Recognition of a driver's gaze for vehicle headlamp control

Jae Hyun Oh and Nojun Kwak
Department of Electrical & Computer Engineering, Ajou University
Suwon 443-749, KOREA

***Abstract*—In this paper, we propose a novel method for gaze recognition of a driver coping with rotation of driver's face. Frontal face images and left half profile images were trained separately using Viola-Jones algorithm to produce classifiers that can detect faces. The right half profile can be detected by mirroring the entire image when neither a frontal face nor a left half profile was detected. As an initial step, this method was used to simultaneously detect the driver's face. Then, we applied a regressional version of linear discriminant analysis (LDAr) to the detected facial region to extract important features for classification. Finally, these features were used to classify the driver's gaze into seven directions. In the feature extraction step, LDAr tries to find features that maximize the ratio of inter-distances among samples with large differences in the target value and those with small differences in the target value. Therefore, the resultant features are more fitted to regression problems than conventional feature extraction methods. Besides LDAr, in this paper, a two-dimensional extension of LDAr is also developed and used as a feature extraction method for gaze recognition. The experimental results show that the proposed method achieves a good gaze recognition rate under various rotation angles of a driver's head resulting in a reliable headlamp control performance.**

***Index Terms*—Gaze recognition, headlamp control, Viola-Jones, dimensionality reduction, LDAr, 2DLDAr.**

I. Introduction

During the last few decades, the field of intelligent vehicles has been rapidly growing [1]. As one component of ITS (intelligent transportation system), intelligent vehicles use many sensors and algorithms to process the information around the vehicle. These systems offer a significant enhancement in safety and operational efficiency to the drivers [2] [3] [4].

In intelligent vehicles, the sensory information may be used to detect obstacles and pedestrians with the objective of keeping a safe distance between the vehicle and detected objects. GPS, laser scanner and vision devices such as infrared camera have been proposed to sense objects efficiently. For example, in [2], a system for automatic adaptation of the longitudinal speed control is proposed using a combination of different sensor technologies: a GPS unit, RFID readers and a speed measurement system based on differential hall effect sensors [2]. A LIDAR and a single camera was used to detect pedestrian in [3]. A real-time 3D range scanning camera based on time-of-flight techniques was developed in [4]. Among various types of technologies, vision devices and computer vision algorithms play a critical role in the field of intelligent vehicle due to their ability to provide diverse information without much cost. The robust detection and tracking techniques for intelligent vehicle can be utilized to provide a visual tracking modality to a traffic advisory system [3] [4] [5] [6].

Besides detecting and tracking outside objects, monitoring activities of a driver is also very important to avoid accidents and to keep safety. The risk of accident increases when the driver controls devices such as air conditioner and audio because he/she cannot concentrate on driving in these situations. In addition, at night time, there may be sites at which a driver wants to stare but no headlamp light reaches. For these reasons, the system that can recognize driver's gaze and automatically control headlamp is needed.

Over the past years, gaze recognition has been studied by many researchers and many methods have been proposed [7] [8]. The position and shape of iris were used to recognize gaze in [7] and [9]. There are also techniques that use electro-oculography [10], pupil and eyelid tracking [11], corneal and pupil reflection relationship [12] and artificial neural networks [13]. Recently, hybrid methods which combine two or more of these methods are suggested to overcome the limitations of a single approach [14]. This line of research includes the one that combines a geometric method with point tracking method [15] and PCA (principal component analysis) embedded template matching with optical flow [16]. By these methods, we can obtain high recognition rate if the original image is of high resolution and well refined. However, it is very difficult to determine the gaze of eyes by analyzing the eye ball rotations from a typical image in a low resolution, while a low-resolution image is helpful in implementing a real-time gaze recognition system. In addition, it is difficult to fit its contour reliably because the iris is partially occluded by the upper and lower eyelids [8].

In this paper, we deal with the gaze recognition problem for headlamp control of a vehicle. Although most gaze recognition systems focus on detecting and tracking the movement of eyes and irises, they require a lot of computation using high resolution images. On the other hand, in our study, because the gaze of a driver and the direction of the head are almost the same while driving and the operation of a headlamp control system requires to be smooth and real-time, we do not focus

The authors are with the Division of Electrical and Computer Engineering, Ajou University, Suwon, Korea.

This research was supported by the Korea Research Foundation Grant funded by Korean Government (KRF-2011-0005324).

on detecting and tracking irises. Instead, we focus on the detection and recognition of a head position using a relatively low-resolution face image to implement real-time headlamp control system. To this end, firstly, we produce classifiers to detect driver's face using the well known Viola-Jones (V-J) algorithm [17]. Then, to recognize the pose of a head, we use a holistic approach where each pixel of an image is interpreted as an input variable [18]. Because the number of pixels may be quite large and the redundant information has a high possibility of degrading the recognition performance, dimensionality reduction techniques such as feature extraction are typically used for holistic approaches. With feature extraction methods, we can not only handle feature vectors more effectively but also can solve the curse of dimensionality that occurs when the input dimension is high [19].

One of the popular feature extraction methods, LDA (linear discriminant analysis) was proposed for the classification problems which have discrete target variables [20] [21]. It tries to find a set of projections, from the original space to a lower dimensional space, which maximizes the between-class scatter while minimizing the within-class scatter [21]. Compared to classification problems, in regression problems, it is difficult to define between-class scatter and within-class scatter because target variables are continuous and LDA cannot be applied directly to regression problems. The simple concept that the samples with small differences in the target variables can be considered as belonging to the same class and the ones with large differences in the target variables can be considered as belonging to the different classes was applied to extend LDA to regression problems [22]. The algorithm is known as LDAr and in this paper, because the direction of gaze changes continuously, gaze recognition problem is treated as a regression problem and LDAr is applied to this problem.

In addition to the application of LDAr to the gaze recognition problem, in this paper, we also extend LDAr to a two dimensional version (2DLDAr) and applied it to the gaze recognition problem. The proposed two-dimensional feature extraction method can also be considered as an extension of 2DLDA [23] to a regressional version which uses the same concept that was used in extending LDA to LDAr.

After recognizing gaze of a driver by either LDAr or 2DLDAr, we transmit the mapping result of gaze direction to the headlamp controller using RS232c, which is a standard interface for relatively low speed serial data communication between computers and related devices defined by Electronic Industries Association (EIA). Then the headlamp controller controls the headlamp in the direction of the driver's gaze direction.

The paper is organized as follows. In Section II, we present a method of gaze recognition with a review on V-J algorithm and LDAr. In addition to these, we also present a new feature extraction method 2DLDAr that extends LDAr to a two dimensional image matrix and show how the proposed feature extraction method can be applied to recognize gaze of a driver. The gaze recognition performances of LDAr and the proposed 2DLDAr are compared with those of other conventional methods such as SVM [24], LDA, NLDA [25] and 2DLDA in Section III. Finally, Section IV completes the paper with a discussion and conclusions.

II. Methods

A. Overall architecture of the proposed headlamp control system

In this paper, we propose a method that robustly recognizes a driver's gaze to develop a headlamp control system. The overall architecture of the proposed gaze recognition system is shown in Fig. 1. A Microsoft LifeCam VX-1000 web camera was mounted on the front side of a driver to capture the facial image of the driver. To cope with the illumination change and night-time operation, the camera was converted to an infrared camera by attaching photo film in front of the camera. On the circumference of the camera, a number of IR LEDs were also attached. The camera generates 30 frames per second.

V-J algorithm was applied to the infrared images to produce classifiers that can detect faces. In this step, we trained frontal face and left half profile separately because it is difficult to consistently detect both the frontal view and the side view of a face at the same time. This is due to the fact that the Haar-like features of each view of face are quite different [26]. The right half profile could be detected by mirroring the entire image only when neither a frontal face nor a left half profile was detected.

After detecting face, the pose of the driver's head was recognized based on the feature extraction methods LDAr and 2DLDAr. Although the rotation angle of a face is a continuous value, the exact rotation angle was very difficult to measure. Therefore, on the deck in front of the driver's seat, seven positions with pre-defined angles were marked at which the driver was asked to stare. In this way, the training and test data were collected.

Finally, the output of driver's gaze direction was transmitted to the control box using RS-232c. In our system, the headlamp can rotate up to 15° in both left and right directions. Therefore, we set the output of driver's gaze direction as an integer number from -15 to 15. We map the output as -15, -10, -5, 0, +5, +10, and +15 which were used in classifying seven gaze directions. The dynamic bending headlamp rotates left and right by swivel Act and LIN (Local Interconnect Network) 1.3 was used to transmit angle data from the control box to the swivel Act.

In the following subsections, we describe each step of Fig. 1 in more detail.

B. Face Detection

V-J algorithm is a learning algorithm that produce classifiers that can classify objects based on Adaboost which uses Haar-like wavelet features. A Haar-like feature is composed of positive and negative rectangular image regions and a threshold. The difference between sums of positive and negative regions is compared with the threshold to weakly classify an image into two classes (object vs. non-object). In doing so, the so called *integral image* enables rapid computation of the sums of the positive and negative rectangular regions. With Haar-like features, V-J makes use of *Adaboost* to create strong binary classifiers that keep a high detection rate with a low false

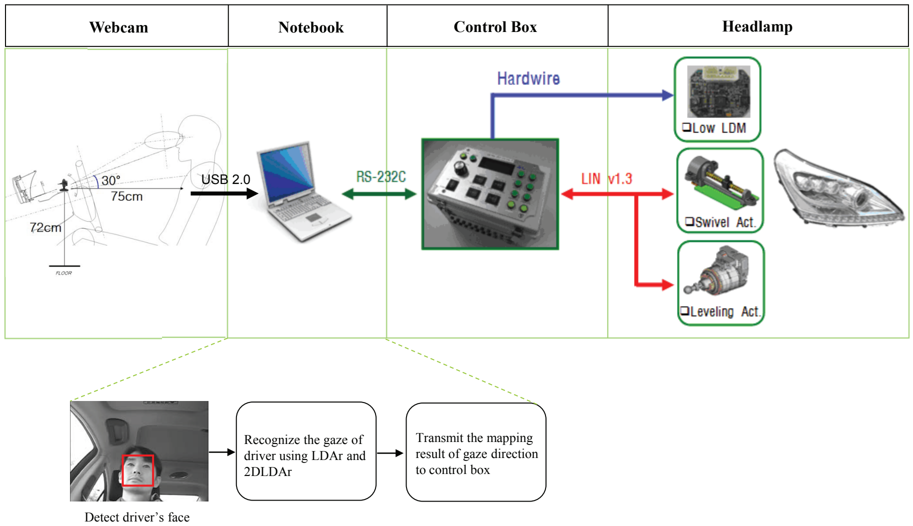

Fig. 1. The overall architecture of the proposed gaze recognition system

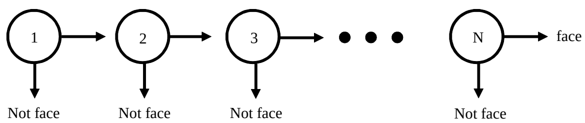

Fig. 2. Cascade of a series of strong classifiers

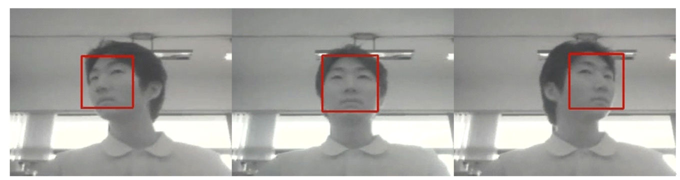

Fig. 3. Output of face detector

positive rate. It also achieves very fast detection of an object by cascading a series of strong classifiers through which most of the negative samples are rejected very fast at the first few stages.

In the training procedure, we compute an integral image which allows a very fast evaluation of Haar-like features in one pass over the entire image containing driver's face. Then, Haar-like features are used by the classifier. For a 20×20 subwindow, there could be more than 100,000 Haar-like features each of which corresponds to one weak classifier. Every Haar-like feature is evaluated using the integral images of the training images and important Haar-like features are selected and combined using Adaboost algorithm. In doing so, Adaboost algorithm selects a small number of critical Haar-like features and yields an efficient strong classifier.

Once a test image presented, the score of the image is computed by Adaboost using the selected Haar-like features. Finally, the decision of whether the area corresponds to a face or not is made based on the comparison of the score with the threshold.

The speed of V-J algorithm is accelerated by using the concept of cascading as shown in Fig. 2. In the figure, each node corresponds to a strong classifier which is made up of a group of weak classifiers combined by Adaboost algorithm. A positive result (face) from the previous node triggers the evaluation of a next node which has been adjusted to achieve high detection rates (e.g. over 99.8%) and low false positive rates (e.g. under 50%). We can detect most rigid objects such as faces, cars, bikes and human body by training new detectors [17] [27] [28] [29].

Although there are methods that detects various faces with different poses simultaneously [30] [31], they are slower than the sequential detection of frontal and half profile faces. Therefore, to speed up the face detection process, we applied different detectors sequentially. More specifically, in our implementation, we trained frontal face and left half profile separately because it is difficult to consistently detect both the frontal view and the side view of a face at the same time. The right half profile was detected by mirroring the entire image only when neither a frontal face nor a left half profile was detected. Fig. 3 shows some results of face detection.

C. Gaze recognition using feature extraction for regression problems

In many pattern recognition problems, feature extraction methods are typically used to reduce the dimensionality of the input space achieving better generalization performance with lower computational complexity. Even when the dimension of the input space is not so high, it is still useful to produce better generalization performance by reducing the effect of irrelevant or redundant variables [32].

A lot of feature extraction methods have been proposed among which subspace methods such as PCA [33] [34], ICA (independent component analysis) [35], and LDA [20] are very popular. These methods are further categorized into supervised and unsupervised methods based on whether target information is utilized or not. PCA and ICA are representative unsupervised methods, while LDA, NLDA (nullspace LDA) [25] and ICA-FX [36] are supervised ones. Conventional supervised feature extraction methods such as LDA and NLDA are mostly

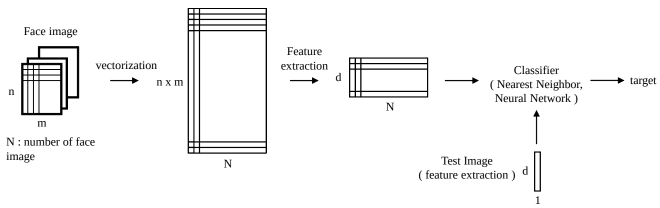

Fig. 4. Flowchart of gaze recognition by LDAr

used for classification problems taking class information attached to each training instance into account. On the other hand, LDAr which is designed for regression problems makes use of numerical output attached to each training instance.

Although the target value of our data is one of the seven categories due to technical difficulties of obtaining ground truth rotation angle and the gaze recognition problem can be considered as a classification problem with seven classes, considering that the rotation angle of a face is in continuous value, gaze recognition problem is more a regression problem than a classification problem. Therefore, in this paper, feature extraction methods designed for regression problems are used for gaze recognition. In the following, LDAr which is a linear feature extraction method for regression problems is described in detail and it is extended to a new two-dimensional feature extraction method for regression problems.

1) LDAr: Fig. 4 shows the flowchart of gaze recognition by LDAr. Firstly, N grayscale training face images of a size $n \times m$ are vectorized to form a $nm \times N$ matix. Because the dimension of a face image is generally so large that it cannot be processed easily, PCA is typically performed on the vectorized data. After PCA, we reduce the dimension to $d(< nm)$ by applying LDAr.

After extracting d features from an image, we obtain a $d \times N$ matrix from N training images as shown in the figure. When a new face image is presented, we can estimate the target value (gaze direction) with a classifier trained by this training matrix. Various estimation methods such as the nearest neighbor [37], neural networks or support vector machines [24] can be used in this step.

LDA finds projective directions by maximizing the ratio of between-class scatter matrix to within-class scatter matrix. In [22], this idea was extended to solve regression problems and a new feature extraction method LDAr for regression problems was proposed.

Consider a set of input/target pairs $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$. In this regression setting, we are to find a set of features f_i's ($= \boldsymbol{w_i}^T \boldsymbol{x}$), which are linear transformations of $\boldsymbol{x}$ such that they contain much information about the target variable y. To achieve this goal, Fisher's criterion was modified appropriately to fit in the regression problems in [22].

The classification problems have discrete target variables, or classes. Compared to classification problem, in regression problems, it is difficult to define between-class scatter and within-class scatter because target variables are continuous. In [22], the simple concept that the samples with small differences in the target variables can be considered as belonging to the same class while the ones with large differences in the target variables can be considered as belonging to the different classes was applied. By this idea of soft class, the between-class scatter matrix and within-class scatter matrix in LDA were modified as:

$$\begin{aligned} S_{wr} &= \frac{1}{n_w} \sum_{(i,j)\in A_{wr}} f(y_i - y_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \\ S_{br} &= \frac{1}{n_b} \sum_{(i,j)\in A_{br}} f(y_i - y_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T. \end{aligned} \tag{1}$$

where

$$\begin{aligned} A_{br} &= \{(i,j)| \quad |y_i - y_j| \geq \tau, i < j\} \\ A_{wr} &= \{(i,j)| \quad |y_i - y_j| < \tau, i < j\}. \end{aligned} \tag{2}$$

Here, the function $f(\cdot)$ is a weight function which takes on a positive value and τ is a threshold separating an index pair (i,j) into A_{br} and A_{wr}. The parameter n_b and n_w are the number of elements of A_{br} and A_{wr} respectively. The threshold τ can be represented as a multiple of the standard deviation of the target variable, i.e. $\tau = \alpha\sigma_y$. In [22], various weight functions were tested with $\alpha \in \{0.1, ..., 1.0\}$ and the performance did not depend much on the choice of weight function and α. As such, in this paper, α is set to 0.3 and the weight function $f(x) = \sqrt{||x| - \tau|}$ is used as in [22].

With the modified within and between-class scatter matrices, Fisher's criterion can be modified as:

$$W_{LDAr} = \underset{W}{\operatorname{argmax}} \frac{|W^T S_{br} W|}{|W^T S_{wr} W|}. \tag{3}$$

Maximizing the above Fisher's criterion is equivalent to solving the following generalized eigenvalue decomposition problem:

$$S_{br}\boldsymbol{w}_k = \lambda_k S_{wr}\boldsymbol{w}_k \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d. \tag{4}$$

Then, linear projections w_k's can be found.

2) 2DLDAr: In this part, we present a new method of feature extraction that extends LDAr to a two-dimensional version.

The 2DLDA (two-dimensional LDA) [23] which directly extracts features from image matrices without transforming the input image into one dimensional vector can also be extended to regressional version of feature extraction method named 2DLDAr (two-dimensional LDA for regression). Compared with LDA, 2DLDA can preserve the underlying two-dimensional data structures while LDA ignores the underlying structure. Furthermore, 2DLDA overcomes the singularity problem which is the limitation of LDA. In LDA, when the number of input space is larger than the number of training samples, a singularity problem occurs and to avoid this singularity, an intermediate dimensionality reduction stage using PCA is normally used. On the other hand, 2DLDA does not suffer from this singularity and it can be directly applied to image data.

The derivation of 2DLDAr is as follows. Firstly, the between-class scatter and within-class scatter matrices of 2DLDA are modified using the idea of soft class which is used in the derivation of LDAr. Now, the between-class scatter

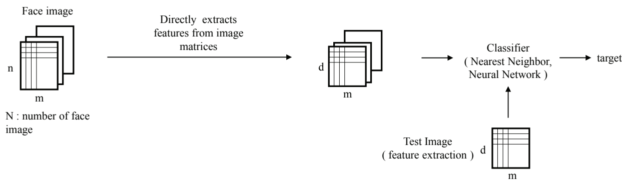

Fig. 5. Flowchart of gaze recognition by 2DLDAr

matrix and within-class scatter matirx in 2DLDAr are defined as:

$$\begin{aligned} R_{wr} &= \frac{1}{n_w} \sum_{(i,j)\in A_{wr}} f(y_i - y_j)(X_i - X_j)(X_i - X_j)^T \\ R_{br} &= \frac{1}{n_b} \sum_{(i,j)\in A_{br}} f(y_i - y_j)(X_i - X_j)(X_i - X_j)^T. \end{aligned} \tag{5}$$

where $X_i \in \Re^{n\times m}$ is the i-th image matrix and

$$\begin{aligned} A_{br} &= \{(i,j)| \quad |y_i - y_j| \geq \tau, i < j\} \\ A_{wr} &= \{(i,j)| \quad |y_i - y_j| < \tau, i < j\} \end{aligned} \tag{6}$$

are the sets of index pairs with different soft class labels and the same labels respectively. Given face images of size $n \times m$, the between-class scatter matrix and the within-class scatter matrix are $m \times m$.

With the modified within and between-class scatter matrices, Fisher's criterion can be modified as:

$$W_{2DLDAr} = \underset{W}{\text{argmax}} \frac{|W^T R_{br} W|}{|W^T R_{wr} W|}. \tag{7}$$

The linear projection w_k's can be obtained by solving the generalized eigenvalue decomposition problem in the same way as shown in equation 4.

Note that $W_{2DLDAr} \in \Re^{n\times d}$ where d is the number of extracted features and for an image matrix $X \in \Re^{n\times m}$, the projection becomes $F = W_{2DLDAr}^T X \in \Re^{d\times m}$. Therefore, for an image, the extracted feature is a matrix, not a vector.

Fig. 5 shows the flowchart of gaze recognition by 2DLDAr. Compared to LDAr, features are directly extracted from N grayscale image matrices of a size $n \times m$. Then, we obtain N feature matrices of a size $d \times m$ after extracting d features from image. In classification, the distance between given feature matrix P and Q is defined as:

$$d^2(P,Q) = \sum_{k=1}^{d} ||p_k - q_k||^2 \tag{8}$$

where $||p_k - q_k||$ denotes the Euclidean distance between the two feature vectors p_k and q_k.

3) Gaze recognition system: In this paper, we estimate the rotation angle of a driver's head using the features extracted by LDAr and 2DLDAr. The regression process is explained in Fig. 6. In the figure, there are two paths in both the training and the testing procedures which correspond to LDAr and 2DLDAr respectively. In the LDAr path, firstly, we normalize the training data so as to make each pixel have zero mean and unit variance. Then, PCA is performed on the vectorized data to project the higher-dimensional image space onto a lower dimensional sub-space. The weight W_{pca} which is obtained from the training process is used as an intermediate dimensionality reduction method for both training and test data. Note that if the original input dimension is not so high, we can omit the PCA step. After PCA, LDAr is applied. To obtain the projection matrix W_{LDAr} of LDAr, we first calculate the between-class scatter matrix S_{br} and the within-class scatter matrix S_{wr}. Then a set of generalized eigenvectors corresponding to the largest eigenvalues are used to form the most discriminant set of projection vectors W_{LDAr} for LDAr. The projection of test data is performed using the set of projection vectors W_{LDAr} obtained from the training process. Finally, the projected samples of training data and test data are obtained and regression is performed on these datasets using regression methods such as the k-nearest neighbor, multilayer perceptron (MLP) or support vector machine (SVM) [24].

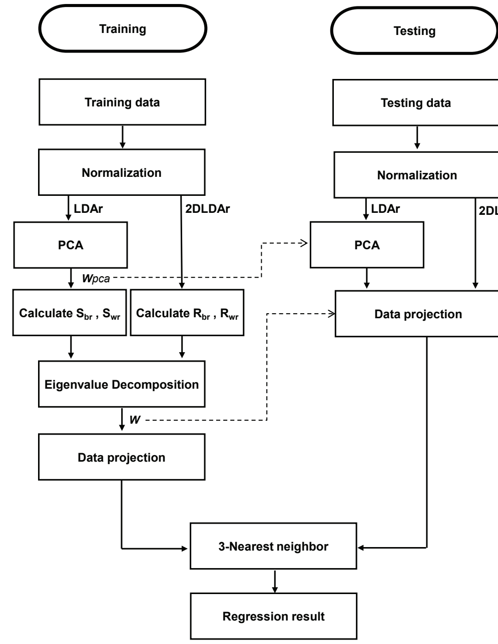

Fig. 6. Flowchart of regression process using LDAr and 2DLDAr

In 2DLDAr, instead of S_{wr} and S_{br}, R_{wr} and R_{br} are calculated directly from the image matrix without vectorization. In addition, PCA is not needed because 2DLDAr extracts features that form the most discriminant set of projection vectors from image matrices directly. The rest steps of the 2DLDAr process are the same as those of LDAr as shown in Fig.6.

As mentioned before, because the exact rotation angle of a head was very difficult to measure, we divided driver's gaze direction into seven categories as shown in Fig. 7. The samples of training data corresponding to the seven categories are shown in Fig. 8.

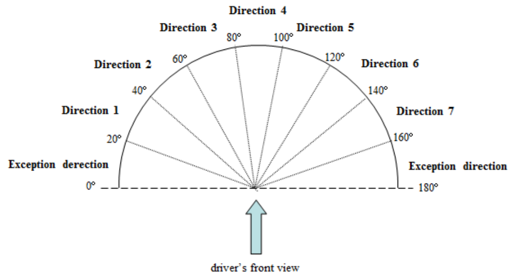

Fig. 7. Seven gaze directions

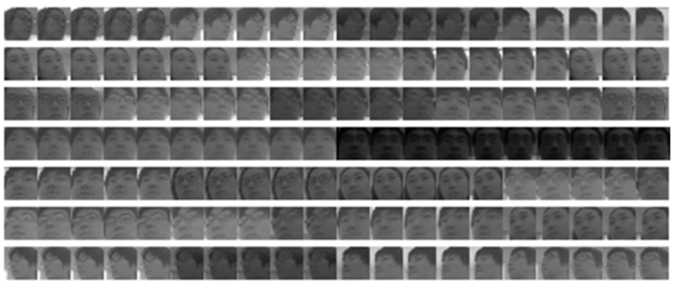

Fig. 8. Training data of gaze directions

III. Experimental results

In this section, we present the experimental results of the proposed gaze recognition system. The system was implemented on a Windows7 Home Premium / SP9300 (2.26 GHz) / 3 GB DDR3 / Geforce G 105M / 1.99 GB RAM platform and the size of the recorded image was 640 $\times$ 480 pixels per frame which was captured up to 30 frames per second. A camera film was attached to a Microsoft LifeCam VX-1000 web camera to make use of the infrared image. The camera was mounted on the opposite side of a driver to capture face images of the driver.

A. Face detection

Two face detectors were used in the experiment, each of which was trained by using either the frontal faces only or the left half profile images only. The third face detector which was designed to detect right half profile can be obtained by applying the left half profile detector on the mirrored face images.

Each of the two face detectors was trained using a set of training faces consisting of 1,000 face images (frontal and left half profile) and 10,000 non-face images scaled and aligned to a base resolution of 20×20 pixels. The non-face images were collected by selecting random area from background images which did not contain faces. The cascade was set to have 22 stages, where every stage is trained to have a high detection rate (over 99.8%) and low false alarm rate (under 50%). We chose the number of stages considering two perspectives; final false alarm rate and processing time. The false alarm rate of cascade of classifiers is below $0.5^{22}(2.3 \times 10^{-7})$. As the number of cascade of classifiers increases, the processing time increases accordingly.

To detect faces consistently regardless of the pose of a face, we applied the frontal face detector, the left half profile detector, and the right half profile detector sequentially. The second and the third detectors were not activated when a face was detected by the previous detector.

The experimental results of detection rates (correctly detected frames/total number of frames) and false alarm rates (false alarms/total number of frames) using V-J algorithm under various poses of face images are reported in Table I. Once a face is detected by the V-J algorithm, it is regarded as a correct detection if the location and size of the face is correctly found, otherwise, it is regarded as a false alarm. The detection rate is computed by the ratio of the number of correct detections and the total number of frames in the dataset. The false alarm rate is computed by the ratio of the number of false alarms and the total number of frames in the dataset. In the experiment, we captured the infrared images of six individuals with various poses. The number of test images per each individual might be different because of random selection. In the first three rows, the detection rates of each of the three detector are reported. The detection rates of the proposed combined face detector are shown in the last row. We also report the numbers of correctly detected faces versus the total numbers of test images in the parentheses. Likewise, false alarm rates for each detector are also included. The detection rates were obtained by checking the output of the detector manually and there are two types of error: the first one being the case where the detector cannot find face from the face image (miss) and the second being the case where the detector finds a face in a wrong position (false alarm).

As can be seen in the table, each single detector has high recognition rate for the designed pose. However, as the difference between the designed pose and the pose of the test image increases, the performance degrades drastically to less than 10%. This is because the Haar-like features which are used to detect different poses of a face are quite different. Even though the false alarm rates of proposed method is slightly higher than each of the three single detectors, the classifier applied in sequence has many advantages compared to simultaneous application of multiple detectors. Compared to the simultaneous approach, the proposed sequential detectors obtain high detection rates for each pose with a reduced processing time. Furthermore, if we use simultaneous approach, it can be difficult to determine gaze direction when more than one detector detect a face at the same time.

In Table I, the detection rates of the proposed method for pose 1 and 2 are slightly lower than those of the right face detector. The reason can be attributed to the fact that in the combined detector, the frontal face detector and the left face detector are activated before the right face detector and if the wrong face is detected in these steps, the right face detector is not activated. If the right face detector was applied to these wrong detected face images, the correct facial region could have been located. The same argument can explain a slight lower detection rate of the proposed method than that of the left face detector for pose 7.

By using the combined detector, we can obtain detection rates ranging from 93.69% to 99.98% for each pose with the

TABLE I
DETECTION AND FALSE ALARM RATES UNDER VARIOUS POSES OF FACE IMAGES (%)

Type of detectors \ Pose		1	2	3	4	5	6	7
Frontal face detector	Detection rate	0.13 (5/3724)	20.42 (704/3448)	65.99 (2652/4019)	99.67 (4810/4826)	70.05 (2454/3503)	22.32 (905/4054)	0.06 (2/3423)
	False alarm rate	0.24 (9/3724)	0.09 (3/3448)	0.02 (1/4019)	0.08 (4/4826)	0.17 (6/3503)	0.91 (37/4054)	1.29 (44/3423)
Left face detector	Detection rate	0.05 (2/3724)	0.00 (0/3448)	1.52 (61/4019)	12.64 (610/4826)	93.58 (3278/3503)	99.43 (4031/4054)	99.15 (3394/3423)
	False alarm rate	0.21 (8/3724)	0.00 (0/3448)	0.00 (0/4019)	0.06 (3/4826)	0.14 (5/3503)	0.39 (16/4054)	0.32 (11/3423)
Right face detector	Detection rate	94.12 (3505/3724)	99.51 (3431/3448)	97.19 (3906/4019)	24.37 (1176/4826)	2.88 (101/3503)	0.15 (6/4054)	1.75 (60/3423)
	False alarm rate	0.16 (6/3724)	0.20 (7/3448)	0.02 (1/4019)	0.10 (5/4826)	0.09 (3/3503)	0.05 (2/4054)	0.26 (9/3423)
Proposed method (combined)	Detection rate	93.69 (3489/3724)	99.42 (3428/3448)	99.98 (4018/4019)	99.88 (4820/4826)	99.31 (3479/3503)	98.47 (3992/4054)	97.78 (3347/3423)
	False alarm rate	0.43 (16/3724)	0.23 (8/3448)	0.02 (1/4019)	0.12 (6/4826)	0.26 (9/3503)	1.06 (43/4054)	1.46 (50/3423)

TABLE II
THE PERFORMANCE OF GAZE RECOGNITION USING 3-NEAREST NEIGHBOR FOR 2800 IMAGES(5-FOLD CROSS-VALIDATION)

Total images : 400 × 7 directions = 2800, 5-fold cross-validation							
Methods \ No. of features		1	2	3	4	5	6
NLDA	Recognition rate(%)	19.04 (2.72)	27.54 (4.09)	35.00 (5.69)	48.82 (5.40)	58.71 (5.47)	69.36 (3.88)
	Test rms error	2.69 (0.12)	2.13 (0.51)	1.98 (0.43)	1.73 (0.44)	1.47 (0.35)	1.29 (0.27)
LDA	Recognition rate(%)	52.11 (1.26)	77.32 (2.36)	88.18 (1.12)	92.89 (2.10)	95.89 (0.84)	96.82 (1.11)
	Test rms error	0.94 (0.05)	0.56 (0.05)	0.36 (0.03)	0.29 (0.06)	0.23 (0.05)	0.21 (0.05)
2DLDA	Recognition rate(%)	53.57 (1.80)	76.57 (2.69)	86.07 (1.45)	91.96 (1.61)	94.39 (1.78)	95.68 (1.23)
	Test rms error	0.89 (0.02)	0.54 (0.03)	0.40 (0.03)	0.31 (0.04)	0.27 (0.05)	0.26 (0.05)
LDAr	Recognition rate(%)	61.50 (2.66)	86.61 (1.09)	94.14 (1.59)	96.46 (1.71)	97.64 (0.87)	**98.36** (0.56)
	Test rms error	0.64 (0.02)	0.36 (0.02)	0.25 (0.04)	0.19 (0.05)	0.16 (0.04)	0.15 (0.04)
2DLDAr	Recognition rate(%)	57.96 (1.36)	81.18 (1.83)	90.32 (2.41)	96.07 (1.08)	97.64 (0.51)	**98.39** (0.59)
	Test rms error	0.81 (0.01)	0.49 (0.03)	0.33 (0.05)	0.22 (0.04)	0.18 (0.03)	0.16 (0.04)

TABLE III
THE PERFORMANCE OF GAZE RECOGNITION USING 3-NEAREST NEIGHBOR FOR 1400 IMAGES(5-FOLD CROSS-VALIDATION)

Total images : 200 × 7 directions = 1400, 5-fold cross-validation							
Methods \ No. of features		1	2	3	4	5	6
NLDA	Recognition rate(%)	15.86 (3.17)	24.64 (2.85)	32.07 (2.81)	40.21 (4.97)	50.43 (7.18)	58.29 (9.99)
	Test rms error	2.61 (0.28)	2.50 (0.30)	2.31 (0.26)	2.01 (0.20)	1.83 (0.18)	1.62 (0.30)
LDA	Recognition rate(%)	51.29 (4.33)	78.93 (4.16)	86.71 (1.59)	90.79 (3.49)	95.14 (3.80)	95.43 (3.14)
	Test rms error	0.96 (0.09)	0.51 (0.09)	0.40 (0.05)	0.33 (0.05)	0.23 (0.15)	0.23 (0.14)
2DLDA	Recognition rate(%)	51.00 (2.22)	77.64 (2.64)	85.50 (2.43)	90.00 (3.22)	93.21 (1.80)	93.57 (1.71)
	Test rms error	0.98 (0.06)	0.53 (0.02)	0.42 (0.06)	0.37 (0.06)	0.33 (0.06)	0.31 (0.06)
LDAr	Recognition rate(%)	58.21 (2.15)	85.00 (2.21)	90.50 (2.57)	94.14 (1.49)	96.07 (1.34)	**96.57** (1.09)
	Test rms error	0.73 (0.01)	0.40 (0.04)	0.34 (0.04)	0.29 (0.05)	0.26 (0.05)	0.27 (0.03)
2DLDAr	Recognition rate(%)	54.14 (2.15)	80.21 (3.85)	87.29 (4.22)	93.64 (1.30)	95.86 (1.35)	**96.64** (1.20)
	Test rms error	0.88 (0.03)	0.51 (0.05)	0.41 (0.08)	0.31 (0.04)	0.28 (0.06)	0.28 (0.04)

processing time of each frame being 43.49 msec on average (23 frames per second) on a 2.26 GHz CPU, 1.99 GB RAM.

B. Gaze recognition using LDAr and 2DLDAr

Once the driver's facial region was detected, we classified the gaze into seven directions as shown in Fig. 7. In this step, we used LDAr and 2DLDAr to extract important features and compared the correct classification rates of the methods with those of the conventional LDA, NLDA [25] and 2DLDA in Table II and III. In both tables standard deviations are denoted in the parentheses. The recognition rate is computed by the ratio of the number of frames which estimates gaze direction correctly to that of total frames in the dataset. For classification, we used the k-nearest neighbor (k-NN) algorithm using Euclidean norm. In 2DLDAr, the distance between given feature matrix is defined using Euclidean norm as shown in (8). We varied the number of k in k-NN classifier and observed that $k = 3$ achieves comparatively good performance. As such, 3-NN classifier is used throughout the paper.

The training and test data consist of six individuals and the number of images per each individual might be different for both the training and test data because of random selection. Various numbers of features ranging from 1 to 6 were extracted. For LDAr and 2DLDAr, we set target variable as integer value ranging from 1 to 7. The target variable $i \in \{1, \cdots, 7\}$ corresponds to the Direction i in Fig. 7.

The correct detect rate of gaze recognition using 5-fold cross-validation is shown in Table II. Totally 2800 images were collected for 5-fold cross-valididation which consists of 400 images per each gaze direction. The training data and testing data is randomly selected from 2800 images with same number of images per each gaze direction. The experimental results with 1400 images which contain 200 images per each direction are also shown in the table III.

The LDAr gave recognition rates of 61.50% to 98.36% and 58.21% to 96.57% and 2DLDAr gave recognition rates of 57.96% to 98.39% and 54.14% to 96.64% for different numbers of features. The best recognition rates were obtained by LDAr and 2DLDAr when 6 features were used for both

TABLE IV
THE PERFORMANCE OF GAZE RECOGNITION USING MLP(5-FOLD CROSS-VALIDATION)

Data (total 1,400)	Training (80%)		Validation (10%)		Test (10%)	
Measure	RMSE	R	RMSE	R	RMSE	R
No. of features	2					
LDA	0.39	0.9810	0.37	0.9832	0.40	0.9792
2DLDA	0.45	0.9748	0.46	0.9729	0.46	0.9723
LDAr	0.24	0.9926	0.22	0.9940	0.23	0.9934
2DLDAr	0.28	0.9900	0.28	0.9902	0.28	0.9905
No. of features	3					
LDA	0.32	0.9873	0.33	0.9863	0.34	0.9855
2DLDA	0.35	0.9847	0.35	0.9842	0.37	0.9829
LDAr	0.16	0.9966	0.17	0.9962	0.17	0.9965
2DLDAr	0.20	0.9950	0.19	0.9954	0.21	0.9948
No. of features	4					
LDA	0.30	0.9887	0.30	0.9888	0.29	0.9899
2DLDA	0.35	0.9851	0.33	0.9861	0.36	0.9837
LDAr	0.15	0.9973	0.14	0.9974	0.16	0.9967
2DLDAr	0.14	0.9974	0.16	0.9968	0.16	0.9967
No. of features	5					
LDA	0.27	0.9909	0.28	0.9907	0.36	0.9875
2DLDA	0.34	0.9859	0.35	0.9848	0.33	0.9848
LDAr	0.16	0.9969	0.17	0.9967	0.14	0.9974
2DLDAr	0.13	0.9979	0.14	0.9974	0.14	0.9975
No. of features	6					
LDA	0.27	0.9908	0.25	0.9923	0.27	0.9907
2DLDA	0.32	0.9867	0.31	0.9882	0.32	0.9871
LDAr	0.09	0.9989	0.09	0.9990	0.11	0.9986
2DLDAr	0.10	0.9987	0.12	0.9983	0.11	0.9986

experiments. With 3NN classifier, the best recognition rates were 98.39% and 96.64% respectively. The recognition rate and test rms (root mean squared) error of LDAr and 2DLDAr is superior to NLDA, LDA and 2DLDA for any number of features. In conclusion, the performance of regressional versions of feature extraction are better than the conventional feature extraction methods.

For comparison, SVM [24] was also applied to the original 400 dimensional input space. Before applying SVM, each of the 400 input variables was normalized to have zero mean and unit variance. The SVM-KM toolbox for Matlab [38] was used. Both the polynomial and the Gaussian kernels were used with various kernel parameters (the degree of polynomial for polynomial kernel and σ for Gaussian kernel). The best performance of SVM using 5-fold cross-validation for 2800 images was 98.93% with the polynomial kernel (degree = 2) and 98.93% with Gaussian kernel (σ=1.0). Similarly, for the 1400 images case, the best classification rate was 97.93% with the polynomial kernel (degree=2) and 97.29% with Gaussian kernel (σ=1.0). Although these values are slightly better than those of LDAr and 2DLDAr in Table II and III, the performance of SVM was very sensitive to kernel parameters. Compared to the proposed method of combined feature extraction (LDAr or 2DLDAr) and simple classifier (3-NN), the method of applying SVM directly to the original input variables has a disadvantage that finding appropriate parameters for SVM is difficult and time consuming.

Table IV shows the performance of gaze recognition using multilayer perceptron (MLP). The 5-fold cross-validation is used to evaluate the performance. Two layer MLP was used and the number of hidden neurons was set to four. We trained

TABLE V
ONE TAILED WELCH'S t-TEST

Compare LDA and LDAr						
No. of features	1	2	3	4	5	6
T-value	7.134	7.991	8.175	2.948	3.236	2.770
d.o.f.	6	6	4	8	8	6
T_{99}%	3.413	3.413	3.747	2.896	2.896	3.413
T_{95}%	1.943	1.943	2.132	1.860	1.860	1.943
Accepted (99%)	H_{A1}	H_{A1}	H_{A1}	H_{A1}	H_{A1}	H_{O1}
Accepted (95%)	H_{A1}	H_{A1}	H_{A1}	H_{A1}	H_{A1}	H_{A1}
Compare 2DLDA and 2DLDAr						
No. of features	1	2	3	4	5	6
T-value	4.351	3.168	3.379	4.740	3.925	4.442
d.o.f.	7	7	7	7	5	6
T_{99}%	2.998	2.998	2.998	2.998	3.365	3.413
T_{95}%	1.895	1.895	1.895	1.895	2.015	1.943
Accepted (99%)	H_{A2}	H_{A2}	H_{A2}	H_{A2}	H_{A2}	H_{A2}
Accepted (95%)	H_{A2}	H_{A2}	H_{A2}	H_{A2}	H_{A2}	H_{A2}

MLP using Levenberg-Marquardt algorithm and evaluated performance using root mean squared error (RMSE) and regression (R) value. The RMSE is the square root of the average squared difference between the output and the target values. A lower RMSE values are better and zero RMSE means there is no error. The R values measure the correlation between the output and the target values. R value of 1 means a close linear relationship, while 0 means a random or no linear relationship. As can be seen in Table IV, RMSE values of LDAr and 2DLDAr are much smaller than those of LDA and 2DLDA regardless of the number of extracted features. In the same manner, regardless of the number of extracted features, R values of LDAr and 2DLDAr are greater than those of LDA and 2DLDA. From this, we can see that LDAr and 2DLDAr improves the performance of gaze recognition regardless of the number of features.

We also performed one tailed Welch's t-test [39] to show the statistical significance of our methods as shown in Table V. The null (H_O) and the alternative (H_A) hypotheses are as follows:

- H_{O1}: For a fixed number of features, the performance of LDA and LDAr are the same.
- H_{A1}: For a fixed number of features, LDAr outperforms LDA.
- H_{O2}: For a fixed number of features, the performance of 2DLDA and 2DLDAr are the same.
- H_{A2}: For a fixed number of features, 2DLDAr outperforms 2DLDA.

We calculate the T-value, degree of freedom (d.o.f.) and the corresponding target T-values using the recognition rates in Table II. As shown in Table V, the null hypothesis was rejected, thus the alternative hypothesis was accepted for all the number of features when the confidence level was 95%. If the confidence level was 99%, the null hypothesis was rejected for all the number of features except for the case of 6 extracted features. From this, we can conclude that the regressional version of feature extraction method outperforms other feature extraction methods.

The time complexity of the proposed gaze recognition system is around 1.32 msec/frame on average on a 2.26 GHz CPU, 1.99 GB RAM. The total frame rate of the system which include face detection and gaze recognition is on average around 44.91 msec/frame (43.59 msec to detect face using V-J and 1.32 msec to recognize the direction of face) which is enough for real time demands on a 2.26 GHz CPU, 1.99 GB RAM.

IV. Conclusions

This paper addresses a method for gaze recognition of a driver coping with rotation of a driver's face. Because the gaze of a driver and the direction of the head are almost the same while driving and the operation of a headlamp control system requires to be smooth and real-time, unlike other gaze recognition researches, we focus on the detection of a head and the recognition of the pose of the head using a relatively low-resolution face image.

As a result, our gaze recognition system is mainly divided into two parts: face detection and gaze recognition systems. As a face detection method, we used V-J algorithm to produce classifiers that can detect faces. In doing so, three different face detectors designed to detect frontal face, left profile and the right profile respectively were sequentially used. Compared to the simultaneous application of different detectors, the sequential use of the different face detectors can speed up the detection process without much degrading the detection rate.

After detecting face, to extract good features from the original input variables, we used LDAr, a regressional version of linear discriminant analysis, which tries to maximize the ratio of inter-distances among samples with large differences in target value and those with small differences in the target value. In addition, we also proposed a new two-dimensional feature extraction method 2DLDAr which extends 2DLDA to a regressional version and applied it to gaze recognition problem. The gaze recognition performances of LDAr and 2DLDAr exceed those of the conventional LDA, NLDA and 2DLDA.

Based on various application areas of the gaze recognition such as intelligent vehicles, medical and integrated human computer interface, the resultant systems are quite different in their assumptions and the direct comparison with other gaze recognition methods was out of our ability. Although we could not compare the performance of the proposed method with those of other researches, the performance of the proposed gaze recognition system was good enough to be applied to the headlamp control of a vehicle.

As a future work, we would like to use tracking methods to efficiently cope with false alarm. The proposed approach can also be applied to similar problems such as airbag ignition problem.

References

[1] R.A. Hancck and R. Rarasuraman, "Human factors and safety in the design of intelligent vehicle-highway systems (IVHS)," *Journal of Safety Research*, vol. 23, no. 4, pp. 181 – 198, Winter 1992.

[2] J. Pérez, F. Seco, V. Milanés, A. Jiménez, J. Díaz, and T. de Pedro, "An RFID-based intelligent vehicle speed controller using active traffic signals," *Sensors*, vol. 10, no. 6, pp. 5872 – 5887, 2010.

[3] C. Premebida, O. Ludwig, and U. Nunes, "Lidar and vision-based pedestrian detection system," *Journal of Field Robotics*, vol. 26, no. 9, pp. 696 – 711, 2009.

[4] S. Hsu, S. Acharya, A. Rafii, and R. New, "Performance of a time-of-flight range camera for intelligent vehicle safety applications," *Advanced Microsystems for Automotive Applications 2006*, pp. 205 – 219, 2006.

[5] C.E. Smith, C.A. Richards, S.A. Brandt, and N.P. Pananikolopoulos, "Visual tracking for intelligent vehicle-highway systems," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 4, pp. 744 – 759, Nov. 1996.

[6] R. Bishop, "A survey of intelligent vehicle applications worldwide," in *Proc. IEEE Intelligent Vehicles Symposium 2000*, Oct. 2000, pp. 25 – 30.

[7] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, May 2002, pp. 124 – 129.

[8] J.G. Wang and E. Sung, "Study on eye gaze estimation," *IEEE Trans. System, Man, and Cybernetics-Part B: Cybernetics*, vol. 32, no. 3, pp. 332–350, June 2002.

[9] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive driver gaze tracking with active appearance models," in *Proc. 11th World Congress on Intelligent Transportation Systems*, Oct. 2004.

[10] A.E. Kaufman, A. Bandopadhay, and B.D. Shaviv, "An eye tracking computer user interface," in *Proc. IEEE 1993 Symposium on Research Frontier Virtual Reality Workshop*, Oct. 1993, pp. 120 – 121.

[11] G.A. Myers, K.R. Sherman, and L. Stark, "Eye monitor," *IEEE Computer Magazine*, vol. 3, pp. 14 – 21, 1991.

[12] B. Hu and M.H. Qiu, "A new method for human-computer interaction by using eye gaze," in *Proc. IEEE Int. Conf. System, Man, and Cybernetics*, Oct. 1994, pp. 2723 – 2728.

[13] R. Stiefelhagen and J. Yang, "Gaze tracking for multimodal human-computer interaction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*, 1997, vol. 4, pp. 2617 – 2620.

[14] E. Murphy-Chutorian and M.M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607 – 626, Apr. 2009.

[15] J. Heinzmann and A. Zelinsky, "3-D facial pose and gaze point estimation using a robust real-time tracking paradigm," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 1998, pp. 142 – 147.

[16] Y. Zhu and K. Fujimura, "Head pose estimation for driver monitoring," *IEEE Intelligent Vehicle Symposium.*, pp. 501 – 506, 2004.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition*, vol. 1, pp. 511 – 518, 2001.

[18] K.-M. Lam and H. Yan, "An analytic-to-holistic approach for face recognition based on a single frontal view," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 673–686, 1998.

[19] S. Yan, D. Xu, B. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[20] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997.

[21] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[22] N. Kwak and J.-W. Lee, "Feature extraction based on subspace methods for regression problems," *Neurocomputing*, vol. 73, no. 10, pp. 1740 – 1751, June 2010.

[23] J. Ye, R. Janardan, and Q. Li, "Two-Dimensional Linear Discriminant Analysis," in *NIPS*, 2004.

[24] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, chapter 7–12, pp. 187–404, The MIT Press, 2002.

[25] J. Ye and T. Xiong, "Null space versus orthogonal linear discriminant analysis," in *Proc. ACM Int. Conf. Mach. Learning*, 2006, pp. 1073 – 1080.

[26] Xiao R., Li M.J., and Zhang H.J., "Robust Multi-Pose Face Detection in Images," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 14, no. 1, pp. 31 – 41, 2004.

[27] G. Monteiro, P. Peixoto, and U. Nunes, "Vision-based pedestrian detection using Haar-like features," in *Proc. 6th National Festival of Robotics, Scientific Meeting (ROBOTICA)*, 2006.

[28] B. Stanciulescu, A. Breheret, and F. Moutarde, "Introducing new Adaboost features for real-time vehicle detection," in *Proc. Cognitive systems with Interactive Sensors*, 2007.

[29] F. Moutarde, B. Stanciulescu, and A. Breheret, "Real-time visual detection of vehicles and pedestrians with new efficient adaboost features," in *Proc. IEEE International Conference on Intelligent Robots Systems*, 2008.
[30] M. Viola, M.J. Jones, and P. Viola, "Fast Multi-view Face Detection," in *Proc. of Computer Vision and Pattern Recognition*, 2003.
[31] S.Z. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum, "Statistical learning of multi-view face detection," in *Proc. of the 7th European Conference on Computer Vision*, 2002, pp. 67–81.
[32] P.A. Devijver and J. Kittler, *Pattern recognition: A statistical approach*, Prentice Hall, London, 1982.
[33] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, chapter 9–10, pp. 399–507, Academic Press, second edition, 1990.
[34] M. Turk and A. Pentland, "Face recognition using Eigenfaces," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
[35] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, June 1995.
[36] N. Kwak and C.-H. Choi, "Feature extraction based on ica for binary classification problems,," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 6, pp. 1374 – 1388, Nov. 2003.
[37] E.A. Narayada, "On estimating regression," *Theory Probab. Appl.*, vol. 9, pp. 141–142, 1964.
[38] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and Kernel Methods Matlab Toolbox," Perception Systemes et Information, INSA de Rouen, Rouen, France, 2005.
[39] B. L. Welch, "The generalization of 'Student's' Problem when Several Different Population Variances are Involved," *Biometrika*, vol. 34, no. 1 – 2, pp. 28 – 35, Jan. 1947.

Jae Hyun Oh received the BS and MS degrees from the Department of Electrical Engineering, Ajou University, Suwon, Korea, in 2006 and 2008 respectively. He is currently working towards the Ph.D. degree in the same university. His research interests include machine learning, statistical analysis, pattern recognition and their applications.

Nojun Kwak received the BS, MS, and PhD degrees from the School of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea, in 1997, 1999 and 2003 respectively. From 2003 to 2006, he worked for Samsung Electronics. In 2006, he joined Seoul National University as a BK21 assistant professor. Since 2007, he has been with the Department of Electrical Engineering in Ajou University, Suwon, Korea where he works as an associate professor. His research interests include pattern recognition, machine learning, computer vision, data mining, image processing, and their applications.