A Browsing and Retrieval System for Driving Data

Masashi Naito, Chiyomi Miyajima, Takanori Nishino, Norihide Kitaoka, and Kazuya Takeda

Abstract— **With the increased presence and recent advances of drive recorders, rich driving data that include video, vehicle acceleration signals, driver speech, GPS data, and several sensor signals can be continuously recorded and stored. These advances enable researchers to study driving behavior more extensively for traffic safety. However, increasing the variety and the amount of driving data complicates the simultaneous browsing of various data and finding desired data from large databases. In this study, we develop a browsing and retrieval system for driving data that provides a multi-modal data browser, query- and similarity-based retrieval functions, and a fast browsing function that skips redundant scenes. For sharing data with several users, this system can be used via networks from PCs or smartphones, This system uses a time-series active search, which has been successfully used for fast search of audio and video data, as its retrieval function algorithm. In a few seconds, this system can retrieve driving scenes that are similar to an input scene from 80,000 scenes. Retrieval performance was compared in various retrieval conditions by changing the codebook size of the vector quantization for the histogram features and a combination of driving signals. Experimental results showed that more than 97% retrieval performance was achieved for driving behaviors of left/right turns and curves using a combination of such complementary information as steering angles and lateral acceleration. We also compared the proposed method to a conventional image-based retrieval method using subjective similarity scores of driving scenes. Our proposed system retrieved similar scenes with about a 75% retrieval performance that was five points higher than a conventional image-based retrieval method. It is because image-based method is sensitive to changes of image in the area except in the region of interest for driving data retrieval. The fast browsing function also skipped scenes that could not be skipped by an image-based method.**

I. INTRODUCTION

Drive recorders capture driving data including video and vehicle acceleration signals in actual driving environments [1]. These driving data are utilized in driver risk consulting as feedback about driving performances to reduce risky driving behaviors. A risk consulting company reported that about 30 to 50 % of traffic accidents could be reduced by mounting drive recorders on vehicles and informing drivers with the data analysis results of safe driving diagnosis services (Table I) [2]. The number of fatal accidents has been decreased year after year by employing of vehicle passive and active safety systems. It is expected to be further decreased by utilizing such safe driving diagnosis services in the future (Fig. 1).

With the increased presence and recent advances of drive recorders, richer driving data including speech, GPS data, and several sensor signals can be continuously recorded and stored. These advances enable researchers to study driving behavior more extensively for traffic safety. However, increasing the variety and the amount of driving data complicates the simultaneous browsing of various data and finding desired data from large databases. To solve these problems, we developed a browsing and retrieval system for driving data that provides a multi-modal data browser, query- and similarity-based retrieval functions, and a fast browsing function that skips redundant scenes. This similarity-based retrieval is achieved using the time-series active search technique proposed in [6]. We show the efficiency of this retrieval method by comparing it to a conventional image-based retrieval method. We also discuss which combination of signals is efficient for driving data retrieval by comparing retrieval performances in various combinations of driving signals.

In the next section, we introduce each function of the proposed system, and in Section III we compare retrieval performances as previously mentioned and discuss the results. In Section IV we present conclusions.

The authors are with the Graduate School of Information Science, Nagoya University, JAPAN. {m-naito, miyajima, kitaoka, takeda}@sp.m.is.nagoya-u.ac.jp, nishino@esi.nagoya-u.ac.jp

TABLE I

RESULT OF RISK CONSULTING

Type of vehicle (# of vehicles)	Accident rate		Improvement rate
	Before consulting	After consulting	
Trucks (500)	6.8 %	3.7 %	46 %
Taxis (350)	53.4 %	35.7 %	33 %
Light trucks (700)	36.2 %	25.5 %	30 %
Buses (330)	5.4 %	2.7 %	50 %

II. BROWSING AND RETRIEVAL SYSTEM FOR DRIVING DATA

The driving data recorded by a drive recorder require large amounts of storage space. Some users want to share these data. Thus a system was developed on a server computer as a web application using CGI [4] for easy access via networks from PCs or smartphones such as iPhone.

In a demonstration experiment of our system, we used a subset of driving data collected in the NUDrive project [3], a large scale real-world driving data collection project. The instrumented vehicle shown in Fig. 2 was used in the data collection. Various sensors are mounted on the vehicle for synchronous recording of driving data. About 470 drivers have participated in the project so far. After recording, these data are annotated manually with annotation labels that describe the driving environments and driver's behaviors in a similar fashion at Dallas (USA) and Istanbul (Turkey) [5]. We use these data for retrieval performance evaluation in the

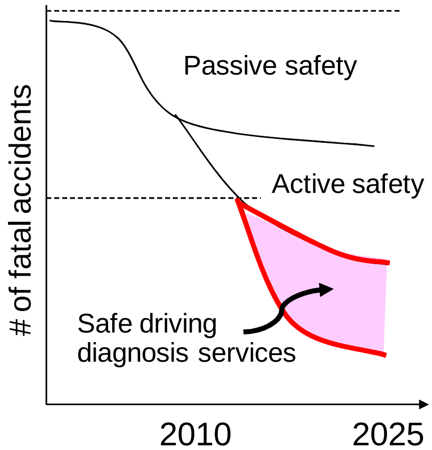

Fig. 1. Expected decrease in the number of fatal accidents by utilizing the safe driving diagnosis services in the future

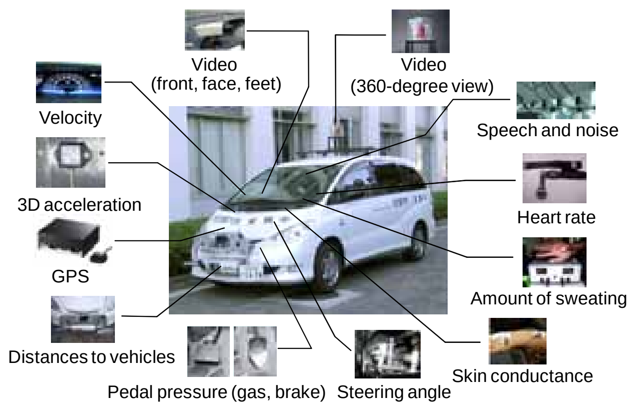

Fig. 2. Instrumented vehicle

next section. An overview of the proposed system is shown in Fig. 3, and the details of each function are described in Section II-A–D.

A. Multi-modal data browsing

Our system supports the browsing of various driving data including video images and sensor signals. When a driver ID and the starting position of a piece of data are specified by a user, the following data will be displayed in pop-up windows (Fig. 3(A)).

1) Videos captured by four cameras: driver's face (two different views), driver's feet, and the road view in front of the vehicle
2) Waveforms of several sensor signals
3) Vehicle's position recorded by GPS and map information
4) List of annotation labels

The browser displays the driving data of x-second length by downloading the data segments from the server. We call this segment of driving data a "scene." Length x of the scene can be flexibly changed based on the situation.

B. Query-based retrieval

Our system provides query-based retrieval of driving data. As shown in Fig. 4, we can retrieve desired driving data by specifying the retrieval queries by the query-choice fields, e.g., the driving data of left turns at less than 40 km/h. Retrieval queries include the type of annotation labels, the signal magnitude specified by a signal type, and a threshold.

C. Similarity-based retrieval

Similarity-based retrieval is also supported by the system. The top N driving scenes with high similarities to the currently viewed driving scene are retrieved from a large database; e.g., the top four scenes are retrieved from 80 hours of driving data in a few seconds (Fig. 3(C)). Unlike the query-based retrieval function, this function retrieves without previous annotation or detail specifications.

As an algorithm of the retrieval function, this system uses time-series active search [6], which has been successfully used for fast search of audio and video data. In this algorithm, the system calculates the similarity between the input scene and the scenes stored in the database as the window for the stored scenes shifts forward in time and detects scenes whose similarity exceeds a threshold value. This algorithm enables high speed retrieval based on the following techniques:

- To reduce the computational cost of similarity evaluations, multi-sensor data sequences are transformed into histogram features, which are low-dimensional vectors.
- By computing the upper bound of the similarity measure in advance, skip all time-step similarity evaluations until the upper bound exceeds the detection threshold.

We choose the detection threshold value with the lowest similarity measure in the interim list of the top N similar scenes in our system. The procedure to obtain histogram features is shown in Fig. 5.

1) Concatenate the multi-sensor data at each sample point into a vector.
2) Quantize the vectors in advance using the LBG algorithm [7].
3) Calculate the histogram of the indices within each scene.

The length of the histogram feature corresponds to the codebook size in the above vector quantization process. Considering the correlation of signals, we use the Mahalanobis distance [8] instead of the Euclidean distance. Since the time order of the data is also important information for driving scene retrieval, we divided a scene into two sub-scenes and calculated the histogram features from each sub-scene based on the time-series active search that incorporates the time order [9]. Moreover, the sensor data used in the retrieval can be chosen by users as the situation demands.

D. Fast browsing

Driving data collected in actual driving environments include such temporally-monotonous sections as "go straight," "long stop," "long curve," and so on, where the difference between contiguous scenes is small. These sections are often

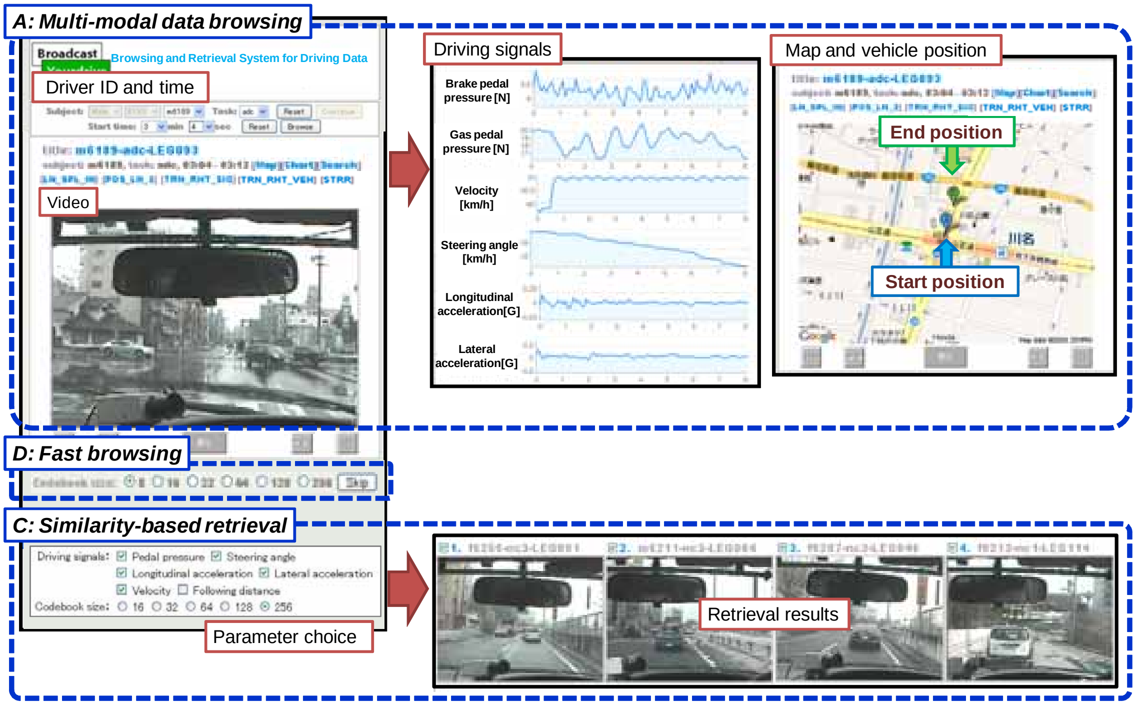

Fig. 3. System overview

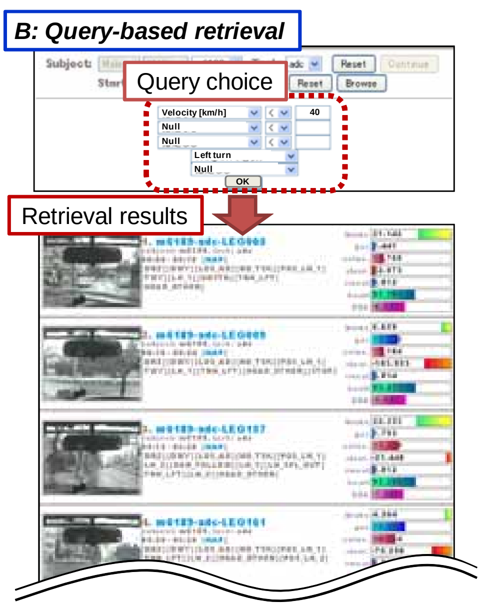

Fig. 4. Interface of query-based retrieval

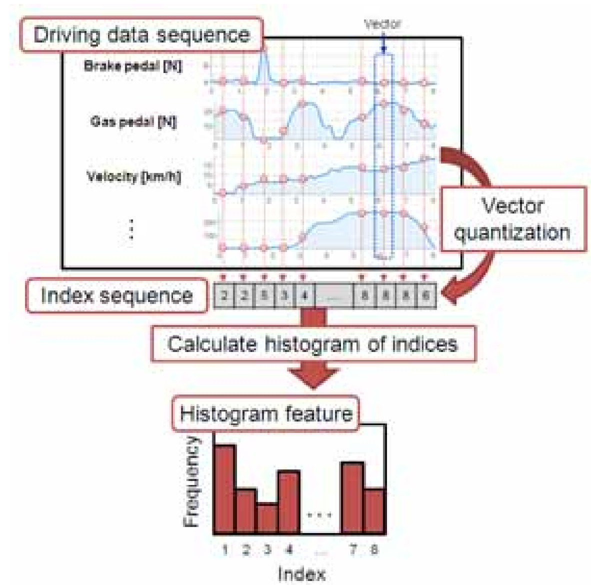

Fig. 5. Transformation procedure into histogram features

redundant when users sequentially browse driving data. By clicking the skip button, this function can easily skip such redundant sections and go to the scene that differs substantially from prior scenes. Such *change-point* is detected based on the similarity. The system calculates the similarity between the current scene and subsequent scenes as the window for them shifts forward in time and detects a scene whose similarity falls below a threshold value (Fig. 6).

An example of fast browsing is shown in Fig. 7, in which we set both the window (scene) length and the window time shift to two seconds and used the vehicle acceleration, the steering angle, the vehicle velocity, and the pedal pressure to calculate the scene similarity. The system skipped redundant sections that seemed monotonous. More than 80 seconds were skipped, including scenes of waiting at stoplights.

Shot detection, which is a related work, is basic video summarization technology [10]. However, shot detection is

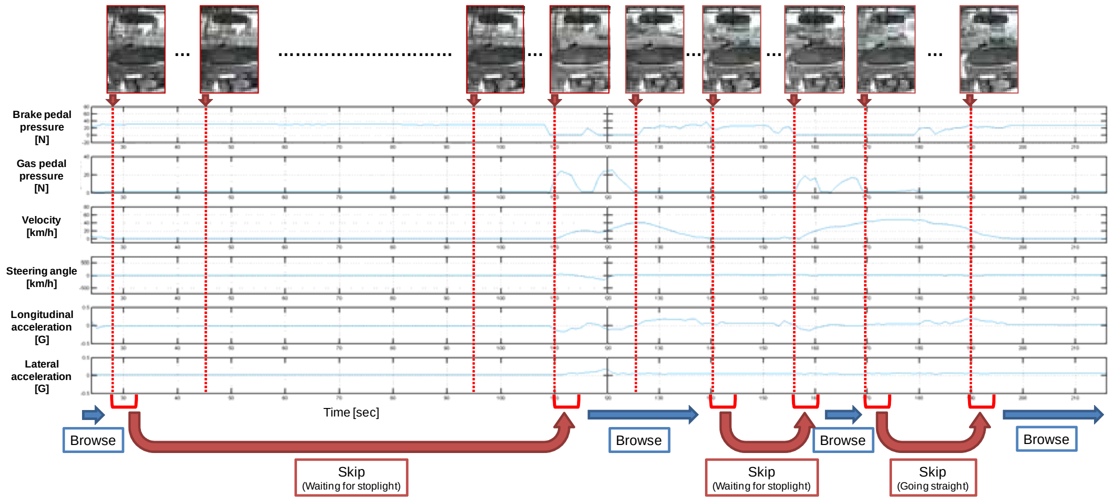

Fig. 7. Example of fast browsing

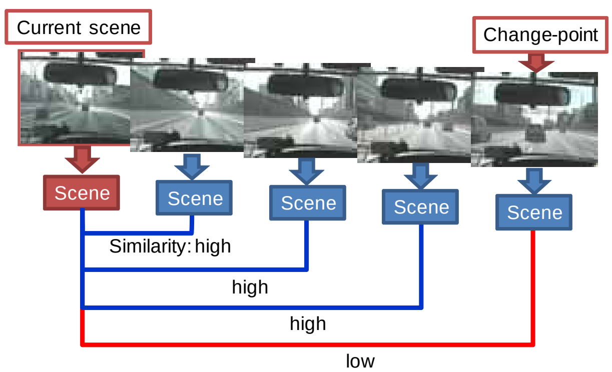

Fig. 6. Overview of fast browsing algorithm

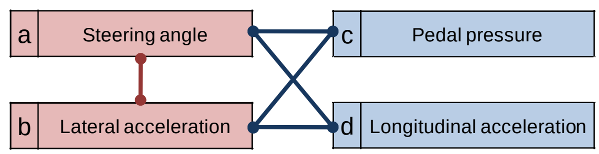

Fig. 8. Signals used in evaluation experiment

sensitive to the changes of images in the area except the region of interest, because shot detection is only based on image features. On the other hand, our proposed method can appropriately detect change-points using various sensor data related to driving behavior.

III. Evaluation of Retrieval Performance

In this section, we evaluate the performances of the similarity-based retrieval described in Section II-C by conducting two experiments.

A. Comparison of retrieval performance in various retrieval conditions

Retrieval performance changes based on the retrieval conditions including the codebook size of the vector quantization for the histogram features and the combination of driving signals. Thus we compared the retrieval performance by changing the retrieval conditions. In the comparison, we used the label data annotated in the NUDrive project [3] to judge whether a retrieved scene is similar.

The experiment was conducted as follows:

1) For each driving behavior of the retrieval target, 30 scenes were selected randomly from scenes with the same label data;
2) For each input scene, the top four similar scenes were retrieved;
3) Retrieval precision was calculated for each input scene and averaged for each target driving behavior.

As target driving behaviors, we chose “left turn,” “right turn,” “left curve,” and “right curve.”

The following signals were used for retrieval.

a) Steering angle
b) Lateral acceleration of vehicle
c) Pedal pressure
d) Longitudinal acceleration of vehicle

We also used the time derivatives of these signals along with the original signals and conducted an evaluation experiment for some combination patterns of these signals: a, b, a+c, a+d, b+c, b+d, and a+b (Fig. 8). Since the target driving behaviors are related to the vehicle’s lateral movement, all combination patterns include a or b. In this experiment, the sampling rate of all signals was 10 Hz, and the codebook size was in the range of 16–256. The experimental results are shown in Figs. 9–12.

From the retrieval results of “left turn” and “right turn” (Figs. 9 and 10), precision exceeding 90% was achieved in most retrieval conditions, except for the combination patterns that didn’t use the steering angles (b, b+c, and b+d). These combination patterns did not work well, especially in the

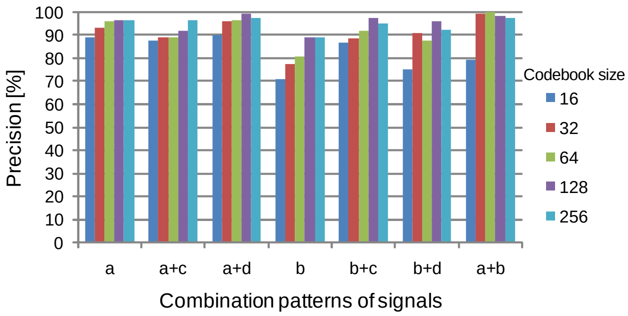

Fig. 9. Retrieval result of “left turn”

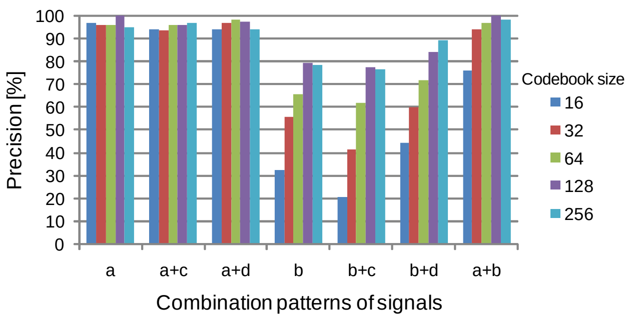

Fig. 10. Retrieval result of “right turn”

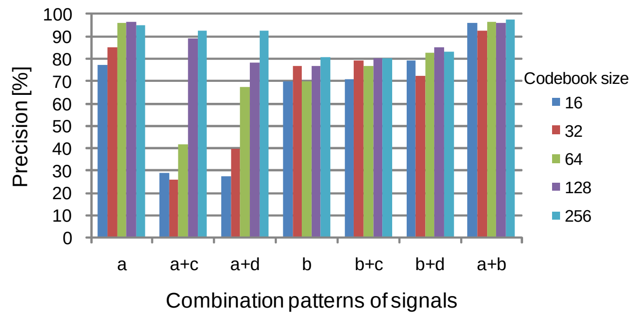

Fig. 11. Retrieval result of “left curve”

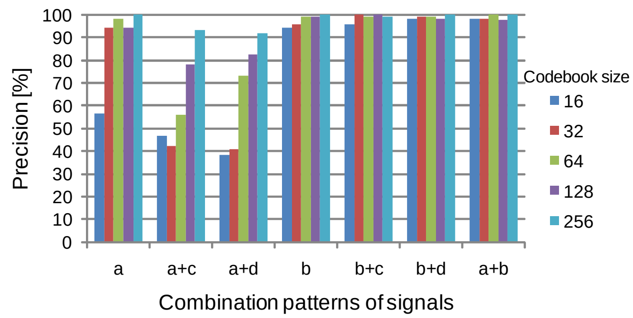

Fig. 12. Retrieval result of “right curve”

“right turn” retrieval.

From the retrieval results of “left curve” and “right curve” (Figs. 11 and 12), precision of about 80% was achieved in most retrieval conditions. The combination of steering angle and lateral acceleration (*a*+*b*) achieved precision that exceeded 90%. On the other hand, combinations of steering angle and a signal unrelated to lateral movement (*a*+*c* and *a*+*d*) did not work well when the codebook size was small.

To determine why some combination patterns did not work well for particular driving behaviors, we compared the steering angle and lateral acceleration characteristics. Examples of the raw signals of the steering angle and lateral acceleration extracted from the database are shown in Fig. 13. Annotation labels are also presented in this figure. This figure shows that the amplitude of the steering angle has a low absolute value when the vehicle is going through a curve, but the amplitude of lateral acceleration takes a relatively high absolute value of about 0.1 G. These low absolute steering angle values could not be trained in the LBG codebook when it was small and the dimensionality of the combined signals was high. This is why the combinations of steering angle and a signal unrelated to lateral movement (*a*+*c* and *a*+*d*) did not work well when the codebook was low.

We can also see from Fig. 13 that lateral acceleration takes a low absolute value when a vehicle is turning at low speed. Thus the lateral acceleration signal is not effective for the retrieval of such a “low speed turn.” In a real driving environment, most right turns are done at low speed because the driver must wait for cars coming from the opposite direction. This is why the combination patterns that failed to use the steering angles (*b*, *b*+*c*, and *b*+*d*) did not work well, especially in the “right turn” retrieval.

In this experiment, the combination of steering angle and lateral acceleration (*a*+*b*) achieved the highest precision for all driving behaviors because this combination pattern complements the weaknesses of each signal described above.

B. Evaluation of retrieval performances for subjective similarity scores

In this evaluation, we used the similarity scores of each pair of driving scenes subjectively scored by human referees to judge whether scenes are similar. These similarity scores were given to all pairs of 42 car following scenes on straight road sections. The similarity of each pair of scenes was scored from 1 to 4. We regarded pairs with scores 3 or 4 as similar. This scoring was only done based on seeing video images of the road view in front of the vehicle.

As reasons for deeming a pair to be similar, referees often pointed how the following distance was maintained. Thus we prepared the following three signal combination patterns (*A*–*C*) that included following distance.

A) Following distance captured by a range sensor
B) Pedal pressure and following distance
C) Velocity, longitudinal acceleration, and following distance

We also used the time derivatives of these signals along with the original signals.

To compare our method with a conventional image-based retrieval method, we prepared an intensity-based image fea-

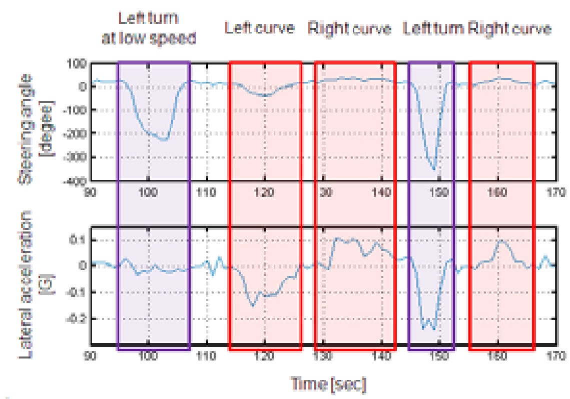

Fig. 13. Raw signals of steering angle and lateral acceleration

ture (combination pattern D) used in the time-series active search [9]. To extract the feature, the image in each video frame was divided into W sub-images. Letting k denote the video frame number and j the sub-image index ($j = 1, 2, \ldots, W$), we express image feature vector $\mathbf{g}(k)$ as

$$\mathbf{g}(k) = (g_1(k), g_2(k), \ldots, g_W(k)), \tag{1}$$

where g_j is the normalized value of the image intensity in jth sub-image given as

$$g_j(k) = \frac{\bar{x}_j(k) - \min_i \bar{x}_i(k)}{\max_i \bar{x}_i(k) - \min_i \bar{x}_i(k)}, \tag{2}$$

where $\bar{x}_i(k)$ denotes the average value of the intensities in the i-th sub-image.

This image feature was extracted from the image of the road in front of the vehicle, captured at 29.41 fps. The capture size was 692 × 480 pixels, and each frame image was divided into 12 sub-images (4 × 3). In calculation of the histogram feature, we quantized the image feature vector using the LBG algorithm [7].

The experiment was conducted as follows:

1) For each of 42 scenes, the top four similar scenes were retrieved from the remaining 41 scenes using the similarity-based retrieval method;
2) Retrieval precision was calculated for each scene and averaged for all 42;
3) Performance was also evaluated for its *success rate*, which considered retrieval successful if at least one of the top four scenes was similar to the input scene.

The precision and success rate are shown in Figs. 14 and 15. The combination patterns that use following distances (A, B, and C) outperform the one using the image feature (D), because the conventional method is sensitive to the changes of image in the area except in the region of interest.

IV. CONCLUSIONS AND FUTURE WORKS

In this study, we developed a browsing and retrieval system for driving data that provides a multi-modal data browser, query- and similarity-based retrieval functions, and a fast browsing function that skips redundant scenes.

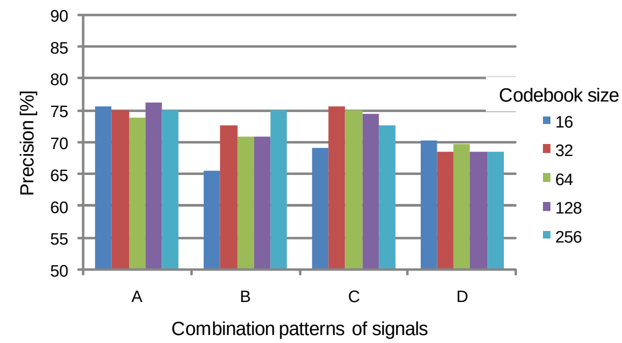

Fig. 14. Precision rate in evaluation experiment based on similarity score

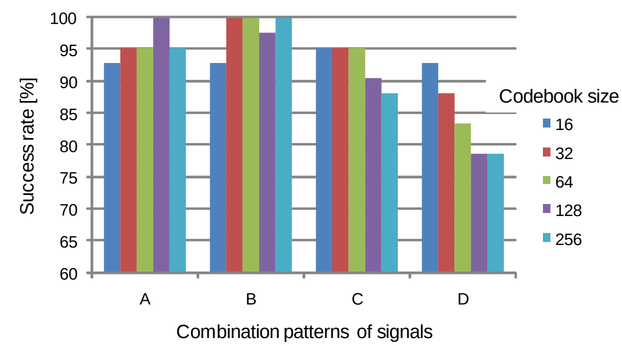

Fig. 15. Success rate in evaluation experiment based on similarity score

From the comparisons of retrieval performances in various retrieval conditions, we found the combination of steering angle and lateral acceleration achieved the highest precision for all driving behaviors, because the combination complemented the weaknesses of each signal.

To show the efficiency of our proposed retrieval system, we also compared the proposed method to a conventional image-based retrieval method using the similarity scores of each pair of driving scenes subjectively scored by human referees. The proposed method outperformed the conventional image-based method in precision and success rate because the conventional method is sensitive to changes of image in the area except in the region of interest.

However, our proposed system has room for improvement. We did not use any feature extraction method for the driving signals. The retrieval performance can be improved by using feature extraction methods like spectral analysis [11]. We must also establish a method to choose the combination patterns of signals that were empirically and manually chosen in the existent system.

V. ACKNOWLEDGMENTS

This work was supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) of Ministry of Internal Affairs and Communications (MIC) Japan under No. 082006002.

REFERENCES

[1] DriveCam, http://www.drivecam.com/.

[2] TOKIO MARINE & NICHIDO RISK CONSULTING CO.,LTD., http://www.tokiorisk.co.jp/consulting/auto_loss/index.html (in Japanese)
[3] A. Ozaki, S. Hara, T. Kusakawa, C. Miyajima, T. Nishino, N. Kitaoka, K. Itou, and K. Takeda, "In-car speech data collection along with various multimodal signals," Proc. LREC 2008, May 2008.
[4] The Common Gateway Interface (CGI) Version 1.1, The Internet Society, 2004, http://hoohoo.ncsa.illinois.edu/cgi/
[5] H. Abut, J. H. L. Hansen, and K. Takeda, "Advances for In-Vehicle and Mobile Systems: Challenges for International Standards," Springer, 2007.
[6] K. Kashino, G. Smith, and H. Murase, "Time-series active search for quick retrieval of audio and video," Proc. ICASSP'99, Vol. 6, pp. 2993–2996, Mar. 1999.
[7] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., Vol. COM-28, No. 1, pp. 84–95, 1980.
[8] P. C. Mahalanobis, "On the generalised distance in statistics," Proceedings of the National Institute of Sciences of India, pp. 49–55, 1936.
[9] K. Kashino, T. Kurozumi, and H. Murase, "A Quick Search Method for Audio and Video Signals Based on Histogram Pruning," IEEE Trans. Multimedia, Vol. 5, No. 3, pp. 348–357, Sept. 2003.
[10] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," Journal of Electronic Imaging, Vol. 5, No. 2, pp. 122–128 Apr. 1996.
[11] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, K. Takeda, and F. Itakura, "Driver modeling based on driving behavior and its evaluation in driver identification," Proceedings of the IEEE, vol.95, no.2, pp.427–437, Feb. 2007.