A Method to Detect Inappropriate Postures causing Distraction via Analysis of Pressure Distribution on the Driving Seat

Makoto Itoh, Toshiyuki Inagaki
Department of Risk Engineering, University of Tsukuba

Abstract

This paper proposes a method to identify driver posture based on pressure distribution on the driving seat. In our method, the Higher-order Local Auto-Correlation (HLAC) features are extracted from an image of a pressure distribution. We conducted an experiment to investigate the effectiveness of our method. The data were collected on a driving simulator. The results show that the method is potentially useful for estimating driver actions. We also tried to find ways to improve the performance of the method. The results show that using two sensor sheets on the seat cushion and the backrest is necessary. The resolution of a sensor sheet can be reduced to half of the original or less. If the training samples have lots of variations, the mean recognition rate goes up to approximately 85%, suggesting the effectiveness of the detection method.

1. Introduction

In order to improve road traffic safety, one vital issue is to establish methods to modify the method of driver assistance depending on traffic conditions and the driver's state (Inagaki, 2007). When a driver is performing an activity that may cause driver distraction, real time distraction "prevention" or "mitigation" is necessary (Engstrom and Victor, 2009). Thus it is important to develop methods for detecting activities which distract the driver.

Methods for detecting activities which distract a driver depend on the type of distraction. Regan, Lee, and Young (to appear) proposed using the terms "visual," "auditory," "olfactory," "tactile," "gustatory," and "internal" distractions in order to differentiate modalities of distraction. This categorization is useful when the impact of distraction on driving is investigated, but may not be useful for development of methods for detecting distracting activity. Distinguishing four categories of distraction, i.e., "visual", "auditory", "biomechanical", and "cognitive" (Ranny, Mazzae, Garrott, and Goodman, 2000) would be enough in this paper.

"Visual distraction" stands for "looking away from the road." Lots of works on detection of visual distractions have been done in the EU and in the USA (see Engstrom and Victor, 2009). Similar research has also been done in Japan. For example, Inayoshi and Kurita (2007) have developed a method of identification of driver’s head position with a single camera. Moreover, some vehicles in Japan are equipped with a system which detects visual distraction (Toyota, 2005).

When a driver is performing a cognitive task (e.g., thinking about some serious issue), this is described as the driver being “cognitively distracted”. “Auditory distraction” refers to the driver paying attention to hearing something (e.g., radio or music) too much other than driving. At least in some cases, the driver’s mental workload increases when the driver is under cognitive and/or auditory distraction. Therefore, distraction due to high mental workload can be detected by evaluating the driver’s mental workload. Facial temperature (see, e.g. Veitman, et al., 2005), and blood pulse wave (see, e.g. Suzuki and Okada, 2008) are candidates for detecting high mental workload non-intrusively and effectively. Cognitive distraction may also be detected by monitoring driver gaze behavior (Engstrom and Victor, 2009; Itoh and Inagaki 2008).

In this paper, biomechanical distraction stands for the situation in which a driver is performing a
task which requires motion of the driver’s body, e.g., reaching somewhere to take something to eat/drink. These kinds of activities can often be observed in automobiles including professional drivers (see, e.g. Barr, et al., 2003; Itoh & Yoshimura, 2007). There is little research on the detection of biomechanical distractions. This might be because it seems easy to detect inappropriate body movement with a video camera for detecting visual and/or cognitive distraction. However, other types of sensing devices are still needed. A camera for detecting visual and/or cognitive distraction needs to be zoomed in to monitor the driver’s eyes precisely. Therefore, it may be difficult to evaluate the driver’s posture based on the images taken by the camera for visual and/or cognitive distraction detection. It would be necessary to use another camera for detection of biomechanical distraction. In that case, the problem is that the camera for detection of biomechanical distraction may not be used for other purposes, i.e., it may not be cost effective.

Itoh (2008) pointed out pressure distribution on the driving seat can be utilized for multiple purposes, such as detection of driver fatigue (Furugori, Yoshizawa, Iname, & Miura, 2005) and sleepiness (Kaneko, Enokizono, Kamei, Fujita, 2008), evaluation of situation awareness (Itoh, Hanyuu, Suzuki, Kurita, and Inagaki, 2008), and detection of biomechanical distraction (Riener et al., 2007). Riener and Ferscha (2008) also tried to develop a person identification method based on pressure distribution information. However, it has not been clarified how a driver’s posture can be identified based on the pressure distribution. This paper proposes a method to identify posture by taking the pressure distribution itself into account.

2. Posture recognition method

This paper proposes using HLAC (Higher-order Local AutoCorrelation) features (Otsu, Kurita, 1988) extracted from an “image” of the pressure distribution on the driving seat. HLAC features are often used for image recognition, since HLAC features are inherently shift-invariant and computationally inexpensive (see, e.g. Kurita and Hayamizu, 2003). These characteristics suggest that the use of HLAC features may be suitable to detect a driver inappropriate posture in an automobile.

In this paper, the pressure distribution sensors shown in Figure 1 are used. The sensor sheet on the seat cushion has 38 * 37 sensing points, and the sensor sheet on the backrest has 25 * 40 sensing points. Sample images of these sensors are shown in Figure 2.

From each figure, HLAC features are extracted. In general, N-th order autocorrelation functions with N displacements \( a_1, a_2, \ldots, a_N \) are defined by

\[
x(a_1, a_2, \ldots, a_N) = \int f(r)f(r + a_1) \cdots f(r + a_N) dr,
\]

where \( f(r) \) denotes the pixel value at the reference point \( r \). The number of these autocorrelation functions is enormous. It is necessary to reduce it for practical image recognition. A typical way to do this is to make restrictions as follows:

- the value of N is not greater than two (i.e., N=0,1,2),
- the range of displacements is within a local 3*3 window, the center of which is the reference point.

In this case, the number of patterns of displacements is reduced to 35 (an equivalent displacement to another one by which the shift is eliminated). Figure 3 depicts the 35 mask patterns. By applying the masks to all pixels in an image, HLAC features are computed and a feature vector whose dimension is 35 is obtained.
Figure 1 Pressure distribution sensors

Figure 2 Sample images of the pressure distributions
Each feature vector $x$ is mapped to a new feature vector $y$ in a discriminant space by Linear Discriminant Analysis (LDA). The new feature $y$ is derived as

$$y = A^T x,$$

where $A = [a_{ij}]$ is a coefficient matrix. The optimal coefficients can be determined from training samples so as to maximize a discriminant criterion $J = \text{tr}(\hat{\Sigma}_W^{-1}\hat{\Sigma}_B)$, where $\hat{\Sigma}_W$ and $\hat{\Sigma}_B$ denote the within-class and between-class covariance matrices defined on $y$. The optimal coefficient matrix $A$ which maximizes $J$ is obtained by solving the eigenvalue problem

$$\Sigma_B A = \Sigma_W A \Lambda \quad (A^T \Sigma_W A = I),$$

where $\Sigma_W$ and $\Sigma_B$ are the within-class and between-class covariance matrices defined on $x$, and $\Lambda$ is the diagonal matrix of the eigenvalues.

In this paper, a simple classifier is used to identify the class of the posture. The classifier checks the distance from an input vector $y$ derived from a test sample to the mean vector $\bar{y}_k$ of Class $k$ in the discriminant space. The input is classified to the nearest class.

3. Experiment

3.1 Purpose

The purposes of the experiment are: (a) to evaluate the effectiveness of the proposed method, and (b) to find a reliable and cost-effective way to realize the detection system. The following factors may be related to the reliability of the detection and the cost.

1. The number of the sensor sheets. If either the sensor sheet on the seat cushion or the sheet on the backrest is enough, the other one can be removed.
2. The resolution of the sensor sheets. If both sensor sheets are required, one way to reduce the cost is to reduce the number of sensing points on a sensor sheet. Reduction of the
resolution might contribute to improvement of the reliability. The extracted HLAC features may not be appropriate if the resolution is too high, since HLAC features are extracted from only very local areas (as stated above, a 3 * 3 local window is usually used).

3) Varieties of training samples. If the training samples contain a wide variation in the data, the detection system may be robust. However, increasing of the variation may raise the cost.

4) Individualization. If the detection system is individualized, the system may be reliable but expensive. If individualization is not necessary, we can also reduce the cost of the detection system.

3.2 Method

Ten persons (five females, aged from 23-38, mean=30.6, SD=7.6, and five males, aged from 22-39, mean=28.0, SD=7.6) participated in the experiment. Even though the participants were recruited through a temporary staffing agency, they were given the same rights as ordinary voluntary participants. The participants were paid according to the guidelines of the University of Tsukuba.

The data were collected on a driving simulator shown in Figure 4. The driving simulator is motion-based. However, motion was not provided in this experiment. The participants were not asked to drive the vehicle. Their task was just to take a posture on the driving seat. Five classes are distinguished in this experiment as follows:

C1. Take the normal driving posture.
C2. Stretch the left hand to the left as far as possible (Figure 5 (C2)). This aims to simulate picking up something on the passenger seat next to the driving seat (note here that the driving seat is on the right hand side in Japan).
C3. Touch the point shown in Figure 5 (C3). This aims to simulate taking something from a pocket on the back of the passenger seat.
C4. Touch a navigation screen which is located in the center of the dashboard (Figure 5(C4)).
C5. With the right hand, touch the floor as shown in Figure 5(C5). The point to touch is near the right heel.

These postures are typical examples of unnecessary actions causing body movement that were frequently observed in real world truck driver behavior in Itoh and Yoshimura (2007). For each posture, a driver was asked to press the gas pedal a little. The reason for this requirement was to simulate driving without driver assistance systems such as ACC (Adaptive Cruise Control). The position of the driving seat and the angle of the backrest were set as the participant felt comfortable. At the first stage of the experiment, those parameters were recorded. Before starting each data collection, the driving seat was configured based on the parameter values. Every driver was also asked to fasten his/her seatbelt.

Two types (type-A and type-B) of samples were collected for this experiment. For type-A samples, the participant was asked to take a posture only once for each category of the postures (C1 to C5). For each posture, 100 snapshots were obtained at one time for a data set. This way of taking data was time-saving, but the data in the same class were similar to one another. For type-B samples, a participant was asked to take a posture twenty times from C2 to C5. The order was randomized. Between each pair of postures, a participant was asked to come back to C1 (Thus, C1 was taken 80 times). At each posture taking, five snapshots were recorded. Thus, we have 100 snapshots for each class (for C1, 100 snapshots are randomly chosen) as a data set. This way of taking data was time-consuming, but the data had wide variation.
The data collection lasted three days for each participant. On each day, one type-A data set and three type-B data sets were recorded. After completion of one data set recording, every participant got off the driving simulator and took a rest. Thus, the initial driving positions differed from one another slightly.

3. Results and Discussions

First, the necessity of both sensor sheets is discussed. Three conditions, i.e., using only the sensor on the seat cushion, using only the sensor on the backrest, and using both, were compared. By calculating the (correct) recognition rate for each participant, the data shown in Figure 6 was obtained (each error bar represents the standard deviation of the associated condition). In this analysis, training was based on a type-A data set. All (nine) type-B data sets were tested to each training data set. Therefore, 27 tests were done for each participant. The recognition rate of a participant used in Figure 6 was the mean recognition rate of the 27 tests.

We conducted a single-factor repeated measures ANOVA on the recognition rate. The main effect was statistically significant ($F(2, 18)=12.53, p<0.0004$). Tukey’s HSD test revealed that there were significant
differences between “both” condition and “seat cushion only” condition (p<0.01), and between “both” condition and “backrest” condition (p<0.01). Therefore, we can claim that using the two sensor sheets is necessary for achieving a high recognition rate.

![Figure 6 Effect of combining the sensor sheets on recognition rate](image)

Second, reduction of the sensor resolution is discussed. Using two sensor sheets may result in an increase of the cost. It is necessary for us to find a way to reduce the cost. One possible way is to reduce the number of sensing points on each sheet. Figure 7 depicts the result of the reducing the resolution. The horizontal axis represents the reduction rate of the resolution. The vertical axis represents the difference in the mean recognition rates between the tests based on the original data and a test based on the resolution-reduced data. According to Figure 7, the resolution reduction did not decrease the recognition rate at least the resolution equals to or was greater than 1/8 of the original one for the “both” condition. This result implies that the total number of sensing points can be reduced from the original setting even if two sensor sheets are used.

![Figure 7 Effect of reducing resolution on detection rate](image)

Third, we discuss increasing the variety of training samples. There is room to improve the recognition rate even when two sensor sheets are used. We investigated the effectiveness of increasing the variety of training samples. Four types of trainings were compared. In each type, training was done with Type-1: a single type-A data set,
Type-2: a single type-B data set,
Type-3: a mixture of two type-B data sets in one day,
Type-4: a mixture of two type-B data sets that were from different days from each other. In types-3 and 4, the number of training samples for each class was kept at 100 by extracting samples randomly. For every type of training, tests were done with the remaining type-B data sets. The total recognition rate for each type was calculated as the mean recognition rate of all tests in the type. In Figure 8, the mean value of all participants’ total recognition rates was shown for each type. The error bars represent the standard deviations. A single-factor repeated-measures ANOVA showed that the main effect of training type was statistically significant \( F(3, 27) = 19.99, p < 0.01 \). According to Tukey’s HSD test, there were significant or nearly significant differences between type-1 and type-2 \( p = 0.08 \), between type-1 and type-3 \( p = 0.0005 \), between type-1 and type-4 \( p = 0.0002 \), between type-2 and type-4 \( p = 0.0004 \), and between type-3 and type-4 \( p = 0.06 \). Thus, it can be claimed that the higher the variety in training samples, the higher the recognition rate.

In our experiment, the final recognition rate was approximately 85%, as shown in Figure 8 (type-4). Even though this is not high enough for practical use, the result suggests that our proposed method is potentially useful. Detection rate depends on the combination of classes to be categorized. For example, as shown in Figure 9, the recognition rate for C3 was relatively low for both 1 and 4 training types. This result suggests that the recognition rate can be higher if we can neglect C3. Whether C3 can be neglected or not depends on the purpose of the detection system. Since body movement in C3 is small, driving maneuvers may not be significantly degraded. If the system aims to hit the automatic brake when a rear-end collision is imminent, for example, the system does not need to detect actions like C3. If the system aims to give some caution when a driver is paying attention to something other than driving itself, however, the system has to detect such actions. The reliability of the detection required of the system is also dependent on the severity of the situation and/or degree of support.

![Figure 8](image1.png) Dependence of recognition rate on learning data

![Figure 9](image2.png) Difference between types 1 and 4 in recognition rate for each class
Finally, the necessity of individualization for training is discussed. In the above analyses, training was done separately for each participant. Here, the following two types of training were taken into account. In each type, training was done with

Type-5: 100 randomly chosen samples for each class from all the type-B data sets of all participants,
Type-6: 100 randomly chosen samples for each class from all the type-B data sets of nine participants other than the tested person.

For tests based on type-5 trainings, 10% of the training samples were from the data of the tested person. On the other hand, no own data were included in the type-6 samples for each person. Tests were with the remaining type-B samples for each participant.

Figure 10 shows the differences among training types 4, 5, and 6 by showing each participant’s recognition rates of those types. A single-factor repeated-measures ANOVA for these data showed that the main effect of the training was significant (F(2, 18)=15.7, p<0.01). According to Tukey’s HSD test, there was no significant difference between type-4 and type-5. However, there were statistically significant differences between type-4 and type-5 (p<0.01), and between type-5 and type-6 (p<0.01). The results showed that training based on a “common” data set was almost the same as the one based on the individualized training data set if at least 10% of the “common” data was from the data of the tested person. However, if the training samples did not have the data of the tested person at all, the recognition was degraded. This result implies that common training can be done if we can categorize drivers into a small number of groups, where differences among drivers in the same group can be neglected from the viewpoint of the pressure distributions.

4. Concluding Remarks

The aim of this study was to propose a method for driver posture identification by measuring pressure distribution on the driving seat. The reason for using the pressure distribution was that it could be used for multiple purposes. From the results of the experiment, we found that the proposed method would be useful for detection of biomechanical distraction. In particular, reaching to the left and picking up were detected with a high correct detection rate. These two actions are known to induce distraction (see Regan, Young, Lee, and Gordon, 2009).

This paper tried to find ways to improve the detection method. The results imply that both the sensor sheets on the seat cushion and the backrest are necessary. The cost that might increase due to using the two sensor sheets can be saved by reducing the sensing points to the half of the original or less. We also found that a wide variety of training samples was needed in order to achieve robust recognition. However, the cost of collecting various training samples may not be so high, since our results suggest that individual
training for each driver may not be necessary if we can categorize drivers into a small number of groups.

For further research, it is necessary to identify which actions must be detected and improve the detection rate for those actions. The postures in this study were important ones to be detected but were not exhaustive. Reorganizing the categorization would be necessary based on Regan et al. (2009). It is also necessary to investigate whether our approach is robust or not when a vehicle is running.

Acknowledgments

This study was supported by the Ministry of Land, Infrastructure, and Tourism. The authors thank Yuuzou Hanyuu and Iori Suzuki for their help in data collection. We would like to extend our appreciation to one of the reviewers for his/her constructive comments on the earlier version of this paper.

References


