The Bayes Decision Rule Induced Similarity Measures

Chengjun Liu

Abstract—This paper first shows that the popular whitened cosine similarity measure is related to the Bayes decision rule under specific assumptions and then presents two new similarity measures: the PRM Whitened Cosine (PWC) similarity measure and the Within-Class Whitened Cosine (WWC) similarity measure. Experiments on face recognition using the Face Recognition Grand Challenge (FRGC) version 2 database show the effectiveness of the new measures.

Index Terms—Face Recognition Grand Challenge (FRGC), PRM Whitened Cosine (PWC) similarity measure, whitened cosine similarity measure, Within-Class Whitened Cosine (WWC) similarity measure.

1 Introduction

SIMILARITY measures play an important role in pattern recognition and computer vision [4], [15], [8], [13], [18]. The whitened cosine similarity measure has been robustly demonstrated for superior pattern recognition performance on large scale experiments [4], [11], [14]. The recent Face Recognition Grand Challenge (FRGC) program, for example, introduces a baseline algorithm to identify challenging face recognition problems [14]. The FRGC baseline algorithm, a Principal Component Analysis (PCA) algorithm that has been optimized for large scale problems, applies the whitened cosine similarity measure for its nearest neighbor classifier [1], [14], [17].

This paper shows that the popular whitened cosine similarity measure is related to the Bayes decision rule for minimum error under some specific assumptions. In particular, the following four assumptions are able to turn the Bayes decision rule into the whitened cosine similarity decision rule:

1. The conditional probability density functions of all the classes are multivariate normal.
2. The prior probabilities of all the classes are equal.
3. The covariance matrices of all the classes are identical to the covariance matrix of all samples regardless of their class membership.
4. The whitened pattern vectors in the Bayes decision rule are normalized to unit norm.

Note that, if only the first three assumptions are satisfied, then the Bayes decision rule becomes the Mahalanobis distance decision rule, whose pattern recognition performance is usually not as good as the whitened cosine similarity decision rule.

The observation that the fourth assumption added to the first three can actually lead to a better classifier—a classifier with the whitened cosine similarity measure—motivates our exploration of new similarity measures with enhanced pattern recognition performance. Specifically, we want to keep the fourth assumption because it helps improve the discriminating power of a pattern classifier, but we need to modify the third assumption because it obscures the class membership information inherent in the Bayes decision rule. By incorporating the class membership information into the whitened cosine similarity measure, we are able to derive two new similarity measures: the PRM[1] Whitened Cosine (PWC) similarity measure and the Within-Class Whitened Cosine (WWC) similarity measure.

The performance of the proposed two new similarity measures together with some other commonly used similarity measures is assessed on a face recognition problem. Face recognition has become a very active research area in pattern recognition and computer vision driven mainly by its broad applications in human-computer interaction, homeland security, and entertainment [2], [3], [9], [6], [19], [12]. The data used in our experiments is the FRGC version 2 database [14]. The FRGC baseline algorithm shows that the FRGC Experiment 4, which is designed for controlled single still image versus uncontrolled single still image, is the most challenging FRGC problem. We therefore use FRGC Experiment 4 to assess these similarity measures. Experimental results show that the proposed two new similarity measures, the WWC and the PWC similarity measures, perform better than the other popular similarity measures, such as the whitened cosine similarity measure.

2 The Whitened Cosine Similarity Measure

This section shows that a pattern classifier based on the whitened cosine similarity measure can be derived from the Bayes classifier under specific assumptions. The whitened cosine similarity measure is among the most commonly used similarity measures in pattern recognition [4], [7]. Let δ_{WC} represent the whitened cosine similarity measure, which may be formulated as follows:

$$\delta_{WC}(\mathcal{U},\mathcal{V}) = \frac{(W^t\mathcal{U})^t(W^t\mathcal{V})}{\|W^t\mathcal{U}\| \ \|W^t\mathcal{V}\|}, \tag{1}$$

where $\mathcal{U},\mathcal{V} \in \mathbb{R}^d$ are two pattern vectors, $\|\cdot\|$ denotes the norm operator, and W is the whitening transformation matrix, which may be specified by means of the covariance matrix. The covariance matrix of all samples, $\Sigma = \mathcal{E}\{(\mathcal{X}-M_0)(\mathcal{X}-M_0)^t\}$, can be factorized using PCA: $\Sigma = \Phi\Lambda\Phi^t$, where $\mathcal{E}(\cdot)$ is the expectation operator, M_0 is the grand mean vector, Φ is an orthogonal eigenvector matrix, and Λ a diagonal eigenvalue matrix. The transformation, $W = \Phi\Lambda^{-1/2}$, is called the whitening transformation [5]. Note that, under this particular whitening transformation, (1) becomes $\delta_{WC}(\mathcal{U},\mathcal{V}) = \frac{\mathcal{U}^t\Sigma^{-1}\mathcal{V}}{\|W^t\mathcal{U}\| \ \|W^t\mathcal{V}\|}$, which is the commonly used form of the whitened cosine similarity measure.

Now, let $\mathcal{X}$ be a pattern vector in a d-dimensional space: $\mathcal{X} \in \mathbb{R}^d$ and $\mathcal{X}$ belongs to one of the predefined L classes: $\omega_1, \omega_2, \ldots, \omega_L$. The conditional probability density functions and the prior probabilities are $p(\mathcal{X}|\omega_1), p(\mathcal{X}|\omega_2), \ldots, p(\mathcal{X}|\omega_L)$ and $P(\omega_1), P(\omega_2), \ldots, P(\omega_L)$, respectively. The multiclass Bayes decision rule for minimum error may be written as follows [5]:

$$ln[p(\mathcal{X}|\omega_k)P(\omega_k)] = \max_{i=1}^{L} ln[p(\mathcal{X}|\omega_i)P(\omega_i)] \longrightarrow \mathcal{X} \in \omega_k, \tag{2}$$

which indicates that, if the natural logarithm of the product of the conditional density function of $\mathcal{X}$ given ω_k and the prior probability of ω_k is the largest among the L classes, then the pattern vector $\mathcal{X}$ is classified to ω_k. When the conditional density function, $p(\mathcal{X}|\omega_i)$, is modeled as a multivariate normal distribution with mean vector, $\mathbf{M}_i \in \mathbb{R}^d$, and covariance matrix, $\mathbf{\Sigma}_i \in \mathbb{R}^{d\times d}$,

$$p(\mathcal{X}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} exp\left\{-\frac{1}{2}(\mathcal{X}-\mathbf{M}_i)^t\mathbf{\Sigma}_i^{-1}(\mathcal{X}-\mathbf{M}_i)\right\}, \tag{3}$$

1. PRM stands for Probability Reasoning Model [10].

- The author is with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102. E-mail: chengjun.liu@njit.edu.

Manuscript received 16 June 2006; revised 29 Aug. 2006; accepted 23 Oct. 2006; published online 18 Jan. 2007.
Recommended for acceptance by B. Triggs.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0452-0606.
Digital Object Identifier no. 10.1109/TPAMI.2007.1063.

0162-8828/07/$25.00 © 2007 IEEE Published by the IEEE Computer Society

the discriminant function $\delta_i(\mathcal{X}) = ln[p(\mathcal{X}|\omega_i)P(\omega_i)]$ becomes

$$\delta_i(\mathcal{X}) = -\frac{1}{2}\{(\mathcal{X} - \mathbf{M}_i)^t \mathbf{\Sigma}_i^{-1}(\mathcal{X} - \mathbf{M}_i) + d ln(2\pi) + ln|\mathbf{\Sigma}_i|\} + ln[P(\omega_i)]. \quad (4)$$

If we assume that the prior probabilities are equal and drop the constant terms from (4) but still use $\delta_i(\mathcal{X})$ to represent the discriminant function (note that dropping constant terms from (4) does not affect the Bayes decision rule of (2)), then (4) is simplified as follows:

$$\delta_i(\mathcal{X}) = -\frac{1}{2}\{(\mathcal{X} - \mathbf{M}_i)^t \mathbf{\Sigma}_i^{-1}(\mathcal{X} - \mathbf{M}_i) + ln|\mathbf{\Sigma}_i|\}. \quad (5)$$

Note that the discriminant function of (5) is valid under two assumptions: 1) The conditional probability density functions, $p(\mathcal{X}|\omega_i)$, are multivariate normal and 2) the prior probabilities, $P(\omega_i)$, are all equal. Now, we further assume that 3) the covariance matrices of the L classes, $\mathbf{\Sigma}_i$, are identical to the covariance matrix of all samples, regardless of their class membership: $\mathbf{\Sigma}_i = \Sigma$. Under these assumptions, the discriminant function of (5) is simplified as follows (again dropping the constant term $ln|\mathbf{\Sigma}|$ does not affect the Bayes decision rule of (2)):

$$\delta_i(\mathcal{X}) = -\frac{1}{2}(\mathcal{X} - \mathbf{M}_i)^t \Sigma^{-1}(\mathcal{X} - \mathbf{M}_i). \quad (6)$$

Equation (6) shows that, under the three assumptions, the Bayes decision rule for minimum error of (2) turns into the Mahalanobis distance decision rule:

$$\delta_{Md}(\mathcal{X}, \mathbf{M}_k) = \min_{i=1}^{L} \delta_{Md}(\mathcal{X}, \mathbf{M}_i) \longrightarrow \mathcal{X} \in \omega_k, \quad (7)$$

where $\delta_{Md}(\mathcal{X}, \mathbf{M}_i) = \left[(\mathcal{X} - \mathbf{M}_i)^t \Sigma^{-1}(\mathcal{X} - \mathbf{M}_i)\right]^{1/2}$ is the Mahalanobis distance measure.

By rearranging terms, (6) may be rewritten as follows:

$$\delta_i(\mathcal{X}) = -\frac{1}{2}\left\{\left[W^t(\mathcal{X} - \mathbf{M}_i)\right]^t\left[W^t(\mathcal{X} - \mathbf{M}_i)\right]\right\}, \quad (8)$$

where W is the whitening transformation matrix: $W = \Phi\Lambda^{-1/2}$ and $\Sigma = \Phi\Lambda\Phi^t$. Equation (8) thus further reveals the connection between the whitened cosine similarity measure and the Bayes discriminant function:

$$\delta_i(\mathcal{X}) = -\frac{1}{2}\left\{\|W^t\mathcal{X}\|^2 + \|W^t\mathbf{M}_i\|^2 - 2\|W^t\mathcal{X}\|\|W^t\mathbf{M}_i\|\delta_{WC}(\mathcal{X}, \mathbf{M}_i)\right\}. \quad (9)$$

Now, if we add another assumption, 4) the whitened pattern vectors, $W^t\mathcal{X}$ and $W^t\mathbf{M}_i$, are normalized to unit norm, then $\delta_i(\mathcal{X}) = \delta_{WC}(\mathcal{X}, \mathbf{M}_i) - 1$ and the Bayes decision rule of (2) becomes the whitened cosine similarity decision rule:

$$\delta_{WC}(\mathcal{X}, \mathbf{M}_k) = \max_{i=1}^{L} \delta_{WC}(\mathcal{X}, \mathbf{M}_i) \longrightarrow \mathcal{X} \in \omega_k. \quad (10)$$

Note that the connection between the whitened cosine similarity measure and the Bayes discriminant function is established by (9), but usually the whitened pattern vectors, $W^t\mathcal{X}$ and $W^t\mathbf{M}_i$, do not have unit norm. That is why the Bayes classifier (corresponding to the Mahalanobis distance decision rule of (7) under the first three assumptions) and the whitened cosine similarity decision rule of (10) display different pattern recognition performance: The latter usually performs better than the former, i.e., the whitened cosine similarity measure usually outperforms the Mahalanobis distance measure. The likely reason that the Bayes classifier does not achieve its optimal pattern classification performance is because of the first three assumptions, which are oversimplified and do not represent the true nature (the statistical distribution) of the data. An interesting observation is that the fourth assumption added to the first three ones can actually lead to a better classifier—a classifier with the whitened cosine similarity measure. Next, we will keep the fourth assumption and modify the third one to derive new similarity measures with enhanced pattern recognition performance.

3 The PRM Whitened Cosine (PWC) Similarity Measure

The third assumption in establishing the connection between the whitened cosine similarity measure and the Bayes decision rule requires that the covariance matrix of every class be identical to the covariance matrix of all samples. While this assumption simplifies the computation of the covariance matrix in concept (computing one covariance matrix rather than L different covariance matrices), it compromises the discriminating power of the whitened cosine similarity measure by discarding the valuable class membership information. The Bayes decision rule for minimum error, however, shows the importance of the class membership information in terms of different covariance matrices for the L classes. Therefore, in order to enhance pattern classification performance, we propose a new similarity measure that considers both the simplicity of computation and the class membership information.

First, we still assume that the covariance matrix of every class is identical to the covariance matrix Σ of all samples and Σ is factorized via PCA: $\Sigma = \Phi\Lambda\Phi^t$. Then, we replace the diagonal eigenvalue matrix Λ with another diagonal matrix Δ, whose ith diagonal element is derived as follows: 1) compute the variance of the ith principal components for every class and 2) average the L variances from all the classes. In fact, the derivation of Δ resembles the Probability Reasoning Model (PRM) [10]:

$$\Delta = diag\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2\}, \quad (11)$$

where the component σ_i^2 is calculated as follows [10]:

$$\sigma_i^2 = \frac{1}{L}\sum_{k=1}^{L}\left\{\frac{1}{N_k - 1}\sum_{j=1}^{N_k}\left(y_{ji}^{(k)} - m_{ki}\right)^2\right\}, \quad (12)$$

where N_k is the number of samples in class ω_k, $y_{ji}^{(k)}$ is the ith element of the projected vector $\Phi^t\mathcal{X}_j^{(k)}$ of class ω_k, and m_{ki} is the ith element of the projected mean vector of class ω_k: $\mathcal{E}(\Phi^t\mathcal{X}|\omega_k)$.

Next, we define a new whitening transformation matrix:

$$W_p = \Phi\Delta^{-1/2}, \quad (13)$$

where Φ is from the factorization of the covariance matrix of all samples ($\Sigma = \Phi\Lambda\Phi^t$) and Δ is from (11) and (12). Finally, we define a new similarity measure as follows:

$$\delta_{PWC}(\mathcal{U}, \mathcal{V}) = \frac{(W_p^t\mathcal{U})^t(W_p^t\mathcal{V})}{\|W_p^t\mathcal{U}\| \; \|W_p^t\mathcal{V}\|}. \quad (14)$$

We call this new similarity measure the PRM Whitened Cosine (PWC) similarity measure.

4 The Within-Class Whitened Cosine (WWC) Similarity Measure

The advantage of the PRM whitened cosine, or the PWC, similarity measure over the whitened cosine similarity measure is due to the incorporation of the class membership information: Even though the assumption of the PWC similarity measure still requires that the covariance matrix of every class be identical to the covariance matrix of all samples, the eigenvalues of the covariance matrix are replaced with the average variances from all the L classes estimated in the PCA space. However, the class membership information in the Bayes decision rule for minimum error is determined by the covariance matrices of all L classes (see (5)). These L covariance matrices, $\mathbf{\Sigma}_i$, collectively define the within-class scatter matrix [5]:

TABLE 1
Number of Images and Image Quality (Controlled, Uncontrolled) of Training, Target, and Query Sets of the FRGC Version 2 Database

Data	Experiment	Set	# Images	Image Quality
FRGC Version 2	Experiment 4	Training	12,776	Controlled & Uncontrolled
		Target	16,028	Controlled
		Query	8,014	Uncontrolled

Fig. 1. Example FRGC images cropped to the size of 128×128. The first row shows two controlled training images (the first two images) and two uncontrolled training images. The second row shows one controlled target image (the first image) and three uncontrolled query images.

$$S_w = \sum_{i=1}^{L} P(\omega_i)\mathcal{E}\{(\mathcal{X} - M_i)(\mathcal{X} - M_i)^t|\omega_i\} = \sum_{i=1}^{L} P(\omega_i)\mathbf{\Sigma}_i. \quad (15)$$

Note that, when the prior probabilities are assumed to be equal, the within-class scatter matrix is just the average of the L class-dependent covariance matrices: $S_w = \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Sigma}_i$. But, the covariance matrix of all samples, Σ, consists of both the within-class scatter matrix, S_w (see (15)), and the between-class scatter matrix [5]. In order to enhance pattern recognition performance, a new similarity measure should model $\mathbf{\Sigma}_i$ in the Bayes decision rule using the within-class scatter matrix, rather than the mixture scatter matrix, i.e., the covariance matrix of all samples.

Now, we assume that the covariance matrix of every class $\mathbf{\Sigma}_i$ in the Bayes decision rule is identical to the within-class scatter matrix, S_w:

$$\mathbf{\Sigma}_i = S_w = \sum_{i=1}^{L} P(\omega_i)\mathcal{E}\{(\mathcal{X} - M_i)(\mathcal{X} - M_i)^t|\omega_i\}. \quad (16)$$

Then, we factorize this within-class scatter matrix, S_w, into the following form using PCA [5]:

$$S_w = \Phi_w \Lambda_w \Phi_w^t. \quad (17)$$

Finally, we define a new whitening transformation as follows:

$$W_w = \Phi_w \Lambda_w^{-1/2}. \quad (18)$$

Based on this new whitening transformation, we define a new similarity measure as follows:

$$\delta_{WWC}(\mathcal{U}, \mathcal{V}) = \frac{(W_w^t\mathcal{U})^t(W_w^t\mathcal{V})}{\|W_w^t\mathcal{U}\| \ \|W_w^t\mathcal{V}\|}. \quad (19)$$

We call this new similarity measure the Within-class Whitened Cosine (WWC) similarity measure.

5 Experiments

The section assesses the whitened cosine similarity measure, δ_{WC}, the PWC similarity measure, δ_{PWC}, and the WWC similarity measure, δ_{WWC} on a face recognition task using the FRGC version 2 database [14]. The FRGC baseline algorithm, which is a PCA algorithm that applies the whitened cosine distance measure for its nearest neighbor classifier, reveals that the FRGC Experiment 4, which is designed for controlled single still image versus uncontrolled single still image, is the most challenging FRGC experiment. We therefore choose FRGC Experiment 4 to assess these similarity measures. Table 1 shows Training, Target, and Query sets of the FRGC version 2 database for Experiment 4. The Training set contains 12,776 images, which are either controlled or uncontrolled. The Target set contains 16,028 controlled images and the Query set has 8,014 uncontrolled images. While the faces in the controlled images have good image resolution and good illumination, the faces in the uncontrolled images have lower image resolution and larger illumination variations. It is these uncontrolled factors that pose grand challenges to face recognition performance.

As the FRGC data consists of high resolution still images, an image preprocessing procedure is applied to extract the facial region. Image preprocessing includes the following steps: First, manual annotation detects the centers of the eyes; second, rotation and scaling transformations align the centers of the eyes to the predefined locations with a fixed interocular distance; finally, a subimage procedure crops the face image to the size of 128×128 to extract the facial region. The facial region thus contains only face and the performance of face recognition is not affected by the factors not related to face, such as hair styles. Fig. 1 shows some example FRGC images used in our experiments that are already cropped to the size of 128×128 to extract the facial region. In particular, the first row displays four training images: two controlled images (the first two images) and two uncontrolled images (the remaining two images). The second row shows a target image (the first image, which is

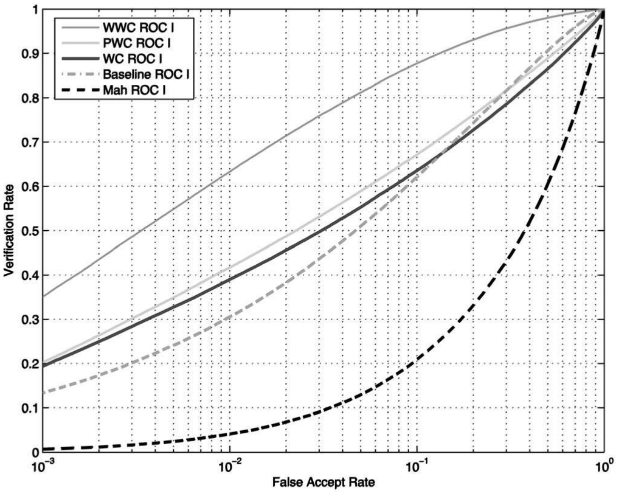

Fig. 2. FRGC version 2 Experiment 4 face recognition performance (the ROC I curves) using the WWC similarity measure (WWC ROC I), the PWC similarity measure (PWC ROC I), the whitened cosine similarity measure (WC ROC I), and the Mahalanobis distance measure (Mah ROC I), respectively. The FRGC baseline performance (Baseline ROC I) using gray-scale images is also included for comparison.

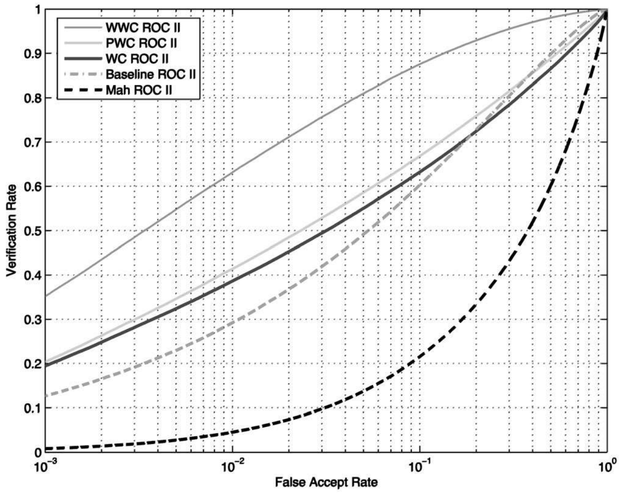

Fig. 3. FRGC version 2 Experiment 4 face recognition performance (the ROC II curves) using the WWC similarity measure (WWC ROC II), the PWC similarity measure (PWC ROC II), the whitened cosine similarity measure (WC ROC II), and the Mahalanobis distance measure (Mah ROC II), respectively. The FRGC baseline performance (Baseline ROC II) using gray-scale images is also included for comparison.

controlled) and three query images (the remaining three images, which are uncontrolled).

FRGC version 2 Experiment 4 contains Training, Target, and Query sets and face recognition performance is reported using the Receiver Operating Characteristic (ROC) curves, which plot the Face Verification Rate (FVR) versus the False Accept Rate (FAR). The ROC curves are generated by the Biometric Experimentation Environment (BEE), when a similarity matrix is provided to the system. The similarity matrix stores the similarity score of every target versus query image pair. As a result, the size of the similarity matrix is $T \times Q$, i.e., the number of target images T times the number of query images Q ($T = 16,028$ and $Q = 8,014$ for FRGC version 2 Experiment 4). Let $\mathcal{T}_i$ be a target pattern vector, $i = 1, 2, \ldots, T$, and let $\mathcal{Q}_j$, $j = 1, 2, \ldots, Q$, be a query pattern vector, then the similarity score $\mathcal{S}(\mathcal{T}_i, \mathcal{Q}_j)$ based on similarity measure δ is calculated as follows:

$$\mathcal{S}(\mathcal{T}_i, \mathcal{Q}_j) = \delta(\mathcal{T}_i, \mathcal{Q}_j). \tag{20}$$

The similarity measure δ can be the WWC similarity measure, δ_{WWC}, the PWC similarity measure, δ_{PWC}, or the whitened cosine similarity measure, δ_{WC}.

Our previous research shows that the R component image carries more discriminating information than the gray-scale image converted from the RGB image [16]. We therefore keep only the R component images (and discard the G and B component images) from the FRGC jpeg images in the Training, Target, and Query sets. The image preprocessing steps described before are applied to these R component images to extract 128×128 face images. Traditional image processing operations, such as image filtering for getting rid of noise and histogram processing for illumination enhancement, are applied to improve image quality.

The BEE system generates three ROC curves (ROC I, ROC II, and ROC III) corresponding to the images collected within semesters, within a year, and between semesters, respectively. Figs. 2, 3, and 4 show the ROC I, ROC II, and ROC III curves of FRGC version 2 Experiment 4 face recognition performance using the WWC similarity measure, δ_{WWC}, the PWC similarity measure, δ_{PWC}, the whitened cosine similarity measure, δ_{WC}, and the Mahalanobis distance measure, respectively. The FRGC baseline performance using gray-scale images is also included in these figures for comparison. All three of these figures show that the two new similarity measures achieve better face recognition performance than the popular whitened cosine similarity measure and the Mahalanobis distance measure and the R component image helps improve face recognition performance when compared against the gray-scale image. In particular, Fig. 4 shows that the WWC similarity measure achieves the face verification rate (ROC III) of 35 percent at the false accept rate of 0.1 percent, compared to the PWC similarity measure with the face verification rate of 20 percent, the whitened cosine similarity measure with the face verification rate of 19 percent, the BEE baseline verification rate of 12 percent (using gray-scale images), and the Mahalanobis distance measure with the face verification rate of 1 percent, at the same false accept rate.

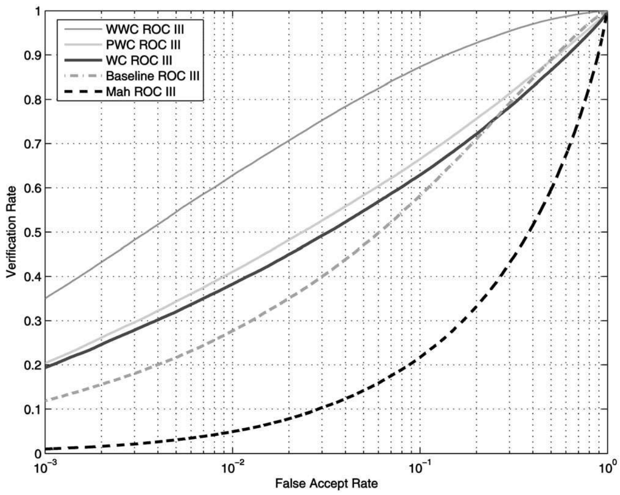

Fig. 4. FRGC version 2 Experiment 4 face recognition performance (the ROC III curves) using the WWC similarity measure (WWC ROC III), the PWC similarity measure (PWC ROC III), the whitened cosine similarity measure (WC ROC III), and the Mahalanobis distance measure (Mah ROC III), respectively. The FRGC baseline performance (Baseline ROC III) using gray-scale images is also included for comparison.

6 CONCLUSIONS

This paper shows that the popular whitened cosine similarity measure is related to the Bayes decision rule under the following four assumptions:

1. The conditional probability density functions of all the classes are multivariate normal.
2. The prior probabilities of all the classes are equal.
3. The covariance matrices of all the classes are identical to the covariance matrix of all samples regardless of their class membership.
4. The whitened pattern vectors in the Bayes decision rule are normalized to unit norm.

By modifying the third assumption to incorporate the class membership information that is inherent in the Bayes decision rule, we are able to develop two new similarity measures with enhanced pattern recognition performance: the PRM Whitened Cosine (PWC) similarity measure and the Within-Class Whitened Cosine (WWC) similarity measure. Experiments using the FRGC version 2 database show that the proposed two new similarity measures, the WWC and the PWC similarity measures, perform better than the other popular similarity measures, such as the whitened cosine similarity measure. Future work will consider incorporating the new similarity measures into complex pattern recognition systems to improve the overall pattern recognition performance.

ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work was partially supported by the NJCST/NJIT grant and by Award No. 2006-IJ-CX-K033 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect those of the US Department of Justice.

REFERENCES

[1] J.R. Beveridge, D. Bolme, B.A. Draper, and M. Teixeira, "The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure," *Machine Vision and Applications*, vol. 16, no. 2, pp. 128-138, 2005.

[2] K.W. Bowyer, K. Chang, and P.J. Flynn, "A Survey of Approaches and Challenges in 3D and Multimodal 3D+2D Face Recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1-15, 2006.

[3] K.W. Bowyer, K.I. Chang, P. Yan, P.J. Flynn, E. Hansley, and S. Sarkar, "Multi-Modal Biometrics: An Overview," *Proc. Second Workshop MultiModal User Authentication*, 2006.

[4] K.W. Bowyer and P.J. Phillips, *Empirical Evaluation Techniques in Computer Vision*. IEEE CS Press, 1998.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed. Academic Press, 1990.

[6] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross, "Biometrics: A Grand Challenge," *Proc. 17th Int'l Conf. Pattern Recognition*, pp. 935-942, 2004.

[7] B.V.K.V. Kumar, M. Savvides, K. Venkataramani, and C. Xie, "Spatial Frequency Domain Image Processing for Biometric Recognition," *Proc. IEEE Int'l Conf. Image Processing*, 2002.

[8] G. Lebanon, "Metric Learning for Text Documents," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 497-508, Apr. 2006.

[9] C. Liu, "Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725-737, May 2006.

[10] C. Liu and H. Wechsler, "Robust Coding Schemes for Indexing and Retrieval from Large Face Databases," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 132-137, 2000.

[11] H. Moon and P.J. Phillips, "Computational and Performance Aspects of PCA-Based Face Recognition Algorithms," *Perception*, vol. 30, pp. 303-321, 2001.

[12] S. Pankanti, R.M. Bolle, and A. Jain, "Guest Editors' Introduction: Biometrics the Future of Identification," *Computer*, vol. 33, no. 2, pp. 46-49, Feb. 2000.

[13] R. Paredes and E. Vidal, "Learning Weighted Metrics to Minimize Nearest-Neighbor Classification Error," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1100-1110, July 2006.

[14] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[15] N. Sebe, M.S. Lew, and D.P. Huijsmans, "Toward Improved Ranking Metrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1132-1143, Oct. 2000.

[16] P. Shih and C. Liu, "Comparative Assessment of Content-Based Face Image Retrieval in Different Color Spaces," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 873-893, 2005.

[17] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 13, no. 1, pp. 71-86, 1991.

[18] H. Wang, "Nearest Neighbors by Neighborhood Counting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 942-953, July 2006.

[19] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.