International Journal of Fuzzy Logic and Intelligent Systems, vol. 10, no. 2, June 2010, pp. 95-100

A Spatial Regularization of LDA for Face Recognition

Lae-Jeong Park

Department of Electronics Engineering, Gangnung-Wonju National University
123 Chibyun-Dong, Kangnung, 210-702, Korea
Tel : +82-33-640-2389, Fax : +82-33-646-0740, E-mail : ljpark@gwnu.ac.kr

Abstract

This paper proposes a new spatial regularization of Fisher linear discriminant analysis (LDA) to reduce the overfitting due to small size sample (SSS) problem in face recognition. Many regularized LDAs have been proposed to alleviate the overfitting by regularizing an estimate of the within-class scatter matrix. Spatial regularization methods have been suggested that make the discriminant vectors spatially smooth, leading to mitigation of the overfitting. As a generalized version of the spatially regularized LDA, the proposed regularized LDA utilizes the non-uniformity of spatial correlation structures in face images in adding a spatial smoothness constraint into an LDA framework. The region-dependent spatial regularization is advantageous for capturing the non-flat spatial correlation structure within face image as well as obtaining a spatially smooth projection of LDA. Experimental results on public face databases such as ORL and CMU PIE show that the proposed regularized LDA performs well especially when the number of training images per individual is quite small, compared with other regularized LDAs.

Key Words : Face recognition, LDA, regularization

1. Introduction

Fisher linear discriminant analysis (LDA)[1] is a very popular subspace algorithm in face recognition to accomplish the tasks of dimensional reduction and feature extraction in face recognition[2]. LDA is a supervised learning algorithm that maximizes a ratio of the determinant of the between-class scatter matrix S_B to the within-class scatter matrix S_W of the training samples. Unfortunately, LDA suffers from the so-called small sample size (SSS) problem[3] that arises when the number of training samples is smaller than the dimensionality of the sample space, which often happens in face recognition. In the case, S_W becomes singular, resulting in difficulty in calculating discriminant vectors that maximize the Fisher criterion. One solution for the singularity is to reduce the dimensionality of the original sample space. Several LDA algorithms such as Fisherface have been proposed to solve the singularity problem[4-6]. The algorithms make the samples in the original high dimensional space be mapped into a lower dimensional space in which S_W is nonsingular.

Another direction is to regularize S_W. Friedman[7] proposed a regularized LDA (R-LDA) that adds some small constant values to the diagonal elements of S_W. R-LDA makes it possible to resolve the singularity and, at the same time to use discriminatory information in the null space of S_W. Even though R-LDA overcomes the singularity of S_W, there remains another problem of LDA in face recognition. LDA algorithms are plagued by the overfitting problem that occurs when the number of training images per individual is small, which is not unusual in face recognition. For example, given 30 persons and 3 facial images of 40×40 size for each person, LDA algorithms usually find a set of 29 discriminant vectors in the 1600-dimensional space. Ninety samples are too small in size to determine a total of 46400 (=29×1600) parameters, hence overfitting is unavoidable. Since advent of R-LDA, several regularized LDA algorithms have been proposed to attack the overiftting [8-10]. Their approach is to regularize an estimate of S_W that is inaccurate due to a small-sized training data set in a such way that it gets less overfitted, insensitive to the data set. They have shown good generalization performance when the number of training samples is small.

Some LDA algorithms have also been introduced that take advantage of the two-dimensionality or spatial correlations of neighboring pixels in face images[11,12]. They are quite reasonable because a face image is originally a two-dimensional data. They benefit from mitigation of the overfitting in terms of reduction of the number of parameters. The spatially regularized LDA algorithms (S-LDA) make it possible to produce the discriminant vectors spatially smooth in a matrix form by incorporating a two-dimensional smoothness constraint into the Fisher criterion. S-LDA has been reported to outperform LDA algorithms such as R-LDA and 2D-LDA[13].

In this paper, a new spatial regularization method is introduced to alleviate the overfitting in face recognition. The proposed regularized LDA, a generalized version of S-LDA, utilizes the nonuniformity of spatial correlation structures in face images, which has not been considered in previous regularization methods. In general, spatial frequency

Manuscript received Oct. 15. 2009; revised May. 6, 2010; Accepted May. 12, 2010.
This research was supported by Gangnung-Wonju National University Research Grant of 2008.

information in a face image differs from a local region to another. Local regions of eyes and nose have more high-frequency information than other regions such as cheek and hair. Therefore, it would be useful to take advantage of the nonuniform spatial characteristic of face images in spatially regularized LDA algorithms because it enables us to make the discriminant vectors capture the non-flat spatial correlation structures explicitly. In a nonuniform spatially regularized LDA (NS-LDA), a two-dimension smoothness constraint whose strength is different from a pixel to another is incorporated into a LDA framework. The pixel-by-pixel regularization strengths are determined based on the within-class variation measured from training samples.

The next Section briefly reviews LDA and regularized LDA. In Section 3, a formulation of NS-LDA is presented, along with how to determine the nonuniform regularization parameters. Experimental results on the face databases are presented in Section 4 before drawing conclusions in Section 5.

2. A review of LDA and regularized LDA

2.1 LDA

Given a set of N d-dimensional samples $\{x_1,...,x_N\}$ belonging to C different classes, with N_i samples in the subset D_i, $i=1,...,C$. The objective of LDA is to find a linear transform Φ or a set of ϕ's that maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix. The criterion of LDA is given by

$$\max_{\phi} \frac{\left|\phi^T S_B \phi\right|}{\left|\phi^T S_W \phi\right|}, \tag{1}$$

where S_B and S_W are the between-class and within-class scatter matrices, respectively. The optimal linear projection Φ of Eq. (1) is a set of the generalized eigenvectors ϕ's of S_B and S_W, i.e.,

$$S_B \phi = \lambda S_W \phi . \tag{2}$$

When S_W is nonsingular, Φ is a set of the eigenvectors of $S_W^{-1} S_B$. However, S_W is singular when N is smaller than d. Fisherface[4] has been developed to resolve the singularity, followed by D-LDA[6] and N-LDA[5] that have been proposed to overcome the weak point of Fisherface.

2.2 Regularized LDA(R-LDA)

The singularity of LDA due to the SSS problem can be resolved by regularization of S_W. In [7], a simple regularized LDA (R-LDA) has been introduced in which a small value $\alpha(\alpha > 0)$ is added to each of the diagonal elements of S_W so that Φ results in a set of eigenvectors of $(S_W + \alpha I)^{-1} S_B$, where I is an identity matrix. The R-LDA resolves the singularity because $(S_W + \alpha I)$ is full rank and thereby uses discriminatory information in the null space of S_W. The smaller α is, the more ϕ relies on the null space of S_W. In this way, α is able to determine an appropriate trade-off between the range space and the null space.

Some regularization methods have been introduced to modify or regularize some unreliable eigenvalues for reduction of overfitting and better generalization. A two-parameter regularized LDA has been proposed to exploit the range and null space of S_W by separately regularizing the eigenvalues of the two spaces[9]. Jiang *et. al.* proposed an eigenvalue regularization that differently regularizes eigenvalues based on an eigenspectrum model to alleviate the problems of instability and overfitting[10].

2.3 Spatially regularized LDA(S-LDA)

Unlike the eigenvalue regularization methods above, a spatial smoothness regularization approach[11,12] aim to regularize the eigenvectors of S_W explicitly by using a two-dimensional (2-D) smoothness constraint. T. Hastie, *et. al.* proposed a penalized version of LDA designed for classification problems such as handwritten digit recognition[11]. The spatial smoothness regularization based on a 2-D smoothing penalty for ϕ's has recently been revisited and successfully applied in face recognition[12]. The reason why S-LDA is effective over other regularized LDAs is that it can reduce the number of degree of freedom by utilizing high correlation structure in face images. As there exists high spatial correlation of neighboring pixels in face images, it is desirable in terms of generalization performance that discriminant vectors φ's also have high spatial correlation between neighboring pixels, i.e., they are 2-D spatially smooth. Given unseen samples that are a bit spatially different from training samples, projections whose discriminant vectors are spatially smooth are more likely to produce small variation of features than others with spatially rough ones.

3. Nonuniform Spatially Regularized LDA

The proposed spatially regularized LDA takes advantage of nonuniform spatial common correlation structures of face images. Local regions in face images have different spatial correlation structures from each other. For example, regions corresponding to eyes, nose, and mouth have short spatial correlation, i.e., spatially rough, while other regions such as cheek and forehead have relatively long spatial correlation. The nonuniform spatial characteristic of face images is incorporated in spatially regularizing the discriminant vectors of LDA. The region-dependent spatial regularization is more advantageous for making a projection capture the spatial characteristic of face images than the uniform spatial regularization that relies on identical roughness constraints across local regions. The nonuniform smoothness penalty is described in detail in the following section.

3.1 Nonuniform smoothness constraint

The 2-D smoothness constraint adopted is based on the discrete Laplacian-based measure[14] that calculates local smoothness at each pixel. The Laplacian measure involves the derivatives in the horizontal and vertical directions only. For a discriminant vector φ of $n \times n\,(=d)$ size, the nonuniform smoothness penalty is defined as

$$\Theta(\varphi) = \sum_{i}^{n}\sum_{j}^{n} \gamma_{ij} \cdot \left(\varphi_{(i+1,j)} + \varphi_{(i-1,j)} + \varphi_{(i,j+1)} + \varphi_{(i,j-1)} - 4\varphi_{(i,j)}\right)^2 \quad (3)$$

where $\varphi_{(i,j)}$ represents a pixel of φ on (i,j) in the $n \times n$ matrix form and γ_{ij} is a regularization parameter on the pixel. The parameters γ_{ij} 's determine pixel-by-pixel nonuniform regularization strengths for the coefficients of φ. Eq. (3) is expressed in a matrix form for LDA framework as

$$\Theta(\varphi) = \left\|\Gamma^{1/2} D\varphi\right\|^2 = \varphi^T D^T \Gamma D \varphi, \quad (4)$$

where ϕ is a $d \times 1$ vector and D is a $d \times d$ matrix given by

$$\begin{bmatrix}
-2 & 1 & & & & & 1 & & & & & & & \\
1 & -3 & 1 & & & & & 1 & & & & & & \\
 & 1 & -3 & 1 & & & & & 1 & & & & & \\
 & & \cdot & \cdot & \cdot & & & & & \cdot & & & & \\
 & & & \cdot & \cdot & \cdot & & & & & \cdot & & & \\
1 & & & & 1 & -4 & 1 & & & & & 1 & & \\
 & 1 & & & & 1 & -4 & 1 & & & & & 1 & \\
 & & \cdot & & & & \cdot & \cdot & \cdot & & & & & \cdot \\
 & & & & & & & & & & & & &
\end{bmatrix}, \quad (5)$$

where the positive elements of each row correspond to 4 neighbors at each pixel (2 or 3 neighbors for pixels at boundaries), and I is a $d \times d$ diagonal matrix, $[\gamma_i > 0 \mid i = 1,...,d]$ that consists of the pixel-by-pixel spatial regularization parameters, which is the core of nonuniform spatially regularized LDA (NS-LDA). Varying the values of diagonal elements of I separately makes it possible to apply the strengths of spectral regularization differently from a pixel to another. The parameters can also control the scale of nonuniformity of spatial regularization, thereby ranging from pixel-by-pixel to region-by-region regularization. The nonuniform spatial regularization of Eq. (4) can be viewed as a generalized version of a spectrally regularized LDA in [11,12] in which an uniform smoothness constraint is adopted.

3.2 NS-LDA algorithm

There are several ways to combine two criteria of Eq. (1) and Eq. (4). Here, the nonuniform smoothness penalty, $\min \Theta(\varphi)$ is combined with the LDA framework, as in regularization methods of [11]. The modified Fisher criterion of NS-LDA is defined as

$$\max_{\varphi} \frac{\left|\varphi^T S_B \varphi\right|}{\left|\varphi^T (S_W + \alpha D^T \Gamma D)\varphi\right|}, \quad (6)$$

where α is a positive regularization factor that controls the overall strength of spatial regularization. The solution of Eq. (6) is obtained by solving the following generalized eigenvalue problem

$$S_B \phi = \lambda (S_W + \alpha D^T \Gamma D)\phi . \quad (7)$$

Since $D^T \Gamma D$ is symmetric and positive definite, the matrix $S_W + \alpha D^T \Gamma D$ is also symmetric and positive definite. Consequently, $S_W + \alpha D^T \Gamma D$ is invertible.

3.3 The nonuniform regularization parameters

This section explains how to determine the nonuniform spatial regularization matrix I, which is a key component of the NS-LDA algorithm. The underlying principle of the selection scheme is that, the larger spatial variation a local region or pixel has, the stronger spatial regularization applied is. The degree of spatial variation on a pixel is measured from an average of within-class variation of the training images. Consequently, NS-LDA leads to a regularized LDA that imposes strong smoothness regularization to local regions where deviation within face images of each person is relatively larger. The reason is as follows. The discriminant vectors that maximize Fisher criterion, that is, minimize $\left|\varphi^T S_W \varphi\right|$, often get overfitted to the "difference"' between face images of each class while transforming the training samples as close to each other in a low dimensional space as possible especially when the training samples are not enough. Discriminant vectors with higher smoothness on the local regions having large within-class variation in the training images are very likely to be advantageous than those with roughness on the local regions in terms of generalization performance.

Given the mean of within-class standard deviation on each pixel $\bar{s}_i$, the nonuniform regularization parameter for each pixel γ_i is determined by a linear scaling which is given by

$$\gamma_i = \frac{\gamma_{\max} - \gamma_{\min}}{\bar{s}_{\max} - \bar{s}_{\min}}(\bar{s} - \bar{s}_{\min}) + \gamma_{\min},\ i = 1,...,d, \quad (8)$$

where $\bar{s}_{\max} = \max \bar{s}_i$, $\bar{s}_{\min} = \min \bar{s}_i$, and $\gamma_{\max}$ and $\gamma_{\min}$ are control parameters that determine the maximum and minimum of the scaled $\bar{s}_i$. In the experiment, $\gamma_{\max}$ is set to 1 and $\gamma_{\min}\,(0 < \gamma_{\min} \le 1)$ is used as a single control parameter for controlling the non-flat regularization strength. When $\gamma_{\min}$ is 1, NS-LDA becomes a uniformly regularized LDA. The procedure of NS-LDA is summarized as follows.

1. Calculate the pixel-by-pixel standard deviation of face images of each person in the training data set.
2. Average the standard deviation over all classes.

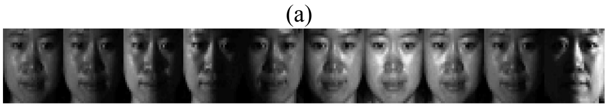
3. Determine Γ by using Eq. (8).
4. Extract discriminant vectors by solving Eq. (7).

4. Experimental Results and Discussions

4.1 Datasets and experiments

The face data sets are used to evaluate NS-LDA compared with other LDA methods. They are all standard publicly available databases. The ORL database[15] consists of 10 images of each of 40 individuals, each face captured in frontal position with slight tilt of the head. The CMU PIE database[16] consists of 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. Five near-frontal poses (C05, C07, C09, C27, C29) under all different illuminations and expressions are chosen for evaluation and 170 images for each individual are used. The CMU PIE dataset is adopted to assess NS-LDA with respect to illumination variation because spatial regularization may depend on illumination variation in face images. In all experiments, original images are normalized. The eyes are detected manually and aligned horizontally by rotation. Then the face area is cropped and re-scaled into a size of 32×32, with 256 gray levels per pixel. Each cropped face image is normalized into the unit length. Fig. 1 shows several sample images of the databases after normalization. For comparison purpose, LDA-based algorithms such as the Fisherface, R-LDA[7] and S-LDA[12] were implemented and compared with NS-LDA. For simplicity, a nearest neighbor classifier with Euclidean distance is used for classification.

(a)



(b)

Fig. 1 Samples of the normalized face images from the two databases: (a) ORL database, (b) CMU PIE database.

4.2 Experimental results

Tables 1-2 show the results of four LDA algorithms on the two databases. For each database, the algorithms are evaluated with respect to the number of training images per individual. "n training samples per individual" in each table means that n images per individual are chosen randomly as training samples and the others remained are used as test samples. For each case, error rates of 50 runs are averaged, each with a training data set chosen randomly. The experiments are run by varying values of the regularization parameters of each algorithm. Each error rate in the tables is the best result of each algorithm.

Table 1 reveals that NS-LDA outperforms Fisherface, R-LDA, and S-LDA on the ORL database. The improvement of NS-LDA over R-LDA and S-LDA was achieved considerably when the number of training images per individual is small, 2 to 3, which is the worst cases in terms of the SSS problem. This implies that NS-LDA is a little more effective in regularizing the poor estimates of S_W due to the SSS problem in face recognition than the other regularized LDA algorithms. As expected, the improvement diminishes as the number of training images per individual increases.

Table 1. Error rates(%) of LDA algorithms on the ORL database.

Methods	training samples per individual						
	2	3	4	5	6	7	8
Fisherface	42.4	21.4	11.5	7.4	5.3	4.0	3.6
R-LDA	20.5	10.8	6.3	3.6	2.6	2.0	1.3
S-LDA	17.0	8.1	4.2	2.3	1.6	1.2	0.8
NS-LDA	16.0	7.4	3.7	1.9	1.4	1.1	0.7

Table 2. Error rates(%) of LDA algorithms on the CMU PIE database.

Methods	training samples per individual						
	2	3	4	5	6	7	8
Fisherface	68.6	50.8	39.4	28.3	24.2	21.7	15.6
R-LDA	51.4	36.8	27.7	18.4	13.6	10.6	5.4
S-LDA	50.7	36.3	27.6	18.1	13.3	10.5	5.2
NS-LDA	49.4	35.0	26.6	17.4	12.9	10.1	5.1

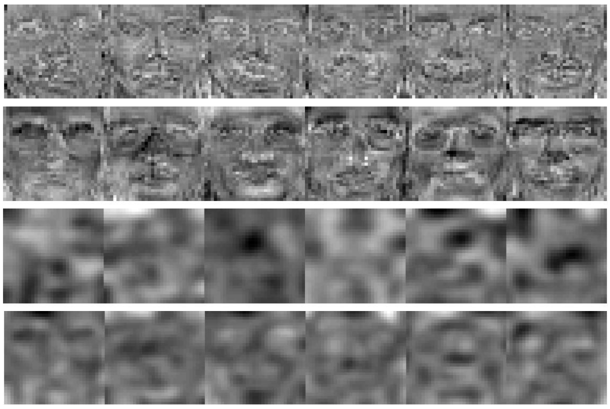


Fig. 2 Discriminant vectors of Fisherface, R-LDA, S-LDA, and NS-LDA (from top to bottom).

The error rates on the CMU PIE database are shown in table 2. High error rates of all LDA algorithms seem to be caused by large illumination variation, as shown in Fig. 1(b). Even though some illumination compensation such as Gabor filters can enhance recognition rate significantly, the improvement of LDA is considered only in this paper. As in the previous database, NS-LDA is also more effective than R-LDA and S-LDA for small sample-sized training data sets. It should be noted that the improvement of recognition rates by 1.0%~2.0% in case of a small number of training samples per individual corresponds to reduction of misclassification of about 110~220 face images. The results imply that the nonuniform spatial regularization is relatively effective in large variation in illumination in face images.

On the other hand, it is noteworthy to compare spatial smoothness of the discriminant vectors of the LDA algorithms. Fig. 2 shows top 6 discriminant vectors (from left to right) of each LDA algorithm from ORL database where training data set with 3 training images per individual was used to construct them. It can be observed that the discriminant vectors of Fisherface are spatially rough, R-LDA has smoother discriminant vectors than those of Fisherface, S-LDA has the smoothest ones, and NS-LDA lies between S-LDA and R-LDA in terms of the smoothness. The observation corresponds exactly to that the average roughness in terms of $\|D\phi\|^2$ of the six discriminant vectors of Fisherface, R-LDA, S-LDA, and NS-LDA are 9.51, 4.26, 0.06, and 0.17, respectively. It could be explained based on the performance results of table 1 that pepper-and-salt patterns on the discriminant vectors of Fisherface are indicative of the overfitting to local spatial features of the training face images, resulting in poor generalization performance. The discriminant vectors of R-LDA are locally smoother than those of Fisherface, which contributes to mitigation of the overfitting, accounting for the superiority of R-LDA over Fisherface. On the other hand, S-LDA and NS-LDA have by far smoother discriminant vectors than those of Fisherface and R-LDA because a spatial smoothness constraint is applied to the discriminant vectors directly. No pepper-and-salt patterns are found unlike in Fisherface and R-LDA, which means that differences between neighboring pixels or local regions are not over-stressed in extracting the discriminant vectors. Note that NS-LDA has smooth discriminant vectors that are more descriptive of the facial contours than those of S-LDA by virtue of the nonuniform regularization. For example, the local regions of eyes, nose, mouth that are crucial local features in face discrimination are discernible, unlike in S-LDA. The spatial characteristic of discriminant vectors of NS-LDA is believed to be highly related to outperformance of NS-LDA over S-LDA.

5. Conclusions

In this paper, a new spatial regularization on LDA has been proposed to reduce the overfitting in face recognition. The proposed regularized LDA, called Nonuniform Spatial Regularized LDA (NS-LDA) incorporates a two-dimensional spatial constraint that utilizes the nonuniformity of spatial correlation structures in face images into an LDA framework. An effective method of determining the nonuniform spatial regularization has also been presented. NS-LDA makes the discriminant vectors spatially smooth as well as capturing the facial contours, leading to better recognition rates especially when the number of training images per individual is quite small, which is not rare in practical face recognition applications. Experimental results on the two distinct face databases have shown that NS-LDA is able to achieve better performance than LDA algorithms.

References

[1] R. A. Fisher, "The use of multiple measures in taxonmic problems," *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.

[2] R. Chellapa, C. Whilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. of IEEE,* vol 83, no. 5, pp. 705-740, 1995.

[3] K. Fukunga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1991.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenface vs. fisherface: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18. no, 7, pp. 711-720, 1997.

[5] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition,* vol. 33, no. 10, pp. 1713-1726, 2000.

[6] K. Liu, Y. Cheng, J. Yang, H. Yu, and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition,* vol. 34, no. 10, pp. 2067-2070, 2001.

[7] J. H. Friedman, "Regularized discriminant analysis," *J. of the American Statistical Association,* vol. 84, no. 405, pp. 165-175, 1989.

[8] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulo, " Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letters,* vol. 26, no. 2, pp. 181-191, 2005.

[9] D.-Q. Dai and P. C. Yuen, "Face recognition by regularized discriminant analysis," *IEEE Trans. on SMC-PART B,* vol. 37, no. 4, pp. 1080-1085, 2007.

[10] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 383-394, 2008.

[11] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis", *The Annals of Statisticis*, vol. 23, no. 1, pp. 73-102, 1995.

[12] D. Cai, X. He, Y. Hu, J. Han, and T. Husang, "Learning a spatially smooth subspace for face recognition", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-7, 2007.

[13] J. Ye, R. Janardan, and Q. Li, "Two-dimensional Linear Discriminant Analysis", *Proc. of Neural Information Processing Systems*, pp. 1569-1576, 2004.

[14] F. O'Sullivan, "Discretized Laplacian smoothing by Fourier methods," *J. of the American Statistical Association,* vol. 86, pp. 634-642, 1991.

[15] ORL Face Database, AT&T Laboratories Cambridge, 2005.

[16] T. Sim, S. Baker, and M. Bsat, "The CMU pose," *Proc. of Conf. on Automatic Face and Gesture Recognition*, pp. 46-51, 2002.

Lae-Jeong Park received the B.S. degree in Electrical Engineering from Seoul National University, Seoul, Korea, in 1991, and the M.S. and Ph.D. degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology, Taejon, Korea, in 1993 and 1997, respectively. From 1997 to 1999, he was with the Information Technology Lab. at LG Corporate Institute of Technology, Seoul, Korea. He is currently an associate professor in the Department of Electronics Engineering at Gangneung-Wonju National University, Gangneung, Korea. His current research interests are machine learning, intelligent mobile apps, and wireless remote monitoring system.

Phone : +82-33-640-2389
Fax : +82-33-646-0740
E-mail : ljpark@gwnu.ac.kr