Computational Analysis of PCA-based Face Recognition Algorithms

Hyeon-Joon Moon[†] and Sang-Hoon Kim[††]

ABSTRACT

Principal component analysis (PCA) based algorithms form the basis of numerous algorithms and studies in the face recognition literature. PCA is a statistical technique and its incorporation into a face recognition system requires numerous design decisions. We explicitly take the design decisions by introducing a generic modular PCA-algorithm since some of these decision are not documented in the literature. We experiment with different implementations of each module, and evaluate the different implementations using the September 1996 FERET evaluation protocol (the de facto standard method for evaluating face recognition algorithms). We experiment with (1) changing the illumination normalization procedure; (2) studying effects on algorithm performance of compressing images using JPEG and wavelet compression algorithms; (3) varying the number of eigenvectors in the representation; and (4) changing the similarity measure in classification process. We perform two experiments. In the first experiment, we report performance results on the standard September 1996 FERET large gallery image sets. The result shows that empirical analysis of preprocessing, feature extraction, and matching performance is extremely important in order to produce optimized performance. In the second experiment, we examine variations in algorithm performance based on 100 randomly generated image sets (galleries) of the same size. The result shows that a reasonable threshold for measuring significant difference in performance for the classifiers is 0.10.

PCA기반의 얼굴인식 알고리즘들에 대한 연산방법 분석

문현준[†] · 김상훈[††]

요 약

얼굴인식 기술 분야에 있어서 Principal component analysis (PCA)기반 알고리즘은 많은 관련 알고리즘의 기초가 되고 있다. PCA는 매우 통계적인 접근이며 얼굴인식 분야에 응용하기 위해서는 많은 설계 결정요인(design decision)을 필요로 한다. 본 논문에서는 일반적인 modular PCA알고리즘을 소개하면서 design decision을 얻는다. 얼굴인식 알고리즘 평가에 대한 표준 접근 방법인 September 1996 FERET evaluation protocol을 활용하여 각 모듈에 대한 서로 다른 구현방법을 실험하고 평가한다. 실험조건으로는 (1) 조도의 정규화 과정을 변화 (2) JPEG과 wavelet compression 알고리즘 사용에 대한 성능효과를 분석 (3) 표현방법에서 eigenvectors의 수를 조절 (4) 분류과정에서 유사도 측정방법을 변경 하는 등이다. 본 논문에서는 standard September 1996 FERET의 대용량 gallery image set에 대해 적용해 본 결과에 대해 정리하며, 100개의 무작위로 발생된 image set에 대해서도 알고리즘의 성능 변화를 평가한다.

Key words: face recognition, PCA, eigenvector, illumination normalization

"This work was supported (in part) by Biometrics Engineering Research Center, (KOSEF)"

접수일 : 2002년 10월 19일, 완료일 : 2003년 1월 24일

[†] 연세대학교 전기전자공학부 조교수

[††] 정회원, 국립한경대학교 정보제어공학과 조교수

1. Introduction

Principal Component analysis(PCA) is a statistical method for reducing the dimensionality of high dimensional data, where the data is represented as a vector. PCA is popular because it is easy to implement, is a natural dimensionality reduction method.

The PCA algorithm is widely used in face recognition since it can be easily implemented and its dimensionality reduction capability.

Also, PCA achieves reasonable performance levels[1] and the bases for algorithms[1,2], and serves as a computational model in psycho-physics [4,5,9]. In designing an algorithm around PCA, a number of critical design issues have to be addressed. Each of these design decisions has an impact on the overall performance of the algorithm. Some of these design decisions have been explicitly stated in the literature; for example, the distance function in the nearest neighbor classifier. However, a large number of decisions are not mentioned and are passed from researcher to researcher by word of mouth. For example, illumination normalization and number of eigenfeatures included in the representation.

Because the design details are not explicitly stated, a reader cannot assess the merits of a particular implementation and the associated claims. This can unnecessarily caste a shadow on performance claims of a new algorithm when a PCA-based algorithm is used as a strawman. Knowledge of the basic strengths and weaknesses of different implementations can provide insight and guidance in developing algorithms that build on PCA. We present a generic modular PCA-based face recognition algorithm. The algorithm consists of preprocessing, PCA decomposition, and recognition modules. Each module consists of a series of basic steps, where the purpose of each step is fixed. However, we systematically vary the algorithm in each step. For example, the classifier step will always recognize a face, but we will experiment with different classifiers. The selection of which algorithm is in each step is a design decision. Based on the generic model for PCA-based algorithms, we evaluate different implementations. Because we use the generic model, we can change the implementation in an orderly manner and assess the impact on performance of these modification. The algorithms are evaluated with the FERET testing procedure[1]. Experiment one does a detailed evaluation of variations in the implementation. Results are reported for standard galleries and probe sets described in[1]. The *gallery* is the set of known individuals. An image of unknown face presented to the algorithm is called a *probe*, and the collection of probes is called the *probe set*. By testing on standard galleries and probe sets, the reader can compare the performance of our PCA implementations with the algorithms tested under the FERET program. In experiment one, we vary the illumination normalization procedure, the number of eigenvectors in the representation, and the distance in the classifier; and we study the effects of compressing facial images on algorithm performance. The effects of image compression on recognition has not previously been studied. This is of interest in applications where image storage space or image transmission time are critical parameters.

In algorithm evaluation, the following two critical questions are often ignored. First, how does performance vary with different galleries and probe sets. Second, when is the difference in performance between two algorithms significant. In experiment two, we look at this question by randomly generating 100 galleries of the same size. We then calculate performance on each of the galleries against two different categories of probes (probes taken on the same day as the gallery images and probes taken on different days than the gallery images). Because we have 100 scores for each category of probe, we can examine the range of

scores, and the overlap in scores among different implementations of the PCA algorithm.

2. PCA-based Face Recognition System

2.1 Principal component analysis (PCA)

Principal component analysis (PCA) is a statistical dimensionality reduction method, which produces the optimal linear least squared decomposition of a training set. Kirby and Sirovich[6] applied PCA to representing faces and Turk and Pentland[5,10] extended PCA to recognizing faces. In PCA, an image is represented as a point in $R^{n \times m}$, where the image is n by m pixels. The input to the PCA algorithm is a training set of images where the ensemble mean is zero. From a training set of N images, PCA generates a set of $N-1$ eigenvectors and eigenvalues. (In the literature, the eigenvectors are sometimes referred to as eigenfaces.) We normalize the eigenvectors so that they are orthonormal. The eigenvectors are ordered so that $\lambda_i > \lambda_{(i+1)}$, where λ_i are the eigenvalues. The λ_i's are equal to the variance of the projection of the training set onto the i th eigenvector. Thus, the low order eigenvectors encode the larger variations in the training set (low order refers to the index of the eigenvectors and eigenvalues).

In a PCA-based face recognition algorithm, the face is represented by its projection onto a subset of M eigenvectors, which we will call *face space*. This is represented as a point in R^M. A gallery of K individuals is represented as K points g_j in face space. Let p be the representation of a probe in face space. The probe is identified as person j^* in the gallery, if distance $d(p, g_{j^*}) = \min_j d(p, g_j)$; i.e., the probe is identified as the person that minimizes the distance between p and the g_j's. Selecting the distance function d is one of the algorithm design decisions.

2.2 System modules

Our face recognition system is shown in Fig. 1.

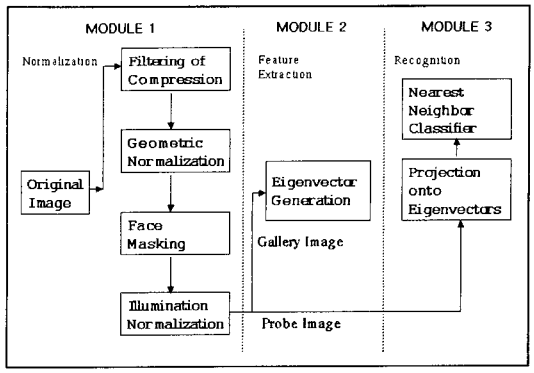


Fig. 1. Block Diagram of PCA-based Face Recognition System.

and consists of three modules and each module is composed of a sequence of steps. The first module preprocesses the input image. The goal of the preprocessing is to transform the facial image into a standard format that removes variations that can affect recognition performance. This module consists of four steps. The first step filters or compresses the original image. The image is filtered to remove high frequency noise in the image. An image is compressed to save storage space and reduce transmission time. The second step places the face in a standard geometric position by rotating, scaling, and translating the center of eyes to standard locations. The goal of this step is to remove variations in size, orientation, and location of the face. The third step masks out background pixels, hair, and clothes. The masked regions vary from day to day and can interfere with identification. The fourth step removes some of the variations in illumination between images. Changes in illumination are critical factors in algorithm performance[1].

The second module performs the PCA decomposition on the training set. This produces the eigenvectors (eigenfaces) and eigenvalues. The third module identifies the face from a preprocessed image, and consists of two steps. The first step projects the image onto the eigenvectors that represent the face. The critical parameter in this step is the subset of eigenvectors that

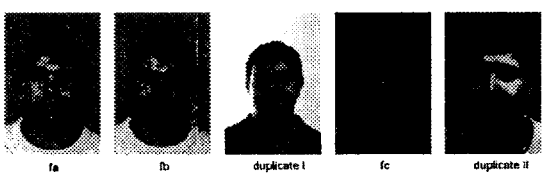
represent the face. The second step recognizes faces using a nearest neighbor classifier. The critical parameter in this step is the distance function in the classifier.

3. Test Design

3.1 FERET Database

The FERET database provides a common database of facial images for both development and testing of face recognition algorithms and has become the de facto standard for face recognition of still images[1,2]. In the FERET database, images of individual were acquired in sets of 5 to 11 images. Each set includes two frontal views **fa** and **fb**; a different facial expression was requested for the second frontal image. For 200 sets of images, a third frontal image was taken with a different camera and different lighting (**fc**). (The remaining images were non-frontal images and were not used in this study.) One emphasis of the database collection was on duplicate image sets. *Duplicate* is defined as an image of a person whose corresponding gallery image was from a different image set (usually taken on a different date). The database contains 365 duplicate sets of images. For 91 duplicate sets, the time between the first and last sittings was at least 18 months. In an effort to maintain a degree of consistency throughout the database, the photographer used the same physical setup in each session. However, because the equipment had to be reassembled for each session, there was variation from session to session. This results in variations in scale, pose, expression, and illumination of the face. For details of the FERET database, refer to[7].

Sample images are shown in following Figures.



3.2 Design rule

To obtain a robust comparison of algorithms, it is necessary to calculate performance on a large number of galleries and probe sets. To allow scoring on multiple galleries and probe sets, we have to adopt an appropriate protocol. In the new protocol, an algorithm is given two sets of images: the *target set* and the *query set*. We introduce this terminology to distinguish these sets from the galleries and probe sets that are used in computing performance statistics. The target set is given to the algorithm as the set of known facial images. The images in the query set are the unknown facial images to be identified.

Definition : For each image q_i in the query set $\boldsymbol{Q}$, an algorithm reports the similarity $s_{i(k)}$ between q_i and each image t_k in the target set $\boldsymbol{T}$. The key property, which allows for greater flexibility in scoring, is that for any two images s_i and t_k, we know $s_{i(k)}$. From the output files, algorithm performance can be computed for virtual galleries and probe sets. Thus, we can create and score results for any gallery $G \subset P$ and any probe set $P \subset Q$. For a given gallery G and probe set P, the performance scores are computed by examination of the similarity measures $s_{i(k)}$ such that $q_i \in P$ and $t_k \in G$. We refer to such galleries and probe sets as virtual galleries and virtual probe sets (because they are subsets of the target set and query set).

In our experiments, we report performance for four different categories of probes. The **FB** probes are the second frontal images from the same session as the frontal images that are in the gallery. (In a gallery, either the **fa** or **fb** images are placed in the gallery, and the other image is placed in the probe set. We will denote those frontal images that are placed in the probe set by **FB**.) The **fc** probes are the **fc** image from the same session as the frontal image in the gallery. (Note: there are only 200 sets of images with **fc** images.) The *duplicate I* probes are the duplicate frontal images of the

images in the gallery (frontal images from different image sets---usually different days). The *duplicate II* probes are images where there is at least a year and half between the acquisition of the gallery and probe images. We report identification results using a closed universe model. In the closed universe, every probe is in the gallery. The complement to the closed universe is the open universe where some probes are not in the gallery. The open universe model is used in verification and authentification scenarios. The results for these scenarios can be found in Rizvi et al.[8].

For the closed universe, we report performance as cumulative match scores. The rank is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches. The cumulative match scores can be calculated for any subset of the probe set. We calculated this score to evaluate an algorithm's performance on different categories of probes. The computation of the score is quite simple. Let P be the number of probes to be scored and R_k the number of these probes in the subset that are in the top k. The fraction is reported as R_k/P. The top rank match (or score) is R_1 (the fraction of probes correctly identified).

4. Experiment I

The purpose of experiment I is to examine the effects of changing the steps in our generic PCA-based face recognition system. We do this by establishing a baseline algorithm and then varying the implementation of selected steps one at a time. Ideally, we would test all possible combination of variations. However, because of the number of combinations, this is not practical and we vary the steps individually. The baseline algorithm has the following configuration: The images are not filtered or compressed. Geometric normalization consists of rotating, translating, and scaling the images so the center of the eyes are on standard pixels. This is followed by masking the hair and background from the images. In the illumination normalization step, the non-masked facial pixels were normalized by a histogram equalization algorithm. Then, the non-masked facial pixels were transformed so that the mean is equal to 0.0 and standard deviation is equal to 1.0. The geometric normalization and masking steps are not varied in the experiments in this paper. The training set for the PCA consists of 501 images (one image per person), which produces 500 eigenvectors. The training set is not varied in the experiments in the paper. In the recognition module, faces are represented by their projection onto the first 200 eigenvectors and the classifier uses the L_1 norm.

4.1 Test sets, galleries, and probe sets

All images are from the FERET database, and the testing was done with the September 1996 FERET protocol. In this protocol, the target set contained 3323 images and the query set 3816 images. All the images in the target set were frontal images. The query set consisted of all the images in the target set plus non-frontal images and digitally modified images. (The non-frontal and digitally modified images were not included in our analysis.) We report results for four different probe categories. The size of the galleries and probe sets for the four probe categories are presented in Table 1. The **FB**, **fc**, and duplicate I galleries are the same. The duplicate II gallery is a subset of the other galleries.

Table 1. Size of galleries and probe sets for different probe categories.

Probe category	duplicate I	duplicate II	FB	fc
Gallery size	1196	864	1196	1196
Probe set size	722	234	1195	194

4.2 Variations in the normalization module

4.2.1 Illumination normalization

We experimented with three variations to the

illumination normalization step. For the baseline algorithm, the non-masked facial pixels were transformed so that the mean was equal to 0.0 and standard deviation was equal to 1.0 followed by a histogram equalization algorithm. First variation, the non-masked pixels were not normalized (original image). Second variation, the non-masked facial pixels were normalized with a histogram equalization algorithm. Third variation, the non-masked facial pixels were transformed so that the mean was equal to 0.0 and variance equal to 1.0. The performance results from the illumination normalization methods are presented in Table 2.

Table 2. Performance results for illumination normalization methods. Performance score are the top rank match.

Illumination normalization method	Probe category			
	duplicate I	duplicate II	FB probe	fc probe
Baseline	0.35	0.13	0.77	0.26
Original image	0.32	0.11	0.75	0.21
Hidtogram Eq. only	0.34	0.12	0.77	0.24
μ=0.0, σ=1.0 only	0.33	0.14	0.76	0.25

4.2.2 Compressing and filtering the images

We examined the effects of JPEG and wavelet compression, and low pass filtering (LPF) on recognition. For this experiment, the original images were compressed and then uncompress prior to being feed into the geometric normalization step of the normalization module. For both compression methods, the images were compressed approximately 16:1 (0.5 bits per pixel). We experimented with other compression ratios and found that performance was comparable. However, the performance starts to degrade if the compression rate is 64:1 (0.125 bits per pixel) and low bit rate coding scenario. The results are for eigenvectors generated from non-compressed images. We found that performance in this case was slightly better than on eigenvectors trained from compressed images. Because compression algorithms usually low pass filter the images, we decided to examine the effects on performance of low pass filtering the original image. The filter was a 3×3 spatial filter with a center value of 0.2 and the remaining values equal to 0.1. Table 3 reports performance for the baseline algorithm, JPEG and wavelet compression, and low pass filtering.

Table 3. Performance score for low pass filter, JPEG, and wavelet compressed images (0.5 bits/pixel compression). Performance scores are the top rank match.

Normalization	Probe category			
	duplicate I	duplicate II	FB probe	fc probe
Baseline	0.35	0.13	0.77	0.26
JPEG	0.35	0.13	0.78	0.25
Wavelet	0.36	0.15	0.79	0.25
LPF	0.36	0.15	0.79	0.24

4.3 Variations in the recognition module

4.3.1 Number of low order eigenvectors

The higher order eigenvectors which are associated with smaller eigenvalues encode small variations and noise among the images in the training set. One would expect that the higher order eigenvectors would not contribute to recognition. We examined this hypothesis by computing performance as a function of the number of low order eigenvectors in the representation. Fig. 2. shows the top rank score for **FB** and duplicate I probes as the function of the number of low order eigenvectors included in the representation in face space. The representation consisted of $e_1, \ldots, e_n$, n = 50, 100, ...,500, where e_is are the eigenvectors generated by the PCA decomposition.

4.3.2 Removing low order eigenvectors

The low order eigenvectors encode gross differences among the training set. If the low order eigenvectors encode variations such as lighting

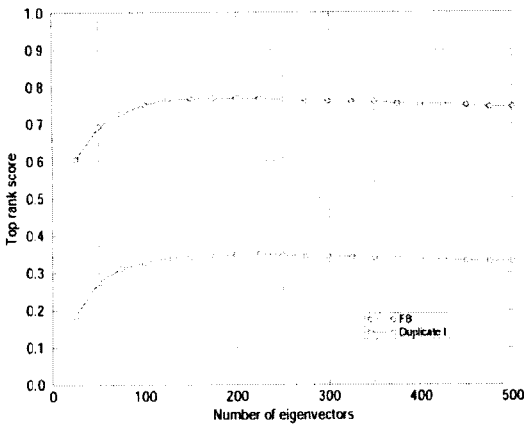


Fig. 2. Performance on duplicate I and FB probes based on number of low order eigenvectors used. (Number of images in gallery = 1196; Number of FB images in probe set = 1195, Number of duplicate I in probe set = 722).

changes, then performance may improve by removing the low order eigenvectors from the representation. We looked at this hypothesis by removing the 1, 2, 3, and 4th eigenvector from the representation; i.e., the representation consisted of $e_i, \ldots, e_{200}$, i = 1,2,3,4,5. The performance results from these variations are given in Table 4. Among the different category of probes, there is a noticeable variation in performance for **fc** probe as shown in Fig. 3.

4.3.3 Nearest neighbor classifier

We experimented with seven similarity measures for the classifier. The detailed classifiers are

Table 4. Performance score with low order eigenvectors removed. Performance scores are the top rank match

Number of low order eigenvectors removed	Probe category			
	duplicate I	duplicate II	FB probe	fc probe
0 (Baseline)	0.35	0.13	0.77	0.26
1	0.35	0.15	0.75	0.38
2	0.34	0.14	0.74	0.36
3	0.31	0.14	0.72	0.37
4	0.20	0.09	0.50	0.22

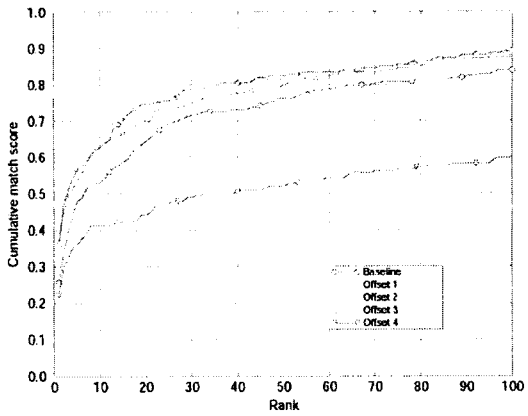


Fig. 3. Performance on fc probes with first 1, 2, 3, and 4th eigenvectors removed.

described in Appendix. They are listed in Table 5, along with the results. The performance score for **fc** probes shows most variation among different category of probes. The results for **fc** probes are presented in Fig. 4.

Table 5. Performance scores based on different nearest neighbor classifier. Performance scores are the top rank match.

Nearest neighbor classifier	Probe category			
	duplicate I	duplicate II	FB probe	fc probe
Baseline(L_1)	0.35	0.13	0.77	0.26
Euclidean(L_2)	0.33	0.14	0.72	0.04
Angle	0.34	0.12	0.70	0.07
Mahalanobis	0.42	0.17	0.74	0.23
L_1+Mahalanobis	0.31	0.13	0.73	0.39
L_2+Mahalanobis	0.35	0.13	0.77	0.31
Angle+ Mahalanobis	0.45	0.21	0.77	0.24

4.4 Discussion

In Experiment I, we conducted experiments that systematically varied the steps in each module based on our PCA-based face recognition system. The goal belongs to understand the effects on performance scores from these variations. In the normalization module, we experimented with varying the illumination normalization and compres-

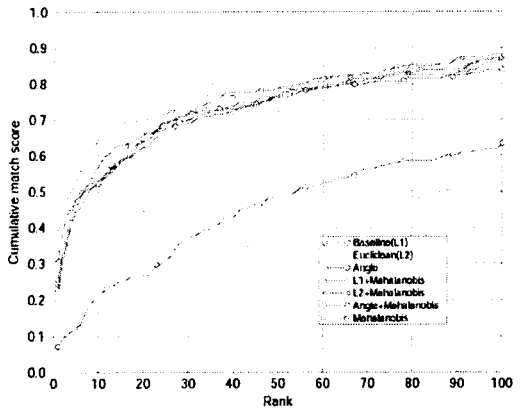


Fig. 4. Effects of nearest neighbor classifier for face recognition. Performance scores for fc probes.

sion steps. The results show that performing an illumination normalization step improves performance, but which implementation that is selected is not critical. The results also show that compressing or filtering the images does not significantly effect performance.

In the recognition module, we experimented with three classes of variations. First, we varied the number of low order eigenvectors in the representation from 50 to 500 by steps of 50. Fig 2. shows that performance increases until approximately 200 eigenvectors are in the representation and then performance decreases slightly. Representing faces by the first 40% of the eigenvectors is consistent with results on other facial image sets that the authors have seen.

Second, low order eigenvectors were removed. Table 4 shows that removing the first eigenvector resulted in an overall increase in performance. The largest increase was observed with the **fc** probes. This increase is further highlighted in Fig. 3. The low order eigenvectors encode the greatest variations among the training set. The most significant difference between the **fc** probes and the gallery images was a change in lighting. If the low order eigenvectors encode lighting differences, then this would explain the substantial increase in performance by removing the first eigenvector.

Third, the similarity measure in the nearest neighbor classifier was changed. This variation showed the largest range of performance. For duplicate I probes, performance ranged from 0.31 to 0.45, and for **fc** probes the range was from 0.07 to 0.39. For duplicate I, duplicate II and **FB** probes, the angle+Mahalanobis distance performed the best. For the **fc** probes, the L_1+Mahalanobis distance performed the best. But, this distance was the worst for the duplicate I probe.

Because of the range of performance, it is clear that selecting the similarity measure for the classifier is the critical decision in designing a PCA-based face recognition system. However, decision of selecting similarity measure is dependent on the type of images in the galleries and probe sets that the system will process.

5. Experiment II

In experiment I, for some variations in components, the change in performance appears to small. Whereas, for others, the range appears to be considerable, i.e., the nearest neighbor classifier. The natural question is, when is the difference in performance between two variations significant? In this experiment we examine this question for the nearest neighbor classifier. We choose to study changing the classifiers because they had the greatest variation in performance. To address this question, we randomly generated 100 galleries of 200 individuals, with one frontal image per person. The galleries were generated without replacement from the **FB** gallery of 1196 individuals in experiment one. (Thus, there is overlap between galleries.) Then we scored each of the galleries against the **FB** and duplicate I probes for each of the seven classifiers in experiment one. (There were not enough **fc** and duplicate II probes to compute performances for these categories.)

For each randomly generated gallery, the corresponding **FB** probe set consisted of the second

frontal image for all images in that gallery; the duplicate I probe set consisted of all duplicate images for each image in the gallery. We measured performance by the top rank score (the fraction of probes that were correctly identified). For an initial look at the range in performance, we examine the baseline algorithm (L_1 similarity measure). There are similar variations for the six remaining distances. For each classifier and probe category, we had 100 different scores. Fig. 5 presents the histogram of top rank scores for the baseline algorithm ($L1$ similarity measure) for both {**FB**} and duplicate I probe sets. This shows a range in performance ranges from 0.80 to 0.92 for {**FB**} probe, from 0.29 to 0.59 for duplicate I probe. This clearly shows a large range in performance of the 100 galleries.

We present a truncated range of top rank scores for the seven different nearest neighbor classifiers in Fig 6. Fig. 6(a) shows the range for {**FB**} probes and Fig. 6(b) for duplicate I probes. For each classifier, score is marked with; the median by $\times$, the 10th percentile by $+$, and 90th percentile by $*$. We plotted these values because they are robust statistics. We selected the 10th and 90th percentile because they mark a robust range of scores and

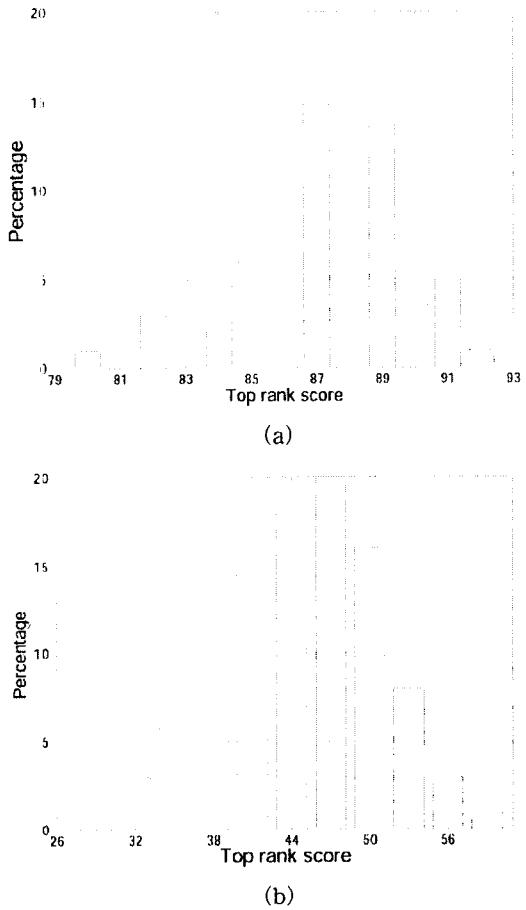


Fig. 5. Histogram of top rank scores of the baseline algorithm (L_1 similarity measure) (a) FB probes and (b) duplicate I probes.

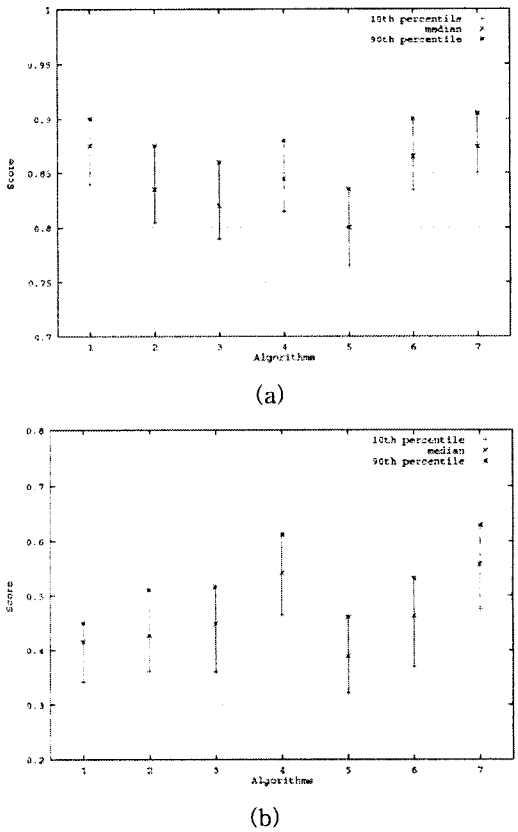


Fig. 6. The range of top rank scores using seven different nearest neighbor classifiers. The nearest neighbor classifiers presented are: (1) $L1$, (2) $L2$, (3) Angle, (4)Mahalanobis, (5) L_1+Mahalanobis, (6) L_2+Mahalanobis, and (7)Angle+Mahalanobis. (a) FB probes and (b) duplicate I probes

outliers are ignored. From these results, we get a robust estimate of the overall performance of each classifier.

5.1 Discussion

The main goal of experiment two was to get a rough estimate of when the difference in performance is significant. From Fig. 6, the range in scores is approximately ± 0.05 about the median for all 14 runs. This suggests a reasonable threshold for measuring significant difference in performance for the classifiers is ~0.10. The results for duplicate I probes in experiment two are consistent with the results in experiment one. In Table 5, the top classifiers were the Mahalanobis and angle+Mahalanobis and these two classifiers produces better performance than the other methods as shown in Table 6. In both experiments, the L_1+Mahalanobis received the lowest scores. This suggest that for duplicate I scores that the angle+Mahalanobis or Mahalanobis distance should be used. Based on the results of this experiment, performance of smaller galleries can predict relative performance on larger galleries.

For **FB** probes, there is not as sharp a division among classifiers. One possible explanation is that in experiment one, the top match scores for the **FB** probes did not vary as much as the duplicate I scores. There is consistency among the best scores (L_1, L_2+Mahalanobis, and angle+Mahalanobis). The remaining classifiers' performances can be grouped together. The performance scores of these classifiers are within each other's error margins. We defined *error margins* as a robust range of performance scores. This suggests that either the L_1, L_2+Mahalanobis, or angle+Mahalanobis distance should be used.

Table 6. Comparison of identification performance scores for Baseline, Proposed I, and Proposed II algorithms. Performance scores are the top rank match

Algorithm	Probe category			
	duplicate I	duplicate II	FB probe	fc probe
Baseline	0.35	0.13	0.77	0.26
Proposed I	0.49	0.26	0.78	0.26
Proposed II	0.40	0.26	0.78	0.33

6. Conclusion

We have presented a design methodology of configuring PCA-based algorithm based on empirical performance results. The heart of the methodology is a generic modular design for PCA-based face recognition systems. This allowed us to systematically vary the components and measure the impact of these variations on performance. Our experiments show that quality and type of images to be processed is the driving factor in determining the design of a PCA-based system. Based on the experiments we propose two algorithms. The algorithms are called proposed I and proposed II. The proposed I algorithm is optimized for duplicate I and II probes. The proposed II algorithm is optimized for **fc** probes.

The components of the proposed I algorithm are:

- illumination normalization (μ=0.0 and σ=1.0),
- Low-pass filtering,
- remove first low order eigenvector, and
- angle+Mahalanobis distance.

The components of the proposed II algorithm are:

- illumination normalization (μ=0.0 and σ=1.0),
- wavelet compression (0.5 bpp),
- remove first low order eigenvector, and
- L_1+Mahalanobis distance.

Table 6 presents the identification scores for both proposed algorithms. For **FB** probes, the scores for all three algorithms are not significantly different. The proposed I algorithm has better performance scores for duplicate I probes than both the baseline and proposed II algorithms. The proposed II algorithm has better performance

scores for **fc** probes than both the baseline and proposed I algorithms. On the duplicate II probes, both proposed algorithms out perform the baseline algorithm.

From the series of experiments with PCA-based face recognition system, we have come to four major conclusions.

First, JPEG and wavelet compression algorithms do not degrade performance. This is important because it indicates that compressing images to save transmission time and storage costs will not reduce algorithm performance. However, the performance starts to degrade if the compression rate is 64:1 (0.125 bits per pixel) and low bit rate coding scenario.

Second, selection of the nearest neighbor classifier is the critical design decision for PCA-based algorithms. The proper selection of nearest neighbor classifier is essential to improve performance scores. Furthermore, our experiments shows similarity measures that achieve the best performance are not generally considered in the literature.

Third, the performance scores vary among the probe categories, and that the design of an algorithm needs to consider the type of images that the algorithm will process. The FB and duplicate I probes are least sensitive to system design decisions, while **fc** and duplicate II probes are the most sensitive.

Fourth, the performance within a category of probes can vary greatly. This recommends that when comparing algorithms, performance scores from a set of galleries and probe sets need to be examined. We generated 100 galleries and calculate performance against **fb** and duplicate probes. Then, we examined the range of scores, and the overlap in scores among different implementations.

For psycho-physics studies, our conclusions have a number of implications. First, face recognition studies should include a range of images in terms of quality. For example, when measuring the concord between algorithm and human performance, the results should be based on experiments on multiple probe categories. Second, the fine details of algorithm implements can have significant impact on results and conclusion. This conclusion can easily extend to other algorithms. Because, like PCA, the majority of the face recognition algorithms in the literature are view-based and have the same basic architecture as our PCA-based system.

Appendix

L_1 distance

$$d(x,y)=|x-y|=\sum_{i=1}^{k}|x_i-y_i|$$

L_2 distance

$$d(x,y)=\|x-y\|^2=\sum_{i=1}^{k}(x_i-y_i)^2$$

Angle between feature vectors

$$d(x,y)=-\frac{x\cdot y}{\|x\|\|y\|}=-\frac{\sum_{k=1}^{k}x_iy_i}{\sqrt{\sum_{k=1}^{k}x_i^2\sum_{k=1}^{k}y_i^2}}$$

Mahalanobis distance

$$d(x,y,z)=-\sum_{i=1}^{k}x_iy_iz_i$$

$$i=\sqrt{\frac{\lambda_i}{\lambda_i+\alpha^2}}\simeq\frac{1}{\sqrt{\lambda_i}},\ \alpha=0.25$$

where λ_i=eigenvalue of ith eigenvector.

L_1 + Mahalanobis distance

$$d(x,y,z)=-\sum_{i=1}^{k}|x_i-y_i|z_i$$

L_2 + Mahalanobis distance

$$d(x,y,z)=-\sum_{i=1}^{k}(x_i-y_i)^2z_i$$

Angle + Mahalanobis distance

$$d(x,y,z)=\frac{\sum_{k=1}^{k}x_iy_iz_i}{\sqrt{\sum_{k=1}^{k}x_i^2\sum_{k=1}^{k}y_i^2}}$$

Reference

[1] Phillips, P.J., Moon, H., Rauss, P., and Rizvi, S. : The FERET evaluation methodology for

face recognition algorithms, In proceedings Computer Vision and Pattern Recognition 97, pp. 137-143, 1997.

[2] Moghaddam, B. and Pentland, A. : Face Recognition using view-based and modular eigenspaces. In Proc. SPIE Conference on automatic Systems for the Identification and Inspection of Human, vol. SPIE Vol.2277, pp12-21, 1994.

[3] Moghaddam, B. and Pentland, A. : Maximum likehood detection of faces and hands. In Bichsel, M., editor, International Workshop on Automatic Face and Gesture Recognition, pp. 122-128, 1995.

[4] Hancock, P.J.B., Burton, A.M., and Bruce, V. : Face Processing : human perception and principal component analysis. Memory and Cognition, 24(1):26-40, 1996.

[5] Turk, M. and Pentland, A. : Eigenfaces for recognition. J. Cognition Neuroscience, 3(1): 71-86, 1991.

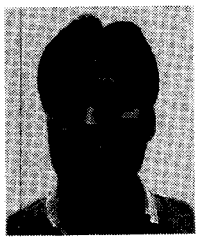
[6] Kirby, M. and Sirovich, L. : Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Trans. PAMI, 12(1), 1990.

[7] Phillips, P.J., Wechsler, H., Huang, J., and Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing Journal, 16(5), pp. 295-306, 1998.

[8] Rizvi, S., Phillips, P.J., Moon, H., : A verification protocol and statistical performance analysis for face recognition algorithms, In proceedings Computer Vision and Pattern Recognition, 1998.

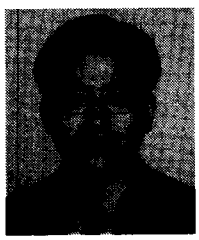
[9] Fukunaga, K. : Introduction to statistical pattern recognition, Academic Press, Orlando, FL., 1972.

[10] Swets, K.-K. and Poggio, T. : Example-based learning for view-based human face detection. IEEE Trans, PAMI, 20, pp. 39-51, 1998.



문 현 준

1990년 고려대학교 전자공학과 학사
1992년 미 State University of New York at Buffalo 석사
1999년 미 State University of New York at Buffalo 박사
1993년 삼성데이타시스템(SDS), 연구원
1996~1999년 미국방성 산하 육군연구소,선임연구원
1999~2003년 Viisage Technology(Boston),책임연구원
2003년 3월~현재 연세대학교 전기전자공학부 조교수, 생체인식연구센터(BERC)연구교수
관심분야 : 이미지처리, 컴퓨터비젼, 패턴인식, 얼굴인식을 포함한 생체인식
E-mail : hmoon@yonsei.ac.kr



김 상 훈

1987년 고려대학교 전자공학과 학사
1989년 고려대학교 대학원 전자공학과 석사
1999년 고려대학교 대학원 전자공학과 박사
2001년 UWA 방문연구원(Australia)
1989년~1994년 LG반도체 연구원
1999년~2001년 KIST 위촉연구원
1999년~현재 국립한경대학교 정보제어공학과 조교수
관심분야 : 3D 영상처리, face detection, real-time object tracking
E-mail : kimsh@hnu.hankyong.ac.kr

교 신 저 자

김 상 훈 456-749 경기도 안성시 석정동 67번지 한경대학교 정보제어공학과