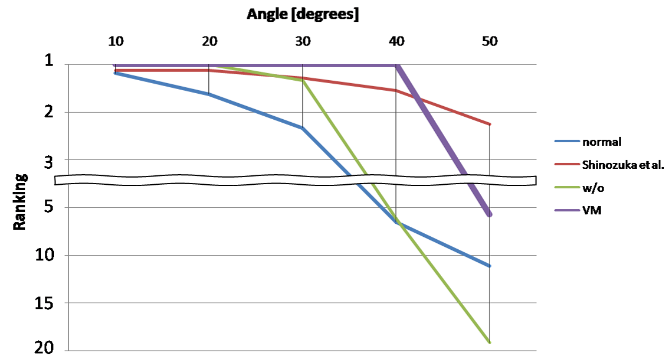

Fig. 6. The Average Ranking in Each Angle

40 degrees, it has some distorted area, though it can still create the frontal image. At 50 degrees, the half of the area is blurred due to the distortion.

Fig.9 shows each process of the view morphing at 50 degrees. As it shows, the epipole is almost in the image. According to [12], view morphing fails when the epipole is within the image. As a result of this, it fails to rectify the images and create blurred morphed images. Compared to Shinozuka *et al*. method at 50 degrees, our method creates more distorted image and license Plate Transformation method creates less distorted one as Fig.9(c)(d) show.

This result shows the less distorted 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.

VI. Conclusion

We proposed a novel vehicle make and model recognition method. Our method showed the efficiency of creating the less distorted 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 view morphing and higher accuracy of the recognition due to reduction of the mismatching.

In the experiments, creating the less distorted frontal images by view morphing 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, we are planning to deal with the automatic identification, reduction of the light conditions and blurring issue.

References

[1] Google Goggles: `http://www.google.com/mobile/goggles/#text`

[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, 2004.

Fig. 7. Keypoint Matching(Prius Alpha)

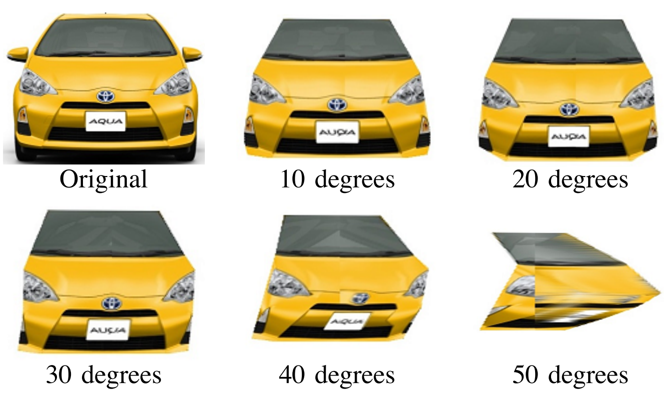

Fig. 8. Morphed Frontal Images(AQUA)

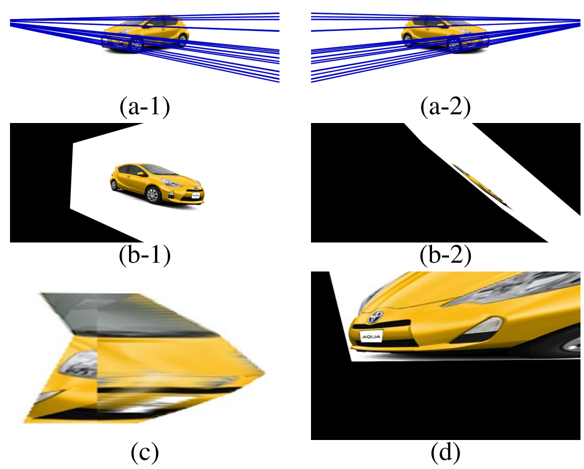

Fig. 9. View Morphing Process at 50 Degrees(AQUA). (a-1)(a-2)Original Images with Epilines (b-1)(b-1)Rectified Images, (c) View Morphed Image (d) Shinozuka *et al*.

[4] J.Matas, O.Chum, M.Urban and T.Pajdal, "Robust wide baseline stereo from maximally stable extremal regions", *Proc. British Machine Vision Conference*, pp.384-393, 2002.

[5] K.Mikolajczyk and C.Schmid, "Scale & affine invariant interest point detectors", *International Journal of Computer Vision*, 60(1), pp.63-86, 2004.

[6] V. Lepetit and P.Fua, "Keypoint recognition using randomized trees", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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*, 2 ,2, pp.438-469, 2009.

[8] A. Psyllos, C.N.Anagnostopoulous and E.Kayafas, "Vehicle model recognition from frontal view image measurements", *Computer Standards & Interfaces*, 33, pp.142-151, 2011.

[9] D. Han, M.J.Leotta, D.B. Cooper and J.L. Mundy, "Vehicle class recognition from video-based on 3D curve probes", *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp.285-292, 2005.

[10] H.Yabushita, J.Shimamura and M.Morimoto, "A framework of three-dimensional object recognition which needs only a few database images", *International Conference on Pattern Recognition*, pp.1375-1378, 2012.

[11] N.Dala and B. Triggs, "Histograms of oriented gradients for human detection", *IEEE Conference on Computer Vision and Pattern Recognition*, vol.1, pp.886-893, 2005.

[12] S.M.Seitz and C.R.Dyer, "View Morphing", *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH*, pp.21-30, 1996.

[13] R.I.Hartley, "Theory and Practice of Projective Rectification", *International Journal of Computer Vision*, vol.35, no.2, pp.115-127, 1999.

[14] Y.Shinozuka, R.Miyano, T.Minagawa, H.Saito, "Vehicle make and model recognition by keypoint matching with pseudo frontal view", *2nd International Workshop on Intelligent Mobile Vision*, To appear 2013.