Elongated Radial Basis Function for Nonlinear Representation of Face Data

Sang-Ki Kim*, Sunjin Yu**, Sangyoun Lee* ***Regular Members***

ABSTRACT

Recently, subspace analysis has raised its performance to a higher level through the adoption of kernel-based nonlinearity. Especially, the radial basis function, based on its nonparametric nature, has shown promising results in face recognition. However, due to the endemic small sample size problem of face data, the conventional kernel-based feature extraction methods have difficulty in data representation. In this paper, we introduce a novel variant of the RBF kernel to alleviate this problem. By adopting the concept of the nearest feature line classifier, we show both effectiveness and generalizability of the proposed method, particularly regarding the small sample size issue.

Key Words : face recognition, subspace learning, kernel feature extraction, RBF kernel function, nearest feature line

I. Introduction

As the security has risen on public concerns, face recognition has received increasing attention over the last decade[1]. This challenging research topic has attracted many researchers for its wide range of potential applications[2]. However, there still remain a lot of issues to be solved for recognizing faces from images in practical use. One of the issues is extraction of proper features.

Among various approaches so far published and suggested, the subspace-based methods, such as principal component analysis (PCA)[3], linear discriminant analysis (LDA)[4], and independent component analysis (ICA)[5], showed promising results. These linear feature extraction methods are further improved via nonlinear generalization through the kernel functions[6], yielding their kernel-based counter parts[7-9]. Especially, owing to its nonparametric nature, radial basis function (RBF) achieves remarkable performance in terms of recognition accuracy. There is conceptual resemblance between RBF and k-nearest neighbor classifier (kNN) where it has been proven that RBF performs the nearest-neighbor mapping when its Gaussian kernel's radius approaches zero[10]. Since face data has heteroscedastic (nonidentical within-class scatter-ness) distribution, this nonparametric nature enables the RBF-based methods to be among the best performers[11].

However, in real application, RBF-based methods suffer from the curse of dimensionality and the small sample size problem[12]. These problems arise whenever the number of samples is not large enough compared to the dimensionality of the samples, which is very common in face recognition. Also, because the nonparametric approaches are highly unstructured, they don't give any understandings of the data[11]. In such situation, it is even harder to represent the distribution of each class/identity of data with RBF's identical and symmetric kernel functions, and it eventually yields unreliable responses[13].

In a similar context, small number of samples

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2011-0004718)
* 연세대학교 전기전자공학과 영상인식 연구실 ({neobox, syleee}@yonsei.ac.kr),
** LG전자 전자기술원/미래IT융합 연구소 (sunjin.yu@lge.com)
논문번호 : KICS2011-06-236, 접수일자 : 2011년 6월 3일, 최종논문접수일자 : 2011년 6월 22일

also makes *k*NN fail to sufficiently represent complex variations of face data[14]. To extend the capacity of data representation, nearest feature line (NFL) was proposed[15]. Here, it is argued that the feature line connecting two feature points of a class gives better generalization than kNN having those points as its prototypes. Recently, NFL is successfully applied to feature extraction in [16, 17] proving its generalization capacity.

Noting the similarities between the immanent nature of RBF and that of *k*NN, we propose a novel kernel function, elongated RBF (eRBF), inspired by the success of NFL. In the proposed eRBF, a Gaussian function stretched along an NFL connecting neighboring sample pair is used as a basis function. Owing to nonidentical kernel functions which capture potential data variations and better preserve local structures, we can achieve more generalizable data representation and higher recognition accuracy.

The organization of this paper is as follows. In section II, we introduce the proposed eRBF in detail. The comparative experiments applying eRBF and other types of kernels to three different facial feature extraction methods are presented in section III. Finally the paper is concluded with discussions in section IV.

Ⅱ. Proposed methodology

2.1 Preliminaries of RBF-based feature extraction

A nonlinear feature based on RBF can be extracted as follows:

$$y = \sum_{i=1}^{M} \omega_i \phi_i(\mathbf{x}), \qquad (1)$$

where

$$\phi_i(\mathbf{x}) = \exp\left(-\parallel \mathbf{x} - \mathbf{x}_i \parallel^2 / \sigma^2\right) \qquad (2)$$

is a RBF kernel function and ω_i is a corresponding weight for extracting feature y. In the conventional kernel trick-based implementations [7-9], the entire training samples are utilized as basis functions which also can be interpreted as a RBF network taking every training sample $\mathbf{x}_i$ as the center of each kernel function with common radial variance σ^2. So, finding suitable features is done only by adjusting the weight ω_i which can be seen as output layer weight of RBF network.

On the other hand, the construction of RBF network generally contains the procedure of finding a proper set of basis functions. This procedure consists of setting the centers of RBFs and finding individually tailored variance size σ for each kernel function via various clustering methods such as *k*-means algorithm[13].

However, directly applying such strategy to facial feature extraction is also problematic because of the small number of samples. When the number of samples per a class is small, approximating their variation by clustering does not help the trained kernels either in better representing the data distribution or in extracting facial features generalizable to unseen data.

Considering these remarks, we pursue following two aims in designing the proposed new kernel function:

To define a new kernel function which captures potential data variation for better generalizability.

To define centroid, orientation, and variance of kernel function so that it is adaptable to heteroscedastic data

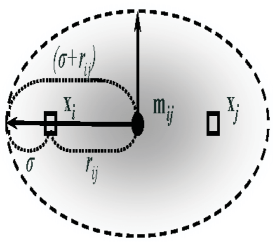


Fig. 1. Ellipsoidal RBF kernel function based on a sample pair, $\mathbf{x}_i$ and $\mathbf{x}_j$. $\mathbf{m}_{ij}$ and r_{ij} are the mean and the standard deviation of the sample pair, and σ is the radius parameter determining overall spread

2.2 Definition of elongated RBF based on a neighboring sample pair

Preceding researches have shown that, a feature line (FL) connecting two samples from a class gives better representativeness than utilizing the samples as individual prototypes in *k*NN[15-17]. This implies that, for a given FL and its corresponding class, data variation along the FL is more likely than other directions. So we have designed the eRBF as shown in Fig. 1, allowing more variance of $(\sigma+r_{ij})^2$ along the FL, and less for other directions with amount of σ^2. Here, x_i and x_j $(\in \mathrm{R}^N)$ are neighboring samples from the same class, m_{ij} and r_{ij} respectively are the mean and the standard deviation of the sample pair, and σ is the radius parameter which determines the overall spread of the basis function. This can be expressed by following equation.

$$\phi_{ij}(\mathrm{x}) = \exp\left(-(\mathrm{x}-\mathrm{m}_{ij})^T \mathrm{U}_{ij} \Lambda_{ij}^{-1} \mathrm{U}_{ij}^T (\mathrm{x}-\mathrm{m}_{ij})\right), \quad (3)$$

where $\mathrm{U}_{ij} = [\mathrm{u}_1^{ij}, \mathrm{u}_2^{ij}, ..., \mathrm{u}_N^{ij}]$ is a N-by-N transform matrix consisting of orthonormal vectors such that only the first vector is specified as

$$\mathrm{u}_1^{ij} = (\mathrm{x}_i - \mathrm{x}_j) / \| \mathrm{x}_i - \mathrm{x}_j \|, \quad (4)$$

and

$$\Lambda_{ij} = diag\left[(\sigma+r_{ij})^2, \sigma^2, ..., \sigma^2\right] \quad (5)$$

is a N-by-N diagonal matrix.

Noting that the vectors $(\mathrm{u}_l^{ij})_{l=2,...,N}$ span the null-space of u_1^{ij}, the computation of Eq. 3 can be simplified as:

$$\phi_{ij}(\mathrm{x}) = \exp\left(-\| \Delta_{ij}^p \|^2 / (\sigma+r_{ij})^2 - \| \Delta_{ij}^o \|^2 / \sigma^2\right), \quad (6)$$

where Δ_{ij}^p and Δ_{ij}^o are the parallel and the orthogonal components of $(\mathrm{x}-\mathrm{m}_{ij})$ to u_1^{ij} respectively. Their L2-norms can be calculated as:

$$\| \Delta_{ij}^p \|^2 = \| \mathrm{u}_1^{ij} (\mathrm{u}_1^{ij})^T (\mathrm{x}-\mathrm{m}_{ij}) \|^2 = \| (\mathrm{u}_1^{ij})^T (\mathrm{x}-\mathrm{m}_{ij}) \|^2 \quad (7)$$

and

$$\| \Delta_{ij}^o \|^2 = \| (\mathrm{x}-\mathrm{m}_{ij}) \|^2 - \| \Delta_{ij}^p \|^2. \quad (8)$$

NFL implementation proposed in [15] constructs the FLs using all possible within-class sample pairs. However, such approach is unsuitable for constructing eRBF in two aspects. Firstly, FLs connecting distant samples may cause undesirable side effects. As two samples are located far from each other, they are more likely to differ in various external conditions, such as pose, illumination, etc., and their effects are so much complicated so that it is hard to be detected linearly[18]. Fig. 2 shows an exemplary case of face data. Here, the dashed-line illustrates a FL, $\overleftrightarrow{\mathrm{x}_1\mathrm{x}_2}$, connecting farthest within-class sample pair, x_1 and x_2. In the context of kNN and NFL, this FL substantially reduces classification margin which means simply interpolated and extrapolated points of distant sample pair might carry altered identity information. Secondly, the number of FLs grows rapidly as the number of available samples increases. Since the number of kernels determines the dimensionality of extracted features, pairing all possible samples leads to an explosion of dimensionality.

In order to prevent these issues, we need a criterion to determine which sample pairs should be used for eRBF. In this paper, we propose to pair nearest neighboring samples of which the proximity is measured by Mahalanobis distance. The strategy

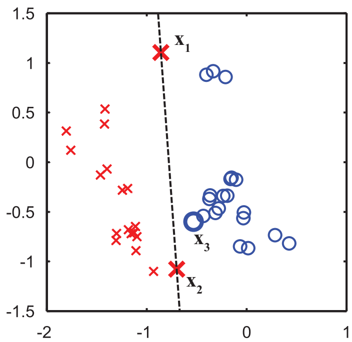


Fig. 2. Distribution of face data from two different individuals marked as 'x' and 'o'. x_1 and x_2 belong to class 'x' and x_3 to class 'o'.

of connecting nearby samples is expected to avoid the construction of undesirable basis functions, and utilizing Mahalanobis distance enables us to find the neighborhoods substantially in the PCA-whitened space. In this way, more meaningful local structures could be captured. Noting all these remarks, we define a set of within-class neighbors as follows:

Definition (Set of within-class neighbors). Given a pair of data indices (i,j), it is the member of the set S^k iff x_j is one of the k nearest within-class neighbors of x_i in terms of Mahalanobis distance or vice versa.

Using these selected data pairs, one can have eRBF-based feature of arbitrary input x as follows:

$$y = \sum_{(i,j)\in S^k} \omega_{ij}\phi_{ij}(\mathrm{x}). \tag{9}$$

Although this equation is quite similar to that of RBF shown in Eq. 1, eRBF is inapplicable directly to conventional kernel trick-based methodologies. This is because eRBF is a nonsymmetric function, and thus does not satisfy Mercer's condition[6].

So, in order to find the feature weights ω_{ij}, we need to calculate eRBF outputs of training data firstly as follows:

$$\Phi_{tr} = [\phi(\mathrm{x}_l)]_{l=1,\ldots,M}, \text{ where} \\ \phi(\mathrm{x}_l) = [\phi_{ij}(\mathrm{x}_l)]^T_{(i,j)\in S^k}. \tag{10}$$

Then, ω_{ij} can be trained by linear feature extraction methods such as PCA and LDA taking Φ_{tr} as input. The procedure of implementing feature extraction based on the proposed eRBF is summarized as follows:

Train PCA to calculate the Mahalanobis distance.

Find the set S^k based on its definition.

Construct eRBFs by calculating u_1^{ij}, m_{ij} and r_{ij}.

Calculate eRBF outputs of the training data, Φ_{tr}, as Eq. 10

Find feature extraction weights, ω_{ij}, by desired feature extraction methods using Φ_{tr} as input data.

Extract features from an arbitrary input x by Eq. 9.

Ⅲ. Experiments

In order to evaluate the proposed eRBF, two publicly available databases, FERET[19] and AR[20], are adopted in this study. All images are normalized to 56 x 46 pixels according to the manually marked eye centers. Then, they are masked, histogram-equalized, and scaled to have zero mean and unit variance pixels. Samples of the images are shown in Fig. 3.

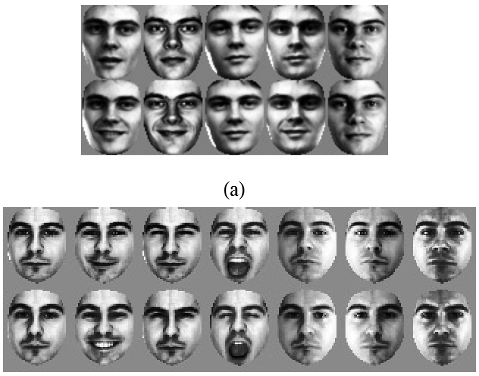
We consider only the verification scenario where the results are reported in terms of the equal error rates (EERs) averaged from the 4-fold cross validation. For the cross validation, the whole data set is divided into 4 subsets so that there are no common individuals, and, for each iteration, 3 subsets are used for training and the remaining one for testing.

The proposed eRBF is tested in comparison with two other kernel functions, RBF and Polynomial, of which the formulations are shown in Eq. 11 and 12, respectively:

$$f(\mathrm{x}_i,\mathrm{x}_j) = \exp\left(-\parallel \mathrm{x}_i - \mathrm{x}_j \parallel^2/\sigma^2\right) \tag{11}$$

and

$$f(\mathrm{x}_i,\mathrm{x}_j) = \left(a(\mathrm{x}_i - \mathrm{x}_j) + b\right)^D. \tag{12}$$



(a)

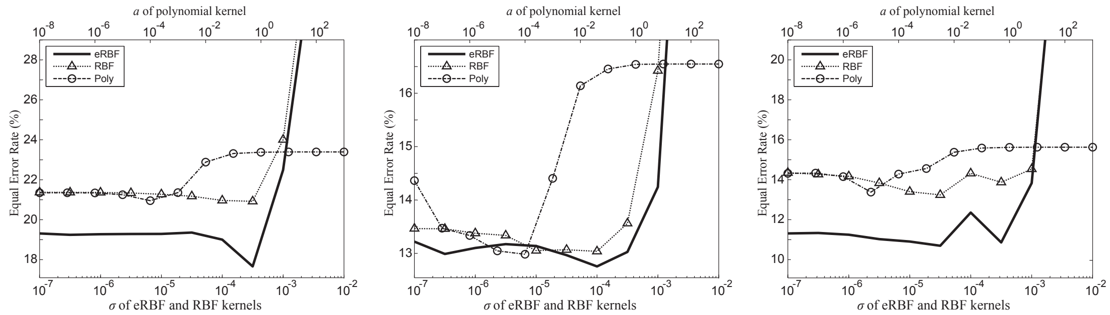
(b)

Fig. 3. Samples of experimental data of (a) FERET and (b) AR databases

These kernels are applied to three feature extraction methods, KPCA, KFD, and KICA. Although eRBF is not implemented via kernel-trick, the main characteristics of the feature extraction methods, such as eigenvalue regularization for whitening the within-class scatter in KFD[29] and the structure of KPCA+ICA in KICA[30], are preserved.

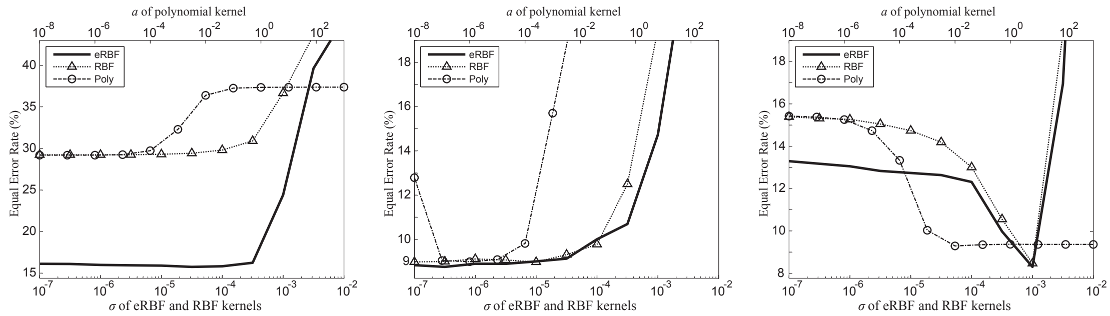
We performed the experiments varying the parameters of the kernels. In RBF, we varied σ logarithmically ranging from 10^{-7} to 10^{-2}, while, in eRBF, we also varied the number of within-class neighbors, k, and the best performance is reported for each setting of σ. In polynomial, for simplicity, we fixed the bias parameter, b, and the degree parameter, D, to one and three respectively, and varied only the weight parameter, a, logarithmically ranging from 10^{-8} to 10^{3}.

Experimental results on FERET and AR varying the kernel parameters are plotted in Fig. 4 and 5, respectively. The best results and corresponding parameter settings are also summarized in Table 1. We can see that the trajectories of EER graphs of RBF and eRBF show similar tendencies as the kernel parameter σ varies. However, eRBF consistently outperforms RBF in all parameter settings. Especially, KPCA gains remarkable performance improvement with eRBF. This is because the eRBF has discriminative power which is obtained by reducing the variance between the connected within-class samples in its response. Another noteworthy observation is the performance gain in KFD. Since KFD is based on the Fisher's discriminant analysis, the discriminating nature of eRBF is not enough to explain the performance gain. This performance improvement can be attributed to the preserved local structure. This is especially true when there are fewer within-class samples, noting that the average number of samples per class are respectively 6.6 and 14 in the FERET and the AR experiments. Since KFD assumes normal and identical within-class scatter, it cannot fully explore the discriminancy of the data only with



(a) (b) (c)

Fig. 4. Resulting EERs of FERET experiments on (a) KPCA, (b) KICA, and (c) KFD



(a) (b) (c)

Fig. 5. Resulting EERs of FERET experiments on (a) KPCA, (b) KICA, and (c) KFD

Table 1. Best EERs and parameter settings thereof

	FERET			AR		
	RBF	Polynomial	eRBF	RBF	Polynomial	eRBF
KPCA	20.9224 % (σ =1e-3.5)	20.9787 % (a = 1e-4)	17.6651 % (σ =1e-3.5, k = 1)	29.2301 % (σ =1e-7)	29.2288 % (a = 1e-8)	15.7538 % (σ =1e-4.5, k = 1)
KICA	13.0377 % (σ =1e-4)	12.9965 % (a=1e-4)	12.7561 % (σ =1e-3.5, k = 2)	8.9715 % (σ =1e-7)	9.0182 % (a = 1e-6)	8.7535 % (σ =1e-6.5, k = 2)
KFD	13.2407 % (σ =1e-4.5)	13.4041 % (a=1e-5)	10.6937 % (σ =1e-3.5, k = 3)	8.4663 % (σ =1e-3)	9.3197 % (a = 1e-2)	8.3021 % (σ =1e-3, k = 2)

RBF. Meanwhile, by capturing the possible local variances based on the concept of NFL, eRBF performs better in classifying heteroscedastic face data.

Ⅳ. Conclusion

In this paper, we presented a novel kernel function, eRBF, for application in facial feature extraction. The proposed eRBF has several advantages over the conventional RBF:

It tries to capture possible variations based on the concept of NFL cultivating the discriminability.

Non-identically designed kernel functions better preserve detailed local structure of data allowing more adaptable feature extraction from heteroscedastic data.

The empirical experiments showed our method outperforms other kernel functions evidencing its claimed effectiveness and generalizability in small sample size situations.

References

〔1〕 W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, Vol.35, No.4, 2003, pp.399-458.

〔2〕 L. Torres, "Is there any hope for face recognition?," *Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisbon, Portugal, Apr. 2004, pp.21-23.

〔3〕 M. Turk and A. Pentland, "Eigenfaces for Recognition," *Cognitive Neuroscience*, Vol.3, No.1, pp.72-86, 1991.

〔4〕 P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* Vol.19, No.7, May 1997, pp.711-720.

〔5〕 M. S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transaction on Neural Networks*, Vol.13, No.6, Nov. 2002, pp.1450-1464.

〔6〕 C.J.C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, Vol.2, No.2, 1998, pp. 121-167.

〔7〕 K.I. Kim, K. Jung, and H.J. Kim, "Face Recognition using Kernel Principal Component Analysis," *IEEE Signal Processing Letters*, Vol.9, No.2, Feb. 2002, pp.40-42.

〔8〕 J. Ma, J.L. Sancho-Gómez, and S.C. Ahalt, "Nonlinear Multiclass Discriminant Analysis," *IEEE Signal Processing Letters*, Vol.10, No.7, Jul. 2003, pp.196-199.

〔9〕 J. Yang, X. Gao, D. Zhang, and J.Y. Yang, "Kernel ICA: An alternative formulation and its application to face recognition," *Pattern Recognition*, Vol.38, No.10, Oct. 2005, pp.1784-1787.

〔10〕 J. Moody and C. Darken, "Learning with Localized Receptive Fields," *Proc. 1988 Connectionist Models Summer Schoo*l, Morgan Kaufmann, San Mateo, CA, 1989, pp.133-143.

〔11〕 T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference and Prediction.* New York: Springer, 2001.

〔12〕 A.K. Jain, R.P.W. Duin, J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.22, No.1, Jan. 2000, pp.4-37.

〔13〕 K. M. Tao, "A closer look at the radial basis function (RBF) networks," *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1993.

〔14〕 Y. Hamamoto, S.Uchimura, S. Tomita, "On the behavior of artificial neural network classifiers in high-dimensional spaces," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.18, No.5, May 1996, pp.571-574.

〔15〕 S. Z. Li, and J. Lu, "Face recognition using the nearest feature line method," *IEEE Transaction on Neural Networks*, Vol.10, No.2, 1999, pp.439-443.

〔16〕 Y. Pang, Y. Yuan, and X. Li, "Generalized nearest feature line for subspace learning," *Electron. Letters*, Vol.43, No.20, 2007, pp.1079-1080.

〔17〕 J. Lu, Yap-Peng Tan, "Uncorrelated Discriminant Nearest Feature Line Analysis for Face Recognition," *IEEE Signal Processing Letters*, Vol.17, No.2, Feb. 2010, pp.185-188

〔18〕 S.-K. Kim, Y.J. Park, K.-A. Toh, and S. Lee, "SVM-based feature extraction for face recognition," *Pattern Recognition*, Vol.43, No.8, Aug. 2010, pp.2871-2881

〔19〕 P.J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.22, No.10, Oct. 2000, pp.1090-1104.

〔20〕 A.R. Martinez and R. Benavente, "The AR face database," *Technical Report #24, Computer Vision Center (CVC)*, Jun. 1998.

김 상 기 (Sang-Ki Kim) 정회원

2004년 8월 연세대학교 전기전자공학과 학사

2004년 8월~현재 연세대학교 전기전자공학과 박사과정

<관심분야> 얼굴인식, 생체인식, pattern recognition, machine learning, computer vision

유 선 진 (Sunjin Yu) 정회원

2003년 8월 고려대학교 전자 및 정보공학과 학사

2006년 2월 연세대학교 생체인식협동과정 석사

2011년 2월 연세대학교 전기전자공학과 박사

2011년 2월~현재 LG전자 선임연구원

<관심분야> 3D vision, HCI, 얼굴 인식

이 상 윤 (Sangyoun Lee) 정회원

1987년 2월 연세대학교 전자공학과 학사

1989년 2월 연세대학교 전자공학과 석사

1999년 2월 Georgia Tech. 전기 및 컴퓨터공학과 박사

1989년~2004년 KT 선임연구원

2004년~현재 연세대학교 전기전자공학부 부교수

<관심분야> 생체인식, 컴퓨터비전, 영상부호화