A New Similarity Measure Based on Intraclass Statistics for Biometric Systems

Kwanyong Lee and Hyeyoung Park

A biometric system determines the identity of a person by measuring physical features that can distinguish that person from others. Since biometric features have many variations and can be easily corrupted by noises and deformations, it is necessary to apply machine learning techniques to treat the data. When applying the conventional machine learning methods in designing a specific biometric system, however, one first runs into the difficulty of collecting sufficient data for each person to be registered to the system. In addition, there can be an almost infinite number of variations of non-registered data. Therefore, it is difficult to analyze and predict the distributional properties of real data that are essential for the system to deal with in practical applications. These difficulties require a new framework of identification and verification that is appropriate and efficient for the specific situations of biometric systems. As a preliminary solution, this paper proposes a simple but theoretically well-defined method based on a statistical test theory. Our computational experiments on real-world data show that the proposed method has potential for coping with the actual difficulties in biometrics.

Keywords: Biometrics, statistical verification, Chi-square distribution, statistical test theory, similarity measure.

I. Introduction

To control the access to secure areas or transact electronically through the internet, a reliable personal identification infrastructure is required. Conventional methods of recognizing the identity of a person by using a password or cards are not altogether reliable. Biometrics refers to automatic identification of a person based on his/her physiological or behavioral characteristics. Biometrics measurements, such as fingerprint, face, or iris patterns are common and reliable ways to achieve verification of an individual's identity with a high level of accuracy. It provides a better way for the increased security requirements of our information society than traditional identification methods such as passwords or ID cards.

In order to make reliable biometric systems, various sophisticated techniques, which involve machine learning, artificial intelligence, or signal processing, have been widely used [1]-[9]. Most studies, however, have tried to optimize the systems to the characteristics of given specific biometric data, and thus the researchers have focused on those problem-dependent aspects, such as the sophistication of feature extraction, for given raw data.

In this paper, we concentrate on the common, core aspect of general biometric systems that is relevant to efficiently selecting the similarity measure and verification threshold. We assume that a set of extracted feature values from raw data is already given for further processing. Moreover, we assume that the feature values are continuous, even though the overall mechanism of the proposed method can be easily extended to binary data.

When new data is given to a trained system, it is necessary to measure the similarity between the new data and the registered

Manuscript received June 4, 2002; revised June 9, 2003.

This work was supported by Korea Research Foundation Grant (KRF-2001-003-E00234).

Kwanyong Lee (phone: +82 2 2102 2222, email: kylee@mail.kcu.ac) is with Department of Computer, Information, and Communication, Korea Cyber University, Seoul, Korea.

Hyeyoung Park (email: hypark@brain.riken.go.jp) is with Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Saitama, Japan.

data. The most basic similarity measure is the Euclidean distance and inner product, but this is too simple for considering the characteristics of data distribution. To improve these methods, Kee et al. tried the normalized Euclidean and simplified Mahalanobis distance [6]. However, in a biometric system, the number of items of data in a class for each person is very small, so it is difficult to get accurate information for the distribution of each class. This problem can cause serious inaccuracy of the distance measure.

To overcome this problem, other researchers have proposed more sophisticated classification methods, such as neural networks and kernel machines [4], [7]. Even in this kind of machine learning system, however, we need a sufficient number of items of data to achieve good performance. As mentioned above, in the field of biometrics, collecting sufficient data is difficult. Furthermore, the main difference between biometric classification problems and general classification problems is that the possibility of the appearance of new data that does not belong to any trained class is very high. (In fact, one of the main purposes of biometric systems is to reject such data.) Since these data have almost infinite variations, the usual classification methods can hardly guarantee robustness in handling them. It is necessary to reconsider the characteristics of the biometric problems and develop more appropriate methods for identification and verification for biometrics. From this point of view, we propose a strategy to extract more robust and essential information of data distributions in biometric problems and apply it to developing a similarity measure. The information obtained by the proposed method does not depend on the distribution of each class for each person, but depends on all the data sets. Accordingly, we can expect to get a more reliable similarity measure.

In addition to the similarity measure, the threshold for determining acceptance is also important for the system performance. The basic method is to find the boundary where the summation of the false rejection rate (FRR) and false acceptance rate (FAR) is minimized using some training data. In this case, the threshold absolutely depends on the distribution of the training data set, which is usually not big enough. Consequently, there is no guarantee to get a good FAR and FRR for new input with noises and deformations. We need to consider the characteristics of the underlying probability distribution of the data.

Daugman's method [5], which exploited the Neyman-Pearson method [10], investigated the data distributions of the imposter class and the authentic class in order to find the threshold for verification. The proposed method takes the same approach, except that the proposed method deals with real (continuous) values whereas Daugman's method handled binary data. In addition, the current paper proposes a method for choosing the threshold based on the assumed data distributions, whereas Daugman just empirically decided an explicit value for the threshold using a given data set. In these senses, the current paper can be considered as a generalization or extension of Daugman's work.

II. Similarity Measure

Let us first describe the conventional approaches for identification from the statistical viewpoint. To make a similarity measure based on the statistics of data, let us represent the data as a random variable $\boldsymbol{x}$=(x_1,…,x_D) with dimension D. The whole data set X={$\boldsymbol{x}_n$|n=1,…,N} can be decomposed into subsets $X_k = \{\boldsymbol{x}_{n_k} | n_k = 1,\ldots, N_k\}$ (k=1,…,K), where each subset X_k consists of data from the class C_k corresponding to a person k. For identification, conventional methods consider the statistical properties of data $\boldsymbol{x}_{n_k}$ (n_k=1,…,N_k) in class C_k. It can be represented by a probability density function $p_k(\boldsymbol{x})$. If we have $p_k(\boldsymbol{x})$ for each k, then the identification process can be done based on the probability; for given data $\boldsymbol{x}$, we calculate $p_k(\boldsymbol{x})$ (or $f(p_k(\boldsymbol{x}))$), where f is a monotonic function and find a class C_k maximizing $p_k(\boldsymbol{x})$. Therefore, the main issue of identification is to find a good estimate of $p_k(\boldsymbol{x})$. The conventional methods can be considered as finding the estimate of $p_k(\boldsymbol{x})$. For example, for K-means clustering, we assume that the probability density $p_k(\boldsymbol{x})$ is defined as a Gaussian distribution with a mean μ_k and an identity covariance matrix, and then we estimate μ_k using data set X_k for each class. Then the similarity measure between a new data item and the center μ_k of class C_k is given by

$$-\log p_k(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \mu_k\|^2. \tag{1}$$

Note that this is the Euclidean distance. If we also estimate the covariance matrix $\sum_\mathrm{k}$ for $p_k(\boldsymbol{x})$, then the similarity measure defined as –log $p_k(\boldsymbol{x})$ is the Mahalanobis distance. In addition, if we assume the covariance matrix is a diagonal matrix, then we get the simplified Mahalanobis distance. For more sophisticated non-linear classifiers, such as neural networks, the $p_k(\boldsymbol{x})$ (or its monotonic mapping $f(p_k(\boldsymbol{x}))$) is estimated by a complicated nonlinear function through learning.

The main problem of these conventional approaches is that we need a sufficient number of data $\boldsymbol{x}$ for each class in order to a get good estimate of $p_k(\boldsymbol{x})$. However, in biometrics, it is costly to get a large enough number of items of data for each person to give meaningful statistics. This problem can cause significant deterioration of system performance.

To solve such a problem, we introduce a new random value

$\boldsymbol{y}$, which can be defined by using a pair of data ($\boldsymbol{x}, \boldsymbol{x}'$) from the same person:

$$\boldsymbol{y} = h(\boldsymbol{x}, \boldsymbol{x}'). \tag{2}$$

We then try to estimate $p(\boldsymbol{y})$ instead of $p_k(\boldsymbol{x})$ (k=1,...,K) and use it to define the similarity measure. An important merit of this approach of not considering $p(\boldsymbol{x})$ is that $\boldsymbol{y}$ does not depend on the distribution of class C_k. This is justified by the fact that the stochastic uncertainty existing in data does not originate from the bio-signal itself, but comes from the measuring process, such as the properties of the equipment, the measuring environment, and so forth. Therefore, we can represent the stochastic uncertainty using some function of the pair of data items $\boldsymbol{x}$ and $\boldsymbol{x}'$, which we define using $\boldsymbol{y} = h(\boldsymbol{x}, \boldsymbol{x}')$. This is the basic concept of the proposed method based on intraclass information.

We give a primary application of the proposed concept in this paper. However, it is also possible to define the shape of h and a model of $p(\boldsymbol{y})$ in various forms according to the properties of the equipment and the characteristics of a given bio-signal. Let us define

$$\boldsymbol{y} = h(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x} - \boldsymbol{x}', \tag{3}$$

and the probability distribution of $\boldsymbol{y}$ is a multivariate Gaussian distribution. This is given under the assumption that the difference between each pair of samples from the same individual, $\boldsymbol{x}$–$\boldsymbol{x}'$, originates from some additive Gaussian noises. Even though this assumption is somewhat ideal, it can still be applied to real data if the data item $\boldsymbol{x}$ is well pre-processed, as we show later. The basic assumption used in this paper shows how the concept of intraclass statistics can be used for biometrics.

Here, since we assume $p(\boldsymbol{y})$ is Gaussian, we need to estimate its mean μ and covariance $\sum$ in order to know $p(\boldsymbol{y})$. In this paper, for simplicity, we assume that each feature element x_d, d=1,…,D is independent of each other, and thus we need to estimate only its diagonal elements, σ_d^2, d=1,…,D.

Let us first construct the set Y of $\boldsymbol{y}$ from the original data set X. It can be simply given by the following two steps.

Step 1. For each subset X_k, for all possible combinations of two different data items $\boldsymbol{x}$ and $\boldsymbol{x}'$, calculate $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}'$, and put it into the set Y.

Step 2. Repeat step 1 for all subsets X_k, k=1,…,K.

Using this set, we can estimate the statistics of $\boldsymbol{y} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_D)$. The standard deviation σ of $\boldsymbol{y}$ can be estimated by

$$\sigma_d = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y_d^m - \mu_d)^2}, \tag{4}$$

where $\mu = (\mu_1,\ldots,\mu_D)$ is the sample mean given by

$$\mu_d = \frac{1}{M}\sum_{m=1}^{M} y_d^m, \tag{5}$$

and y_d^m is the d-th element of m-th data y^m in Y.

Using these statistics, we can define a similarity measure $s(\boldsymbol{x}, \boldsymbol{x}')$ as

$$s(x, x') = \sum_{d=1}^{D}\frac{(x_d - x'_d - \mu_d)^2}{\sigma_d^2}. \tag{6}$$

For a new item of data $\boldsymbol{x}^{new}$, the similarity between the new one and a registered item of data $\boldsymbol{x}^{reg}$ is measured by the function $s(\boldsymbol{x}^{reg}, \boldsymbol{x}^{new})$.

Note that the statistics $\mu = (\mu_1,\ldots,\mu_D)$ and $\sigma = (\sigma_1,\ldots,\sigma_D)$, which are obtained from the data set Y and used for the similarity measure, are different from those of the conventional distances, such as the normalized Euclidean and the simplified Mahalanobis distance. Since the number of items of data in Y is much more than that of each subset X_k, the obtained estimates of the mean and variance are more accurate and more robust against noises.

III. Verification Threshold

To determine a specific threshold for verifying new input data, we need to consider the distribution of the similarity values. Let us consider (6). If the two items of data $\boldsymbol{x}$ and $\boldsymbol{x}'$ are from a same subset X_k, then each factor $(x_d - x'_d - \mu_d)/\ \sigma_d$ is subject to the standard normal distribution $N(0,1)$ from our assumption on $p(\boldsymbol{y})$. Therefore, the similarity measure $s(\boldsymbol{x}, \boldsymbol{x}')$ can be considered as a random variable that is subject to the χ^2 distribution with D degree of freedom, $p_{\chi^2}(s;D)$, where D is the dimension of $\boldsymbol{x}$.

From this, we can easily apply the likelihood ratio test to the verification process. The overall process of the likelihood test can be summarized by the following steps:

Step 1. For a new item of data $\boldsymbol{x}^{new}$, calculate $s(\boldsymbol{x}^{new}, \boldsymbol{x}_n)$ for all data $\boldsymbol{x}_n$ in X and find the minimum value $\boldsymbol{s}^{min}$ and the most similar data $\boldsymbol{x}^{min}$ that can be written as

$$\boldsymbol{x}^{min} = \arg\min\{s(\boldsymbol{x}^{new},\ \boldsymbol{x}_n)|\ \ x_n \in X\}, \tag{7}$$

$$s^{min} = s(\boldsymbol{x}^{new}, \boldsymbol{x}^{min}). \tag{8}$$

Step 2. Under the assumption that the two data $\boldsymbol{x}^{new}$ and $\boldsymbol{x}^{min}$ are given from a same subset, calculate probability $P(s > s^{min})$ using χ^2 distribution.

Step 3. If probability $P(s > s^{min})$ is smaller than a pre-defined small probability, α, then reject the data. Otherwise, accept it.

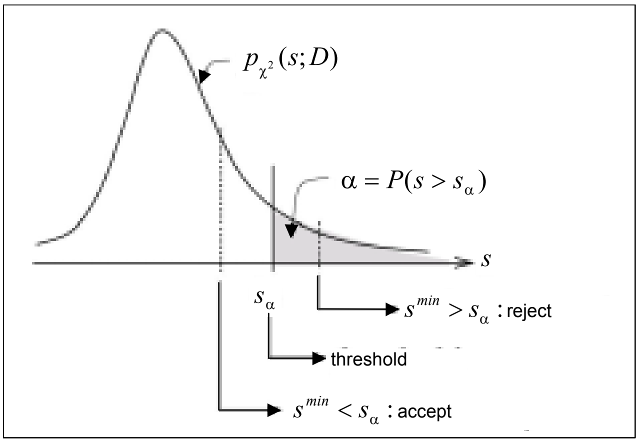
$p_{\chi^2}(s;D)$
$\alpha = P(s > s_\alpha)$
s
s_α
$s^{min} > s_\alpha$: reject
threshold
$s^{min} < s_\alpha$: accept

Fig. 1. Likelihood test.

Figure 1 illustrates the basic concept of the likelihood test. The likelihood test uses probability value $P(s>s^{min})=1-P(s< s^{min})$ as a criterion, where the value P $(s< s^{min})$ roughly means the probability that s^{min} is obtained from the assumed probability density p_{χ^2} $(s;D)$. If $P(s>s^{min})$ is smaller than a pre-defined probability value α, then we consider that s^{min} is not subject to p_{χ^2} $(s;D)$ and decide that $\boldsymbol{x}^{new}$ and $\boldsymbol{x}^{min}$ are from different persons. On the other hand, if $P(s>s^{min})$ is larger than α, then we decide that $\boldsymbol{x}^{new}$ and $\boldsymbol{x}^{min}$ are from the same person.

The value α corresponding to the threshold is determined by the policy of a system builder. If we determine a desired error rate of the test $(P(s>s_\alpha))$, then α can be obtained by solving the equation,

$$P(s > s_\alpha) = 1 - \int_0^{s_\alpha} p(s)ds \, , \tag{9}$$

or easily using the lookup table of percentile values for the χ^2 distribution.

In practical cases, the values s_α and s^{min} are used instead of α and $P(s>s_\alpha)$, which are difficult to calculate. Thus, s_α plays the role of the specific threshold we want in the practical verification tasks. Daugman [5] proposed a similar strategy for binary values using binomial distribution.

However, to apply this strategy, we need a revision. We need to pay attention to the first step of the verification process. As shown in (8), we select s^{min}, which is the minimum of $s_i = s(\boldsymbol{x}^{new}, \boldsymbol{x}_n)$. Therefore, s^{min} is not just a sample from χ^2 distribution, but the order statistics obtained from n samples $\{s_1,\ldots,s_n\}$. Accordingly, s^{min} is not subject to p_{χ^2} $(s;D)$ but subject to $d\{1-(1-p_{\chi^2}$ $(s<s^{min}))\}/ds$ (See [11] for details). Considering this fact, we obtain a revised method for determining the threshold s_α, which is defined by

$$P_{\chi^2}(s < s_\alpha) = 1 - \exp\left\{\frac{1}{n}\log \alpha\right\}. \tag{10}$$

Using (10) and the desired value of α, we can find the corresponding threshold s_α from the lookup table. Here, one can easily see that the value α means the FRR in the test. If we want to decrease the FRR, we can decrease α and find the corresponding s_α. Therefore, the threshold can be determined and can be easily changed according to our goal for the FRR.

The overall verification process can be summarized by the following steps:

Step 1. Set the goal of α and find the corresponding value of s_α using (10) and the lookup table for χ^2 distribution.

Step 2. For new data $\boldsymbol{x}^{new}$, calculate s^{min} and $\boldsymbol{x}^{min}$.

Step 3. If $s^{min} < s_\alpha$, then accept $\boldsymbol{x}^{new}$ as registered and identify the new person as corresponding to $\boldsymbol{x}^{min}$.

Step 4. If $s^{min} > s_\alpha$, then reject x^{new}.

IV. Computational Experiments

In order to check if the proposed method is suitable for a real biometric system, we conducted some computational experiments using real human iris images, which is one of the representative physiological features for biometrics [3], [5], [6], [9]. We collected 375 iris images from 21 persons. From the total data set, we selected 14 persons for the registered individuals, and randomly chose 5 items of data from each registered person. These 70 data items were used for building the identification and verification system. The other 305 data items were used to test the system. The test data set was composed of two groups: the authentic group with 190 data items for the 14 registered persons and the imposter group with 115 data items for 7 non-registered persons.

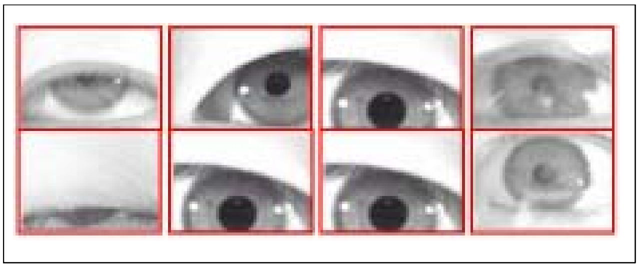

Fig. 2. Examples of images with bad quality.

To obtain the experimental images from the human eye images (Fig. 3), we did the following preprocessing [6]: evaluation of the image quality, iris localization, and normalization on the localized iris area. We first used this method to check the quality of images to determine whether the given iris images were appropriate for the subsequent processing and then to select the proper ones among them in realtime. Some images deemed inappropriate were excluded from the next processing. Figure 2 indicates the type of inappropriate images excluded by the method.

The iris localization on the images deemed proper was required to detect the iris area between pupil and sclera from an eye image. To determine that area exactly, it was important to precisely detect the inner boundary (between the pupil and iris) and the outer boundary (between the iris and sclera). At first, we needed to get the exact reference point, the center of the pupil, and then compute the distance from that point to the boundaries as the radius. We used a three-step technique for detecting the reference point and localizing the iris area from an eye image. In the first step, we applied the Canny edge detector to the image to extract edge components and then labeled the connected components. In the next step, we used a 2D bisection-based Hough transform, not a 2D gradient-based Hough transform [12], to get the center of the pupil. In the last step for the iris localization, we validated the existence of a circle and calculated its radius with a radius histogram technique.

We used a normalization process on the localized iris area to compensate for size variations due to the possible changes in the camera-to-face distance and to facilitate the feature extraction process by converting the iris area represented by a polar coordinate system into a Cartesian coordinate system.

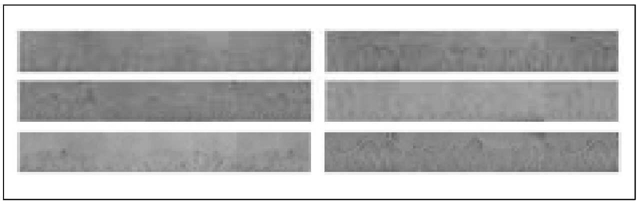

Fig. 3. Examples of human iris images after preprocessing.

Since the iris images in Fig. 3 have 7200 pixels (225×32), we first applied the principal component analysis method to reduce the dimensions, which is a common method for high dimensional data [7], [13]. With the principal component analysis, we obtained 70 dimensional feature vectors for each item of raw data. Using the obtained feature set, we made the set of $\boldsymbol{y}$, the difference vector.

To check the performance of the proposed similarity measure, we first conducted an identification (classification) test for the registered individuals using the authentic group in the test data set. For the test data, we calculated the similarity with all registered data, and found the $\boldsymbol{x}^{min}$ that was the most similar to the test data. Then we assigned the test data to the class in which $\boldsymbol{x}^{min}$ was included. We compared the proposed measure with the simplified Mahalanobis distance and the Euclidean distance (Table 1). The result confirms that the proposed measure is superior to the standard methods.

Table 1. Identification results for the authentic test data.

Similarity measure	Classification rate
Simplified Mahalanobis	83.68%
Euclidean distance	90.53%
Proposed measure	98.95%

Table 2. Identification results using k-nearest neighbor rules (%).

	k=1	k=2	k=3	k=4	k=5
Simplified Mahalanobis	83.68	83.68	26.32	26.32	13.16
Euclidean distance	90.53	90.53	85.26	85.26	82.83
Proposed measure	98.95	98.95	98.42	98.42	96.84

Table 3. Verification results on the test data.

	Threshold = 47.89	Threshold = 46.76
FRR	1.05%	1.58%
FAR	8.70%	6.96%

We also conducted another simple experiment with the k-nearest neighbor rule to compare the proposed method with other standard methods. We can see from Table 2 that the proposed method is better than other methods with regard to the identification performance.

For the verification task, we used two thresholds, 47.89 (for α = 0.05) and 46.76 (for α = 0.1), which was obtained using (10). The verification test was conducted for 190 authentic items of data and for 115 imposter items. The result is shown in Table 3. This result suggests that the proposed method can be applied to practical biometric systems.

The ROC curve [14], [15] of Fig. 4 also demonstrates the relation between the FAR and the FRR by generating a continuously varying threshold. Varying the threshold trades the FAR off against the FRR, so it can be changed according to the security level of application problems. We can get an equal error rate of 4.28% at around the threshold 44.0.

V. Conclusions and Discussions

In this paper, we considered the properties of data sets used in biometrics and proposed a statistical framework of identification and verification for biometric systems. In order to

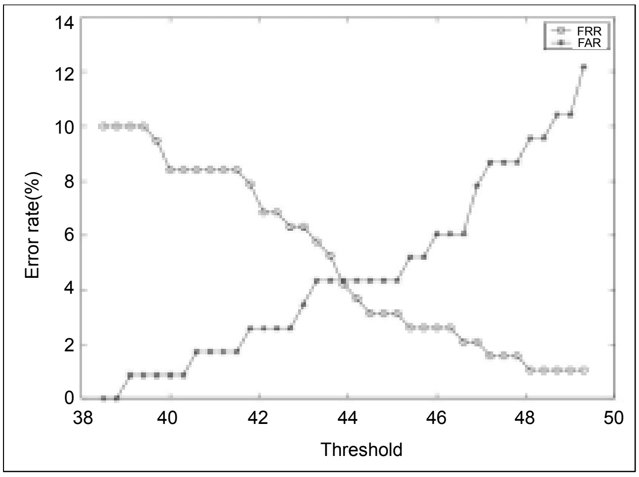

Fig. 4. ROC curve for the proposed method.

get a novel biometric system with good performance, many parts of the system need to be optimized to the specific features of given data. To do so, more sophisticated data processing methods, such as neural networks, can be used for better performance. Nevertheless, we can expect that the proposed method is meaningful as a standard statistical method in treating the data for identification and verification problems. In addition, for the newly introduced random variable $\boldsymbol{y}$ in this paper, more novel data processing tools can be applied to estimate its probability distribution so that we can expect better performance. Even in this case, the proposed method for determining a threshold can still be applied to the estimated density function $p(\boldsymbol{y})$.

The proposed method is based on the assumption that the number of individuals is large and the number of items of data from each individual is small. Therefore, the proposed method can be applied to any other applications with these properties, such as the diagnosis of diseases from pathological data.

References

[1] M. Bartlett and T. Sejnowsky, "Viewpoint Invariant Face Recognition Using Independent Component Analysis and Attractor Networks," *Neural Information Proc. Systems-Natural and Synthetic*, vol. 9, 1997, pp. 817-823.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 19, no. 7, 1997, pp. 711-720.

[3] W.W. Boles, B. Boashash, "A Human Identification Technique Using Images of the Iris and Wavelet Transform," *IEEE Trans. on Signal Proc.*, vol. 46, no. 4, 1998, pp. 1185-1188.

[4] W. Campbell, "A Sequence Kernel and Its Applications to Speaker Recognition," *Advances in Neural Information Proc. Systems*, 2001.

[5] J.G. Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, 1993, pp. 1148-1161.

[6] G. Kee, Y. Byun, K. Lee, and Y. Lee, "Improved Techniques for an Iris Recognition System with High Performance," *AI 2001: Advances in Artificial Intelligence*, LNAI 2256, 2001, pp. 177-188.

[7] O. Lee, H. Park, and S. Choi, "PCA vs. ICA for Face Recognition," *The 2000 Int'l Technical Conf. on Circuits/Systems, Computers, and Commun.*, 2000, pp. 873-876.

[8] S. Lim, K. Lee, O. Byeon, and T. Kim, "Efficient Iris Recognition through Improvement Feature Vector and Classifier," *ETRI J.*, vol. 23, no. 2, 2001, pp. 61-70.

[9] R.P. Wildes, "Iris Recognition: An Emerging Biometric Technology," *Proc. of the IEEE*, vol. 85, no. 9, 1997, pp. 1348-1363.

[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2ed, Academic Press, INC. 1990.

[11] A. Stuart and K. Ord, *Kendall's Advanced Theory of Statistics*, 1, 6ed., Edward Arnold, 1994.

[12] D. Ioammou, W. Huda, A.F. Laine, "Circle Recognition through a 2D Hough Transform and Radius Histogramming," *Image and Vision Computing*, vol. 17, 1999, pp. 15-26.

[13] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[14] R.M. Bolle, S. Pankanti, and N.K. Ratha, "Evaluation Techniques for Biometrics-based Authentication System (FRR)," *Proc. of the Fifteens International Conf. on Pattern Recognition*, vol. 2, 2000, pp. 831-837.

[15] W. Shen, M. Surette, and R. Khanna, "Evaluation of Automated Biometrics-based Identification and Verification," *Proc. of the IEEE*, vol. 85, 1997, pp. 1464-1478.

Kwanyong Lee received his MS and PhD degrees in computer science from Yonsei University in Seoul, Korea in 1991 and 1994. From 1997 to 1999, he joined the Department of Information and Communication Engineering at the University of Tokyo in Japan as a Visiting Researcher. In 1999, he was a Senior Researcher in the EC/CALS division of the Electronics and Telecommunications Research Institute (ETRI) in Daejon, Korea. At present he is a Professor at the Department of Computer, Information, and Communication in the Korea Cyber University. His research interests include pattern recognition, image processing, and biometrics.

Hyeyoung Park got his PhD degree from the Department of Computer Science in Yonsei University in Seoul, Korea in 2000. She is working as a Research Scientist in the Brain Science Institute in RIKEN, Japan. Her main research interests lie in computational learning theory, statistical inference, pattern recognition, and brain science.