A study on image generation with shifted perspective using two in-vehicle camera images

Satoshi Ishikawa*, Yuichi Kobayashi*, Toru Kaneko†, Susumu Ishihara‡ and Atsushi Yamashita†

*Department of Mechanical Engineering, Graduate School of Engineering, Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka, Japan

†Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

‡Department of System Engineering, Graduate School of Engineering, Shizuoka University, Japan

***Abstract*—This paper presents a study on generation of an image using two images taken at shifted locations by in-vehicle cameras. Vehicular Ad Hoc Networks (VANETs) provide one of the most important fields of research for high interest in Intelligent Transportation Systems (ITS). As one of the ways to use VANETs, we propose an image generation method to use two images which were taken by in-vehicle cameras at different locations. The image is generated by shifted perspective projection with multiple planes. 3-D coordinates are measured and simple 3-D model made of multiple vertical planes similarly to Tour Into the Picture (TIP) is constructed by matched points. The matched points of two images are decided by Line Segment Detector (LSD) and template matching. Based on this method, an image between two images is generated. The generated image is expected to help intuitive understanding of traffic situations for drivers through VANETs.**

***Keywords*—Image processing, ITS, VANET**

I. INTRODUCTION

Vehicular Ad Hoc Networks (VANETs) provide one of the most important fields of research for high interest in Intelligent Transportation Systems (ITS). As one of the ways to use VANETs, Ishihara *et al.* proposed a system to exchange images taken by in-vehicle cameras to other vehicles through VANETs [1]. In this system, when a car driver requests an image of a place through the VANETs to understand traffic situations or some views around the road, the driver can see an image which was taken by a camera on another car close to the requested place through the car navigation devices. The advantage of this system is that it is independent of the infrastructure, and hence the system is more robust and cost-effective.

However, it is not guaranteed that a certain car is located appropriately to take the requested image. In this case, the system cannot show any pictures. If all in-vehicle cameras had taken videos, it might solve this problems. But the width of the communication line on the system is too narrow to use a large amount of data to solve this problem. Therefore, solutions which do not require a large amount of communication are desired. As one of such solutions, we propose an image generation method to use two images which were taken by in-vehicle cameras at different locations. Even if no image is available at a requested location, it is possible that VANETs supply drivers the requested image which is generated from the two images.

II. PROBLEM

In this paper, the location of the requested image is between those of two images and the direction of the image is the same to back image (see Fig. 1). The distances between the requested image and back image and between front image and back image are given as S [m] and G [m]. G is supposed to vouch for too long to keep track of some image features like SIFT, SURF, and so on. These two images are taken on the same straight road. As another assumption, it is possible to detect lanes on a road from obtained images. In order to shorten the processing time, image generation is based on simple 3-D model like Tour Into the Picture (TIP) [2] (see Fig. 2). Its model is constructed by several planes onto which views around road are projected.

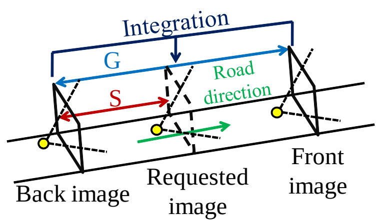


Fig. 1: Situation of image generation

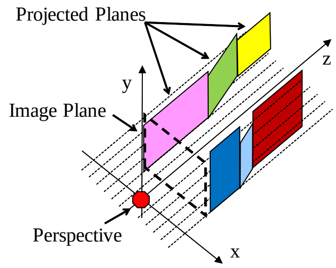


Fig. 2: Situation of 3D model

III. IMAGE GENERATION

A. Outline of Processing

First, the position of template is decided by Line Segment Detector (LSD) [3]. Second, the matching points of two images is found by template matching. By using the matched points, 3-D coordinates are measured and a simple 3-D model which is made of multiple vertical planes similarly to TIP is constructed.

Finally, the requested image is generated by shifted perspective projection with multiple planes.

B. Detection of Matched Points

It is general to use some image features like SIFT to make 3-D model [4]. Although we also need some matching points of two images to construct 3-D model, some image features cannot be kept tracking. So we use the feature points made by LSD and match the features by template matching.

To decide the position of template, first, we use back image (see Fig. 3). Based on the result of LSD on the back image, a list of points are extracted as intersections of two kinds of line segments. One kind is vertical lines, and the other is the lines directed to the vanishing point (see Fig. 4). When the upper side end point of vertical line which is close to either end points of the lines directed to the vanishing point is detected, this upper side end point can be regarded as a corner of the building. After making the list, some points which can correspond to the front image from the list are picked up. As the corners of building, the highest points in their neighborhood from these points on back image are chosen and the points are regarded as centers of templates (see Fig. 5).

With regard to the front image, the list on the same way of the back image is also made. But these points are not picked up from the list because all points of the list on the front image can correspond to back image. To match the corners of building in the back image and ones in the front image, the points of the list in the front image are matched to the centers of templates, where scales of the templates are adjusted according to estimated change of distance. In this way, The matched points of two images are found.

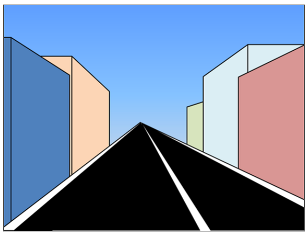


Fig. 3: Back Image

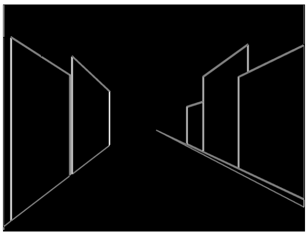


Fig. 4: Two kinds of lines by LSD

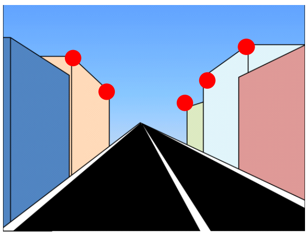


Fig. 5: Detected corners

C. 3-D Measurement

For 3-D measurement, we use triangulation. In triangulation, we need to know some information such as camera parameters in advance. However, it is difficult to get information such as the size of CCD, the distortion of camera, and so on in advance in VANETs. Therefore, 3-D measurement is performed by triangulation using the following assumptions.

1) Perspectives are on the same straight line.
2) Orientations of cameras relative to straight line passing through the perspectives are the same.
3) The heights of two camera are the same.

These assumptions are shown in Fig. 6. G is the distance between two images. f is the focal length. (u,v) is a position of matched point in the image coordinate in the back image and (u',v') is that in the front image. XYZ system is a coordinate that the perspective of the back image is defined as the origin of 3-D coordinates and (X_0,Y_0,Z_0) is a position of this coordinate.. The perspective of the front image is also defined as the origin of $X'Y'Z'$ system and (X_0,Y_0,Z_0) is also a position of this coordinate. The relation of (X_0,Y_0,Z_0) and (X_0',Y_0',Z_0') is expressed as

$$X_0 = X_0' - G\cos\theta \quad (1)$$
$$Y_0 = Y_0' \quad (2)$$
$$Z_0 = Z_0' + G\cos\theta, \quad (3)$$

where θ [rad] is an orientation of cameras relative to straight line passing through the perspectives that is common to the camera of the back image and the one of the front image. Based on these notations, we measure 3-D coordinates of the matched points.

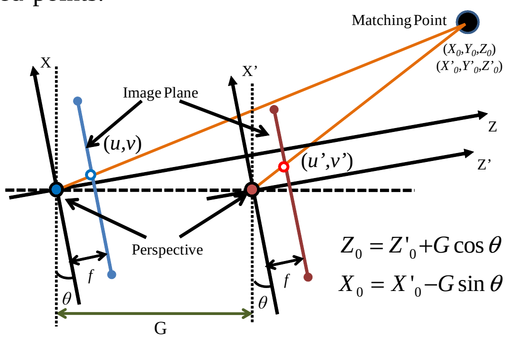


Fig. 6: Triangulation

D. Shift Perspective

Finally, an image is generated from the constructed 3-D model. The way to generate an image is to use the perspective projection like TIP. TIP is a way to generate an image with shifted perspective from an original image and it is based on an assumption that the original picture is taken at a position inside of a box onto which surrounding views are supposed to be projected. We use this idea to generate an image with some assumptions based on section II. Scenery in the back image is supposed to be projected on vertical planes and the edges of these planes is decided by 3-D coordinates of the matched points as shown in Fig. 7. Based on this assumption, we generate an image with shifted perspective from the back image.

The equation of transformation is expressed as

$$u = u' \Big/ \left(1 + \frac{u'S}{fD}\right), \quad v = v' \Big/ \left(1 + \frac{u'S}{fD}\right), \quad (4)$$

where D is the distance between the optical axis and a query point on a plane. To use this equation, we generate an image.

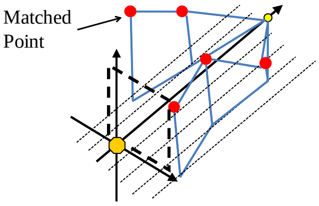


Fig. 7: Simple 3-D model with multiple vertical planes

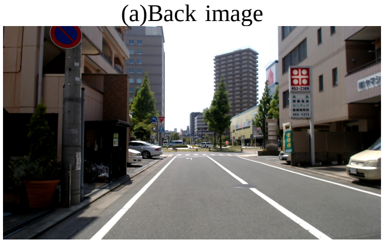
IV. Experiment

To verify the proposed method, we conducted experiment to generate an image from two actual images which are shown in Fig. 8, where Fig. 8 (a) and (b) are back and front image, respectively. The size of images were 1280×720 pixels. This processing was performed on Intel Core i7 processor with 32 GB of RAM. f and S in (4) were 840 and 1000 [pixel].

Extraction of horizontal and vertical line segments by LSD on the back image is shown in Fig. 9. In Fig. 10, blue dots are points of list, green dots are picked up points, and red dots are chosen points, detected by the method described in section III-B, and Fig. 11 shows the matched points of two images. Finally, Fig. 12 shows the generated image from two images.

The processing time was 4.6 seconds. The experiment showed that VANETs can use the image generated by the proposed method, even if no image is available at a requested location.

(a)Back image



(b)Front image

Fig. 8: Obtained image

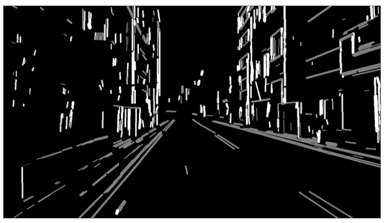


Fig. 9: Two kinds of lines by LSD on back image

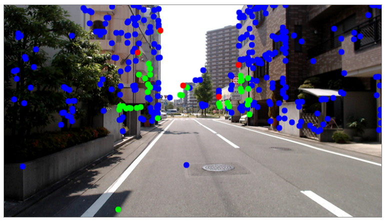


Fig. 10: Detected points on back image

Fig. 11: Result of matching

Fig. 12: Generated image

V. Conclusion

We proposed an image generation method to use two images which taken by in-vehicle cameras at different locations. By using this method, we obtained a generated image from two images and will expect that VANETs will be more serviceable for drivers. As future work, we should experiment with many actual environments and make the method adapted to more complex situations.

Acknowledgment

This work was in part supported by a MEXT KAKENHI Scientific Research (B), 23300024.

References

[1] S. Ishihara, N. Nakamura, and Y. Niimi :“Demand-based Location Dependent Data Disseminaiton in VANETs,” in Proc. of the 19th Annual International Conference on Mobile Computing and Networking, pp. 219-221, 2013.

[2] Y. Horry, K. Anjyo, and K. Arai :“Tour Into the Picture: Using a Spidery Mesh Interface to Make Animation from Single Image,” in Proc. SIGGRAPH, pp. 225-232, 1997.

[3] R. G. von Gioi, J. Jakubowicz, J. Morel, and G. Randall :“LSD: A Fast Line Segment Detector with a False Detection Control,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, no. 4, pp. 722-732, 2010.

[4] N. Snavely, S. M. Seitz, and R. Szeliski :“Photo Tourism: Exploring image collections in 3D,” in Proc. of SIGGRAPH, pp. 835-846, 2006.