Recognition of Road Markings from In-Vehicle Camera Images by a Generative Learning Method

Masafumi NODA*1 Tomokazu TAKAHASHI*1, *2 Daisuke DEGUCHI*1
Ichiro IDE*1 Hiroshi MURASE*1 Yoshiko KOJIMA*3 Takashi NAITO*3

Abstract

In this paper, we propose a method for recognition of road markings from in-vehicle camera images. A road marking in in-vehicle camera images has a variety of appearance due to the change of position and distance between the camera and the road marking, blur effect, and so on. Therefore, accurate recognition of road markings becomes difficult in real environment. Our method generates numerous learning images with the appearance variation in order to recognize road markings in real environment accurately. The effectiveness of the proposed method was confirmed through a recognition experiment using actual in-vehicle camera images.

1 Introduction

With the development in ITS technology, recognition of traffic circumstances by using an in-vehicle camera has attracted a great deal of attention. This enables safety driving, making and updating traffic maps, and so on. In this paper, we propose a method for recognition of road markings from in-vehicle camera images (Figure 1 left). Road markings are painted on the road-plane, which indicate driving directions warnings and so on. Therefore, their recognition is helpful to understand the traffic environment. Li et al. [1] use contour to recognize road markings. The recognition accuracy, however, degrades where a major change of appearance occurs due to change of position and distance between the camera and a road marking, and optical and motion blurring. In our work, we aim at realizing a more robust recognition method by considering such appearance changes caused by such degradation factors.

In this paper, we propose a method that uses an appearance-based approach, which directly learns road marking images with appearance changes. The problem of the appearance-based method is that the cost to collect learning images is very expensive. In order to reduce the cost, a generative learning method for the recognition of traffic signs has been proposed by Ishida et al. [2] [3], which generates learning images. The method constructs generation models considering factors that cause appearance changes so that the generated images should be simulated as close as possible to images captured from the real-world. Such an approach is known as a kind of a virtual learning method [4].

It is important that the generation models reflect accurately real-world factors of appearance changes. Therefore, we tried to construct generation models for road markings considering appearance changes which occur in actual driving environments. It is also important to provide the models with appropriate generation parameters to generate realistic learning images, we determined the generation parameters by analyzing actual driving data.

The paper is organized as follows: The proposed method is introduced in Section 2. Section 3 shows the results of recognition experiments using actual in-vehicle camera images and confirms the effectiveness of our method. Finally, Section 4 concludes the paper.

[1]Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464–8601 Japan
[2]Gifu Shotoku Gakuen University, 1–38 Nakauzura, Gifu, Gifu, 500–8288 Japan
[3]Toyota Central R&D Labs., Inc., Nagakute, Aichi, 480–1192, Japan

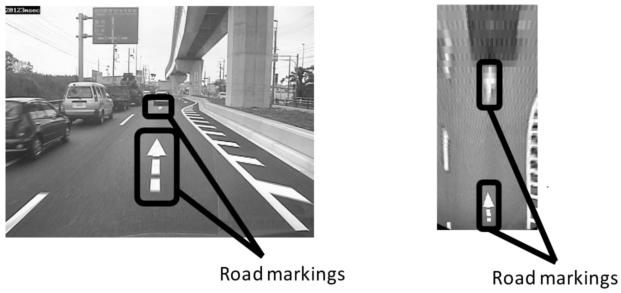

Figure 1: Road markings in an in-vehicle camera image (left) and road markings transformed to the road-plane image (right).

2 Recognition of road markings by a generative learning method

2.1 Overview

Assuming a road plane is flat, a projection transformation using a projection matrix $\mathbf{M}$ is applied to an in-vehicle camera image to obtain a road-plane image which is an image of the road viewed from directly above (Figure 1 right). In the road-plane image, road markings are represented in the same size regardless to the distance from the camera. This transformation simplifies the following operations of the method. The projection matrix $\mathbf{M}$ is calculated from the relation between the position and posture of the camera and a road marking. The parameters that represent the relation between the camera and a road marking are as follows:

- Position of the camera: $\mathbf{p} = (d, l, h)$
  - Distance in the running direction: d [m]
  - Side gap to the running position: l [m]
  - Height from the road plane: h [m]
- Posture of the camera: $\mathbf{r} = (\theta\ [°], \phi\ [°], \psi\ [°])$

Figure 2 illustrates these geometrical parameters.

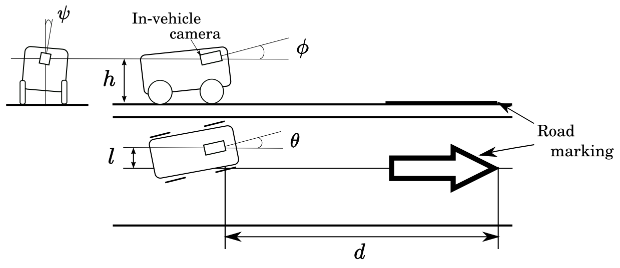

Figure 2: Geometrical position and parameters.

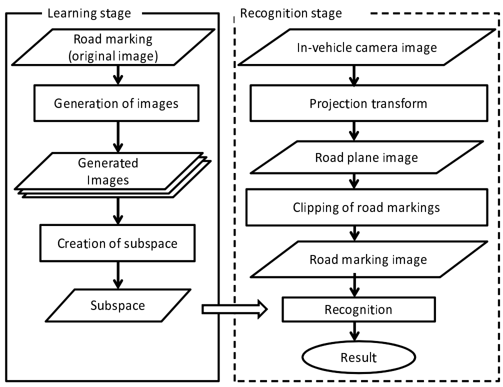

Figure 3: Flow diagram of the proposed method.

Figure 3 shows the flow diagram of the proposed method. A subspace method is used as the recognition method. In the learning stage, we generate numerous learning images with appearance changes for each road marking class from an original image by using generation models. Then we create a subspace from them. In the recognition stage, road marking regions are clipped out from road-plane images, then the similarity between the road marking image and the subspace is calculated for each road marking class.

2.2 Generation model

There are many factors that cause appearance changes of a road marking in an in-vehicle camera image, such as changes of the camera position and camera posture, or internal property of the camera. Figure 4 shows examples of such appearance changes. We classified these factors into the following twelve groups and constructed a generation model for each group.

- Shape deformation
- Resolution degradation
- Optical blur
- Motion blur
- Clipping error

The generation models are used to generate road marking images for learning. In Japan, the shape and size of road markings are standardized, so we prepared a image of the standard road marking image as the original image for generation. In this section, we explain each of the five generation models.

2.2.1 Shape deformation

Because positions and postures of a camera to a road-plane changes constantly while a car runs, we need to estimate the projection matrix $\mathbf{M}$ for every frame in order to project an in-vehicle camera image to a road-plane. To estimate $\mathbf{M}$, we could use objects in the in-vehicle camera image such as the white line.

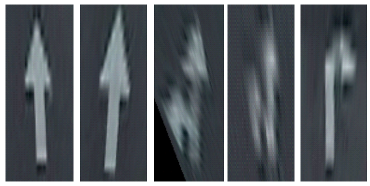

Figure 4: Road markings with variouse appearance.

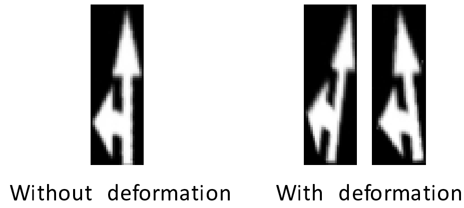

Figure 5: Examples of shape deformation.

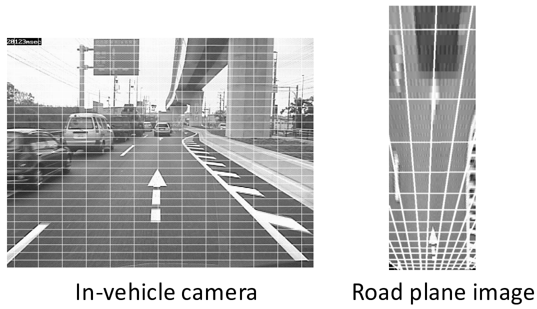

Figure 6: Example of resolution degradation.

However, there is estimation error due to occlusion by other cars, illumination changes by weather, and so on. As a result, shape of a road marking in a road-plane image is deformed as shown in Figure 5.

Therefore, we generate learning images of a road marking whose shapes are deformed by the estimation error rather than estimating an accurate projection matrix.

Let $\mathbf{M}_0$ be a projection matrix of the initial state: $\mathbf{p}_0 = (d_0, h_0, l_0), \mathbf{r}_0 = (\theta_0, \phi_0, \psi_0)$. First, we transform the original image of a road marking to a simulated in-vehicle camera image using $\mathbf{M}$. Then, the image is projected back to the road-plane using $\mathbf{M}_0^{-1}$ instead of $\mathbf{M}^{-1}$ to generate a road marking image with the deformation of shape due to the error between $\mathbf{M}_0$ and $\mathbf{M}$.

2.2.2 Resolution degradation

The resolution of the road plane image obtained by a projection decreases with the distance from the camera. Figure 6 shows how an in-vehicle camera image with grid of equal intervals is projected to a road-plane image. This shows that the resolution of a road-plane image differs depending to the distance from the camera.

2.2.3 Optical blur

The optical blur is often modeled by a point spread function (PSF) [5]. A PSF is represented by the spread of light called the diffusion circle. For in-vehicle cameras used for recognizing traffic environments, the radius of a diffusion circle is assumed as constant regardless to the distance from objects. Therefore, we model

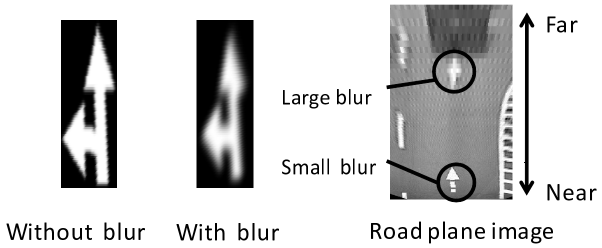

Figure 7: Example of optical blur.

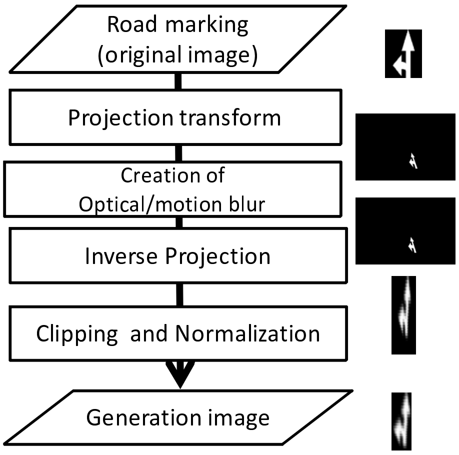

Figure 8: Learning image generation process.

optical blur by a Gaussian filter with a standard deviation r.

2.2.4 Motion blur

The motion blur occurs depending on the exposure time and the motion of the in-vehicle camera. Road markings are fixed on the road-plane. Therefore, we only have to consider how the camera moved during an exposure time T[s]. Because T is sufficiently small, we assume the camera's velocity as uniform. When the position and posture of the camera are represented by a projection matrix $\mathbf{M}_t$ at time t, $\mathbf{M}_{t+\Delta t}$ is calculated by the velocity $\Delta\mathbf{p}, \Delta\mathbf{r}$ of the camera . Then we integrate $F(= T/\Delta t)$ images obtained from $\mathbf{M}_t, \mathbf{M}_{t+\Delta t}, \ldots, \mathbf{M}_{t+(F-1)\Delta t}$ to generate a learning image with motion blur.

2.2.5 Clipping error

Our method use road marking images which are clipped out from road plane images. The position and the size of the clipped images are not precise, so the clipping error is modeled by the positional error $(\Delta x, \Delta y)$, and the size error $(\Delta s_x, \Delta s_y)$.

2.3 Learning image generation

The proposed method generates learning images using the generation models introduced in Section 2.1. Figure 8 shows an overview of the generation process. First, a road-plane image which includes the original road marking image. Then, optical blur and motion blur are added to the image. Next, the image is projected to the road-plane. Finally, road marking images are clipped from the road-plane image with various clipping errors.

In order to generate realistic images, we need to provide adequate generation parameters. The generation parameters can be divided into variables and constants.

Table 1: Generation parameters (constants).

Parameter	Value	
d	10–40	per 1 m
T	1/30	s
r	1.0	pixel
F	3	images

The variables include the position and the posture of a camera whereas the constants include the camera's property. Distribution of occurrence probability of each variable parameter is approximated as a normal distribution. Ishida et al. [3] have estimated the distribution of each parameter from actual images. However, the problem of the cost to collect learning samples still remains because their method need to collect learning images for the estimation of the distribution.

On the otherhand, we determine the distribution of each generation parameter from actual driving data. Because the generation parameters: $l, h, \theta, \phi, \psi, \Delta d, \Delta l, \Delta h, \Delta\theta, \Delta\phi, \Delta\psi$ only depend on car's motional property, we determine these distributions by analyzing positions, posture and velocity in the real driving data. The radius of the diffusion circle r and the exposure time T are obtained from the specification of the camera. As to the distributions of parameters for the clipping error: $\Delta x, \Delta y, \Delta s_x, \Delta s_y$, we determine them experimentally.

2.4 Learning stage

For each road marking class, we create subspaces for different distances from a car using the generated learning images. A matrix $\mathbf{X}_{m,d}$ is formed from N images of a road marking m in a distance of d [m].

$$\mathbf{X}_{m,d} = [\mathbf{x}_{m,d,1}, \cdots, \mathbf{x}_{m,d,n}, \cdots .\mathbf{x}_{m,d,N}] \quad (1)$$

Here, $\mathbf{x}_{m,d,n}$ represents a vector which is composed of pixel values in a generated road marking image. A subspace for class m and distance d is obtained as eigenvectors of $\mathbf{X}_{m,d}\mathbf{X}_{m,d}^T$ corresponding to the L largest eigenvalues.

2.5 Recognition stage

The recognition result $\hat{m}$ for an input vector $\mathbf{y}$ is obtained by

$$\hat{m} = \arg\max_m \max_d \sum_{l=1}^{L} (\mathbf{u}_{m,d,l}^T \mathbf{y})^2, \quad (2)$$

where $\mathbf{u}_{m,d,l}$ represents the l-th eigenvector in the subspace of class m in distance d, and L is the number of eigenvectors which is used for the recognition.

3 Experiment

3.1 Conditions

We conducted a recognition experiment using actual road marking images captured from an in-vehicle camera. As to the settings of the camera, the angle of view was 36 [°], and the resolution was 720 × 480 [pixels]. We mounted the camera adjacent to the windshield at a height of 1.6 [m] from the road-plane, 0.2 [m] on the left side from the back-mirror. The road marking classes

Table 3: Comparison of recognition rates.

Method	Overall		10m–20m		20m–30m		30m–40m	
Proposed	**0.95**	**(949/1000)**	**0.97**	**(485/500)**	**0.94**	**(281/300)**	**0.87**	**(173/200)**
NCC	0.85	(847/1000)	0.88	(440/500)	0.53	(256/300)	0.76	(151/200)
Twice SD	0.91	(912/1000)	0.95	(477/500)	0.92	(276/300)	0.80	(159/200)
Half SD	0.93	(932/1000)	0.96	(479/500)	0.94	(283/300)	0.85	(170/200)

Table 2: Generation parameters (normal distribution).

Parameter	Average		Standard deviation	
h	1.6	m	0.01	m
l	-0.2	m	3.0	m
θ	0.0	°	3.03	°
ϕ	0.0	°	0.64	°
ψ	0.0	°	0.57	°
Δd	5.24	m/s	3/86	m/s
Δh	0.0	m/s	0.01	m/s
Δl	0.0	m/s	0.1	m/s
$\Delta\theta$	0.69	°/s	1.36	°/s
$\Delta\phi$	0.30	°/s	0.38	°/s
$\Delta\psi$	0.30	°/s	0.29	°/s
Δx	0.0	pixel	1.0	pixel
Δy	0.0	pixel	0.3	pixel
Δs_x	0.0	pixel	2.0	pixel
Δs_y	0.0	pixel	0.6	pixel

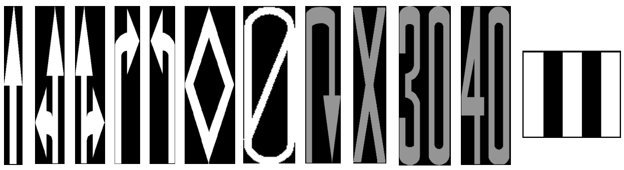

Figure 9: Road marking classes

were the five shown in Figure 9. We created each subspace from $N = 500$ learning images, with a dimension of $L = 11$, which was determined based on the result of a preliminary experiment. Table 1 shows the generation parameters, while Table 2 shows the mean values and the standard deviations of the distribution of the generation parameters.

The proposed method and the following three comparative methods were applied to 1,000 actual in-vehicle camera images.

- Comparative method 1 (NCC) :
  The normalized correlation between an input image and the original image of each class was used as the similarity.

- Comparative method 2 (Twice SD):
  Twice the values of the standard deviations (SD) for $h, l, \theta, \phi, \psi, \Delta d, \Delta h, \Delta l, \Delta\theta, \Delta\phi, \Delta\psi$ shown in Table 2 were used to generate the learning images.

- Comparative method 3 (Half SD) :
  Half the values of the standard deviations (SD) for $h, l, \theta, \phi, \psi, \Delta d, \Delta h, \Delta l, \Delta\theta, \Delta\phi, \Delta\psi$ were used.

3.2 Results and Discussion

Table 3 shows the overall recognition rate and the recognition rates for each group. The rates are divided into three groups according to the distance to the road markings for analysis purpose.

Since the recognition rates of our method ware higher than the comparative methods, we confirmed that it is effective for road marking recognition. Compared with the comparative method 1 (NCC), the recognition rate improved especially when road markings were distant. This demonstrates that the proposed method keeps a relatively high recognition performance even if the appearances of a road marking changes significantly. Since the recognition rate of the proposed method was higher than those of the comparative methods 2 (Twice SD) and 3 (Half SD), we can see that the recognition rates improved by using adequate settings of generation parameters.

4 Conclusion

In this paper, we proposed a method for recognizing road markings from in-vehicle camera images. The proposed method generates learning images using generation models considering appearance changes of road markings, and thus could achieve robust recognition of road markings in actual traffic environments. We confirmed the effectiveness of the proposed method from the results of an experiment using actual in-vehicle camera images. Future works include development of generation models considering other factors, such as flaking and partial occlusions of road markings.

Acknowledgments

The authors would like to thank their colleagues for useful discussions. Parts of this research were supported by the Grant-In-Aid for Scientific Research from the Japan Society for Promotion of Science.

References

[1] Y. Li, K. He, and Peifa Jia, "Road Markers Recognition Based on Shape Information", Proc. 2007 IEEE Intelligent Vehicles Sympo., pp.117–122, Istanbul, Turkey, Jun. 2007.

[2] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, and H. Murase, "Identification of Degraded Traffic Sign Symbols by a Generative Learning Method", Proc. Int. Conf. on Pattern Recognition 2006, Vol.1, pp.531–534, Hongkong, Aug. 2006.

[3] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, and H. Murase,"Generation of Training Data by Degradation Models for Traffic Sign Symbol Recognition", IEICE Trans., Vol.E90-D, No.8, pp.1134–1141, Aug. 2007.

[4] T. Amano, A. Yamaguchi, and S. Inokuchi, "Eigenspace Approach for Object Recognition and its Pose Detection", Systems and Computers in Japan, John Wiley & Sons,Inc., Vol.31, No.11, pp.60–69, Nov. 2000.

[5] B. Okumura, M. Kanbara, and N. Yokoya, "Augmented Reality Based on Estimation of Defocusing and Motion Blurring from Captured Images", Proc. IEEE/ACM Int. Symp. on Mixed and Augmented Reality, pp.97–106, Santa Barbara USA, Oct. 2002.