A Note on Road Narrowing Condition Estimation from In-vehicle Camera Videos Based on Multiple Classifiers Using Integrated Feature Vectors

Sho Takahashi[1], Masahiro Yagi[1], Toshio Yoshii[1], Hiroki Kinoshita[2] and Toru Hagiwra[3]

[1] Faculty of Engineering, Hokkaido University

[2] Graduate School of Engineering, Hokkaido University

[3] Hokkaido Road Management Engineering Center

1. Introduction

In snowy and cold regions, every time snow is removed from the road surface, piled snow on the road shoulder is formed, which narrows the effective width of roads. When a road is narrowed by piled snow, it obstructs urban functions and road traffic services. This obstacle causes serious traffic accidents and traffic congestion. Also, this environment is one of the largest problems for actuating autonomous vehicle systems in snowy and cold regions. Therefore, the observation and removal of the piled snow are carried out in winter to ensure smooth road traffic. Currently, visual inspections by road administrators are performed to ensure smooth road traffic. As denser road networks become, many more roads must be inspected, requiring much time and workforce. However, the number of snow removal workers is decreasing. Therefore, snow removal operations in snowy and cold regions must be more efficient.

The Camera devices have been mounted in many vehicles with miniaturization and cost-reduction advances in recent years. Previous studies in the field of road management include a road condition monitoring system using low-cost MEMS accelerometers to detect abnormalities in road surfaces [1] and a method using portable sensors to monitor road surface conditions [2]. In addition, studies [3,4] that analyze images obtained from inexpensive vehicle-mounted cameras can be cited as methods for estimating road surface deterioration and winter road surface conditions.

In the previous study [5], the authors proposed a method to estimate road narrowing conditions using in-vehicle camera videos into three levels shown in **Fig.1**. Level 3 is defined based on the criteria for dispatching snow removal operation in Sapporo city. Level 1 is defined since it is necessary to grasp if there is piled snow on the road. In reference [5], five features: color, power spectrum, traffic lights, surrounding vehicles, and depth map-based features of the piled snow are utilized to estimate the road narrowing conditions. Also, to reduce the negative effect of some features are not correctly extracted, the method that does not cause a significant decrease in performance based on the late fusion approach was proposed. Late fusion methods are known to be robust to partial errors. Even if one modality has an error

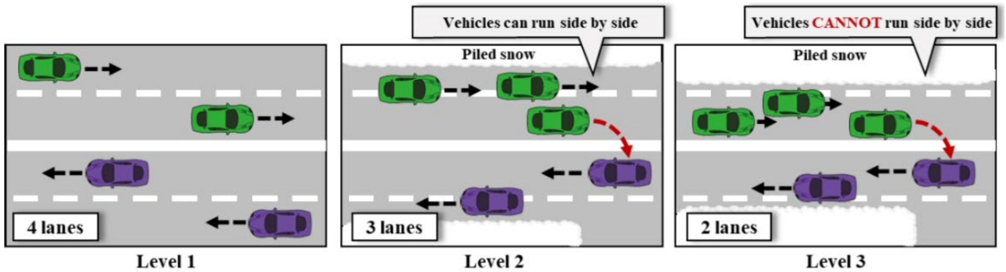

Fig. 1 An Example of the road narrowing condition levels.

it does not necessarily propagate the others [6]. Also, the late fusion approaches can be exceptionally robust when modalities have non-correlated errors, since the chances of all modalities making the same mistake are low [7]. However, the parts of the feature space by using early fusion features were observed from the experimental results in ref. [5]. Therefore, a novel method which additionally use early fusion features is proposed in this paper. The last of this paper the effectiveness of the proposed method confirms by the experiment based on actually observed data.

2. Road Narrowing Condition Estimation Using Integrated Feature Vectors

The overview of the proposed method is shown in **Fig.2**. 5 original features which are color-based features, power spectrum-based features, traffic lights-based features, surrounding vehicles-based features and piled snow-based features, are extracted from the in-vehicle camera videos. The color-based features are obtained by HSV histogram. The power spectrum-based features are obtained by using Gabor filter and Fast Fourier Transform which are compute features related to the spatial frequency. The traffic lights-based features and the surrounding vehicles-based features are location of traffic lights and surrounding vehicles in the image. These locations are obtained by using the object detection based on YOLOv3 [8]. The piled snow-based features are obtained based on the monocular depth estimation [9]. Also, 1 early fusion feature as a novel one is composed by the color-based features, the power spectrum-based features and the piled snow-based features.

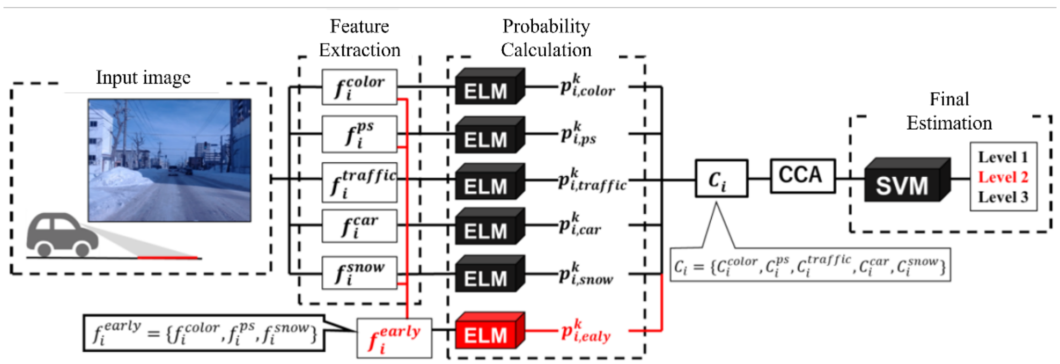

Fig. 2 Overview of the proposed Method

After feature extractions, the proposed method calculates probabilities of each label based on each feature by ELM [10]. ELM is a type of single hidden layer feedforward neural networks (SLFNs), which consist of three layers of neural networks. It enables fast learning speed and universal approximation with a small amount of training data. These calculated probabilities are integrated to confidence level vector C based on the canonical correlation analysis. Finally, the proposed method estimates the final label by SVM which uses the confidence level vector C as input data. In the proposed method, the confidence level is calculated from each feature which include the early fusion feature, and the SVM is used as input to perform an integrated analysis to classify the features. This ensures that the confidence of each feature is optimized independently and that errors in any feature do not affect the other features. This approach was thought to maintain the stability of classification accuracy.

3. Experimental results

An experiment is conducted to verify the proposed method's effectiveness. Video images from an in-vehicle camera mounted on a vehicle driven on a four-lane road in Sapporo, Hokkaido, Japan, are used. The dataset includes roads narrowed by piled snow, and roads widened after snow removal in the same section. The number of training data is 120 videos in each level. The number of test data is 30 videos in each level. The experiment is conducted using five-part cross validation. The number of pixels in each frame of the input video is 640×480, and the video is 3 seconds long at 30 fps. This experiment uses 4 comparative methods. The comparative method 1 (CM1) is used only 5 original features. The comparative method 2 (CM2) is used 5 original features and 5 original features-based early fusion feature. The comparative method 3 (CM3) is used 5 original features and 4 original features which are color, power spectrum, traffic light and piled snow as early fusion feature. The comparative method 4 (CM4) is used 5 original features and 4 original features which are color, power spectrum, surrounding vehicle and piled snow as early fusion feature. Experimental results which are Recall, Precision, F-measure and Average of F-measure, shown in **Table 1**. Table 1 shows the proposed method is better performance in this experiment.

Table 1 Quantitative evaluation of each method (the proposed method is PM).

	Level 1			Level 2			Level 3			Ave.
	R	P	F	R	P	F	R	P	F	
PM	**0.77**	**0.79**	**0.78**	**0.69**	**0.72**	**0.70**	**0.81**	**0.84**	**0.83**	**0.79**
CM1	**0.77**	**0.79**	**0.78**	**0.69**	0.63	0.66	0.77	**0.84**	0.81	0.75
CM2	**0.77**	0.78	0.77	**0.69**	0.70	0.69	0.80	**0.84**	0.82	0.77
CM3	**0.77**	**0.79**	**0.78**	**0.69**	0.70	0.69	**0.81**	**0.84**	**0.83**	0.78
CM4	**0.77**	**0.79**	**0.78**	**0.69**	0.70	0.69	**0.81**	**0.84**	0.82	0.78

4. Conclusions

This paper proposed the method which additionally use early fusion features. Also, the last of this paper the effectiveness of the proposed method confirmed by the experiment based on actually observed data.

Acknowledgement

This work was partly supported by the Committee on Advanced Road Technology under the authority of the Ministry of Land, Infrastructure, Transport, and Tourism in Japan (Project name "Technological development of digital twin-oriented transportation management system in winter road", Principal Investigator: Assoc. Prof. Sho Takahashi, Hokkaido University). Also, this work was partly supported by JSPS KAKENHI Grant Numbers JP23K22877 too.

References

[1] El-Wakeel, A. S., Li, J., Rahman, M., Noureldin, A. and Hassanein, H.: Monitoring road surface anomalies towards dynamic road mapping for future smart cities, 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, pp. 828-832, 2017.

[2] Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S. and Balakrishnan, H.: The pothole patrol: using a mobile sensor network for road surface monitoring, The 6th Inter-national Conference on Mobile Systems, Applications, and Services, IEEE, pp. 29-39, 2008.

[3] Omer, R. and Fu, L.: An automatic image recognition system for winter road surface condition classification, International Conference on Intelligent Transportation Systems, pp. 1375-1379, 2010.

[4] Varadharajan, S., Jose, S., Sharma, K., Wander, L. and Mertz, C.: Vision for road inspection, IEEE Winter Conference on Applications of Computer Vision, IEEE, pp. 115-122, 2014.

[5] Hiroki Kinoshita, Masahiro Yagi, Sho Takahashi, Toru Hagiwara, "A Road Narrowing Condition Estimation from In-vehicle Camera Videos via Late-Fusion based on Confidence Level Integration of Multiple Classifiers," JSCE Intelligence, Informatics and Infrastructure, vol. 5, no. 1, pp. 80-88, 2024.

[6] Snoek, Cees GM, Marcel Worring, and Arnold WM Smeulders. "Early versus late fusion in semantic video analysis." Proceedings of the 13th annual ACM international conference on Multimedia. 2005.

[7] Ergun, Hilal, et al. "Early and late level fusion of deep convolutional neural networks for visual concept recognition." International Journal of Semantic Computing 10.03 (2016): 379-397.

[8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You only look once: Unified, real-time object detection, Computer Vision and Pattern Recognition, pp. 779-788, 2016.

[9] Geiger, A., Lenz, P. and Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite, Computer Vision and Pattern Recognition, IEEE, pp. 3354-3361, 2012.

[10] Huang, G.-B., Zhu, Q.-Y. and Siew, C.-K.: Extreme learning machine: theory and applications, Neurocomputing, Vol. 70, Issues 1-3, pp. 489-501, 2006.