Discriminability and Reliability Indexes: Two New Measures to Enhance Multi-image Face Recognition

Weiwen Zou

Abstract

In order to handle complex face image variations in face recognition, multi-image face recognition has been proposed, instead of using a single still-image-based approach. In many practical scenarios, multiple images can be easily obtained in enrollment and query stages, for example, using video. By accessing these images, a good "quality" image(s) will be used for recognition using conventional still-image-based recognition algorithms so that the recognition performance can be improved. However, existing methods do not fully utilize all images information. In this paper, two new measurements, namely discriminability index (DI) and reliability index (RI), are proposed to evaluate the enrolled and query images respectively. By considering the distribution of enrolled images from individuals, the discriminability index of each image is calculated and a weight is assigned. For testing images, a reliability index is calculated based on matching quality between the testing image and enrolled images. If the reliability index of a testing image is small, the testing image will be discarded as it may degrade the recognition performance. To evaluate and demonstrate the use of DI and RI, we adopt the recognition algorithm using combining classifiers with eigenface representations in input and kernel spaces. CMU-PIE , YaleB and FRGC databases are used for experiments. Experimental results show that the recognition performance, with three popular combination rules, can be increased by more than 10% on average with the use of DI and RI.

1 Introduction and Related Works

Research on face recognition has been performed for more than three decades and a number of encouraging results have been reported [18, 23]. However, some hard

---

problems, such as outdoor illumination, still remain unsolved [12, 20, 23].

To overcome the remaining problems, a natural move is to change from single-frame image-based face recognition to multi-image-based recognition [3, 4, 5, 10, 21] as multiples frames in both enrollment and query phases provide additional information and can be easily obtained, for example, using video. In turn, multi-image face recognition has been proposed for solving the image variations problem. The common practice in multi-image recognition is to select a good visual quality frontal-view face image for recognition. While experimental results show that the recognition performance can be improved [16], non-frontal view good quality images, which can also be used to enhance the recognition performance, are discarded. That means, the available useful information are not fully utilized. To make use of multiple information, multi-image-based recognition methods were proposed.

Kruger *et al.* [15] proposed a new method to select frames from video. By applying an online-clustering algorithm on video, exemplar (representative frame) from each cluster is used to represent the cluster and used it as reference image in recognition system. Based on the cluster size, a weight (importance) is assigned to each representative frame. Hadid *et al.* [10] selected the representative images from video for recognition. They minimized distance between original images and the images to be selected as representative images, so that the most representative images are selected for recognition. However, representative images/exemplars do not guarantee they have high discriminability power.

Zhang *et al.* [22] assigned weights to testing images by evaluating similarity between the testing images and reference images based on the pose and facial expression. Thomas *et al.*[21] used a measurement called *Faceness* to assess the quality of the images. Based on Faceness, the images were selected in several selecting strategies, including N highest Faceness (NHF) and N evenly spaced from M highest faceness (NEHF). Experimental results showed that using the N highest faceness images did not get the best performance. Instead, using multi-images with ensuring face-

ness diversity got the best result. Therefore, evaluating and selecting enrolled and testing images is an important task for multiple images face recognition system.

To overcome the existing problems in evaluating the enrolled and testing images, two new measurements are proposed in this paper. First, a discriminability index (DI) is proposed to measure the discriminative power of enrolled images in each individual. Second, a reliability index (RI) is proposed to measure the matching quality which reflects how good a face image can be used for face recognition. It is important to point out that both DI and RI are generic and can be integrated into most of the existing image-based algorithms.

The rest of this paper is organized as follows. Section 2 reports the proposed method in calculating the discriminability index for enrolled image and reliability index for testing image. Experimental evaluation on the two proposed indexes are given in Section 3. Finally, Section 4 gives the conclusions.

2 Proposed method

Consider the case that there are multiple images in probe ($\mathcal{P}$) and multiple images in gallery ($\mathcal{G}$). This paper proposes two new image measurements to evaluate how good the images are. A weight is then assigned to each image in probe and gallery set. The weight for both reference and testing images can be taken into account in the recognition process. It is important to note that there may have different ways to make use of DI and RI. In the experimental result section, we will demonstrate one of the ways of using DI and RI to improve the performance of multi-image face recognition system.

For multiple images in gallery set, discriminability index (DI) is developed to measure how discriminative an reference image is. A high discriminative (i.e. high DI) reference image means that such image is relatively far away from the classes boundary and has a relatively high tolerance to the intra-class variations. Then, a larger weight will be assigned to those reference images with high DI. Similarly, if a reference image is very close to the images in other classes, it will be easily affected by the intra-class variations. In turn, a smaller weight will be assigned.

For multiple images in probe set (testing set), reliability index (RI) is developed to measure how reliable the image is. A testing image with higher RI means that we have a higher confidence to treat the result is correct. Unlike existing schemes [15, 10, 21] selecting the representative images based on the face appearance, our proposed method calculates the RI by considering the matching quality of the image. This paper considers the distances between the testing image and the images in each class, as well as the distributions of the reference images in each class. The basic idea
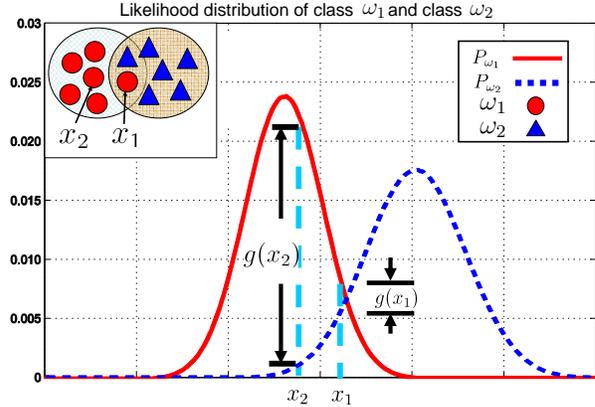

Figure 2. The distribution of ω_1 **and** ω_2**.** x_1 **and** x_2 **are two reference image in** ω_1**.**

of DI and RI is shown in Figure 1. Details in calculating DI and RI are given as follows.

Some useful notations are used in this paper: $\mathcal{P}$ represents the probe set while $\mathcal{G}$ represents the gallery set (reference images), $\mathcal{T}$ represents the training set; the small letters x, y, z represent still images; ω_i represents i-th class of the database.

2.1 Discriminability index (DI)

Discriminability index (DI) measures how much discriminative power of a reference image has. Since there is no generic definition of discriminative power of an image, this paper defines the discriminative power as the ability of an image distinguishing from images from other classes. Therefore, an image which is far away from the classes' boundary will have a high discriminative power. This discriminative power depends on two factors, namely the margin between a reference image and images from other classes, and the tolerance to intra-class variants which are caused by pose, illuminance, expression etc.

To clearly present and illustrate our idea on DI, we consider a two-class problem. As shown in Figure 2, x_1 and x_2 are two reference images which belong to the same class, ω_1. The distributions of ω_1 and ω_2 are overlapped. In practice, the classes distribution is always overlapped with each other due to intra-class variations, such as pose, illuminance, expression, occlusion. So the reference images will have different discriminative powers. Consider the probability likelihood $P(x_1|\omega_1)$, $P(x_1|\omega_2)$, $P(x_2|\omega_1)$ and $P(x_2|\omega_2)$, according to the discriminant function [7]

$$g(x) = P(x|\omega_1) - P(x|\omega_2) \qquad (1)$$

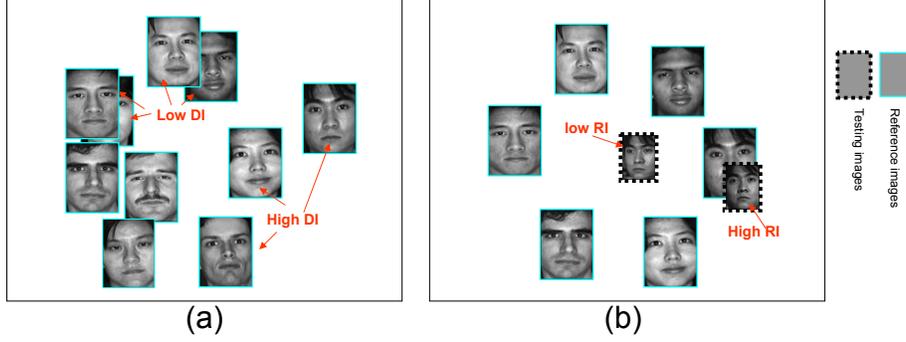we get $g(x_2) > g(x_1)$. This implies that x_2 is better than x_1

Figure 1. Demonstrating the idea of DI and RI. (a) **Ten images from difference classes. Some of them overlap with others (low DI) while some are far away from others (high DI).** (b) **Two testing images matching with the reference images, one only has a short distance to the reference images (high RI) while another has similar distances to each class (low RI).**

as a reference image. The reason is that if a testing image is close to x_2, it has a large gap (far away) from ω_2. That means we have high confidence to classify it to ω_1 but not ω_2. In turn, x_2 has a higher discriminative power. On the other hand, if a testing image is close to x_1, a noisy image of ω_2 may also be very close to x_1 due to relatively small $g(x_2)$. This leads the testing image be misclassified from ω_2 to ω_1. We have low confidence to classify x_1. x_1 has a relatively low discriminative power. Therefore, it is necessary to evaluate the discriminative power (discriminability index) of the enrolled is necessary.

A straightforward method to calculated the discriminability index (DI) is to calculate the involved likelihood probability. However, it is well-known that the likelihood distribution of the face images is not easy to model due to the high dimension of data and relatively small amount of training samples. Inspired by the success of calculating the probability from similarity [2], we estimate likelihood probability for DI calculation based on the similarity. According to Parzen window method, the probability can be estimated by:

$$P(x|\omega) = \frac{1}{n}\sum_{j=1}^{n}(\varphi_\omega(x, y_i)) \qquad (2)$$

where y_i is the i-th neighbor of x in ω. Define the similarity between two images as

$$sim(x, y) = \frac{C_1 - \|x - y\|}{C_1} \qquad (3)$$

where C_1 is a constant to normalize the similarity ranging from 0 to 1. Replace φ with Eq.(3), we have

$$P(x|\omega) = \frac{1}{nC_2}\sum_{j=1}^{n}(sim(x, y_i)). \qquad (4)$$

where C_2 is a normalization constant. So DI can be defined as

$$
\begin{aligned}
\text{DI}_x &= g(x) \\
&= P(x|x \in \omega) - \max_k P(x|\omega_k) \\
&= \frac{1}{nC_2}\min_k(\sum_{i=1}^{n}(\frac{C_1 - \|x - y_i\|}{C_1})) - \sum_{j=1}^{n}(\frac{C_1 - \|x - y_j^k\|}{C_1}))) \\
&= C\min_k\sum_{i=1}^{n}(\|x - y_i^k\| - \|x - y_i\|)
\end{aligned}
$$
$$(5)$$

where ω_k is the class which does not contain x_i, and $y_j^k \in \omega_k$ while $x, y_i \in \omega$, and $C = \frac{1}{nC_1C_2}$.

To alleviate influence of noise, Eq.(5) is performed several times with random selection of reference images of each class in the training set, and the average of DI will be considered as the final DI. Details of DI calculation is shown in Algorithm 1.

2.2 Reliability index (RI)

It is well-known that, in most of the cases, class distributions are overlapped. If testing images are in the overlapping area, they are not suitable for the using in recognition system. A straightforward solution for this problem is to check whether the testing image is suitable before using it for recognition. To evaluate how suitable the testing image is for recognition, matching quality is considered in this paper.

Matching quality reflects the relationship between the testing image and the reference images. High matching quality implies that such testing has a short distance to its

Algorithm 1 DI Calculation
___
Input: Reference images $\mathcal{G} = \{x_{ij}\}$, threshold for termination th, number of neighbors n, number of images to be selected N
Output: DI of each image in $\mathcal{G}$
Initial: $\text{DI}_{ij} \leftarrow 0, \text{DI}'_{ij} \leftarrow 0, \text{DI}^*_{ij} \leftarrow \phi$;
  repeat
     $\text{DI}'_{ij} \leftarrow \text{DI}_{ij}$
     for each x_{ij} **do**
        randomly select N reference images from each class, denote as ω'_k
        search the n-nearest neighbors in each ω'_k, denote as $\mathcal{N}^k_{x_{ij}}$
        $\text{DI}_{ij} \leftarrow \text{DI}^*_{ij} \bigcup \{\frac{1}{n} \min_k \sum_{j=1}^n (\|x_{ij} - y^i_j\| - \|x_{ij} - y^k_j\|)\}$
     end for
     $\text{DI}_{ij} \leftarrow avg(\text{DI}^*_{ij})$
  until $|\text{DI}'_{ij} - \text{DI}_{ij}| < th$
  Output DI_{ij}
___

own class and has relatively long distances to other classes. Figure 3 (a) shows two testing images p_1 and p_2 which have different matching quality from the same class. The distance between p_1 and ω_1 is small, while the distances between p_2 and ω_1, ω_2 and ω_3 are similar. For p_2, it locates in the overlapping area and its location causes ambiguity for classification, because a noisy image from other class (suppose ω_2) may locate at the same position of p_2. It should not be used for recognition. So p_1 has a higher matching quality than that of p_2, therefore it is more suitable for recognition. In this paper, reliability index (RI) is developed to evaluate the matching quality of testing images.
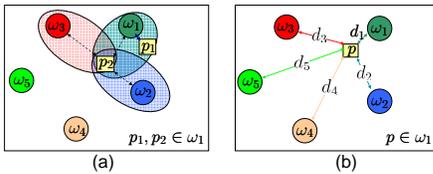

Figure 3. Matching quality and outlier. (a): Testing image p_1 has high matching quality while p_2 has low matching quality; (b)two kind of distances:$d_1 \in D_w$ and $d_2, d_3, d_4, d_5 \in D_b$. p has high matching quality if and only if d_1 is the only statistics outlier of D_b.

Based on the analysis above, the image with high matching quality has high discriminability which we develop in the Section 2.1. Since the class information of the testing images is not available, measuring matching quality by dis-

criminability index is not suitable. Instead we use another cue which can access the image matching quality.

Considering the distances between the testing image (p) and the reference images ($\{x_{kj}\}_{k=1}^C$), define d_k, D_w, D_b as follows:

$$d_k = \arg \min_j \|p - x_{kj}\|$$
$$D_w = \{d|d = x' - x''\}$$
$$D_b = \{d|d = y' - y''\}$$
(6)

where x' and x'' are images from the same class, while y' and y'' are images from different classes respectively. Suppose testing image p is from ω_{k_0}. When matching a testing image p to the reference images $\{x_{kj}\}_{k=1}^C$, there are two kinds of distances, namely d_{k_0} and $d_k, k \neq k_0$. It is easy to shown that

$$d_{k_0} \in D_w \text{ and } d_k \in D_b(k \neq k_0)$$
(7)

If p have high matching quality, d_{k_0} is obviously less than other distances ($d_k, k \neq k_0$). Specifically, d_{k_0} should be the only one distance which is less than others significantly, if p has a high matching quality. As shown in Figure 3(b), $p \in \omega_1$ has high matching quality and d_1 is much smaller than that of d_2 to d_5. So the testing image with high matching quality leads d_{k_0} deviating from D_b. That means d_{k_0} is a statistics outlier [1] of D_b. Here, following the definition of Grubbs [9], a sample is called statistics outlier of a given set of samples, if and only if it appears to deviate markedly from other samples. Therefore, evaluating the matching quality of a testing image is equivalent to evaluating the probability that the corresponding d_{k_0} is the only statistics outlier of D_b.

However, as mentioned before, the class information of testing image p is not available. That means d_{k_0} cannot be determined when p is given. Instead, we propose to make use of d_{min} as a substitute for d_{k_0}, where $d_{min} = \min_k(d_k)$. If p has a high matching quality, d_{min} is equal to d_{k_0}, so d_{min} is the only outlier and replacing d_{k_0} with d_{min} is suitable. In contrary, if p has a low matching quality, there are two scenarios:
1) d_{min} is equal to d_{k_0},
2) d_{min} is less than d_{k_0}.
For 1), d_{min} is not the only statistics outlier of D_b, because p has similar distances to images from ω_{k_0} as well as other classes. For 2), d_{min} definitely is not an outlier of D_b, because d_{min} belongs to D_b. In both scenarios, d_{min} is not the only outlier of D_b. Therefore, p has a high matching quality (high reliability) if and only if d_{min} is the only significant outlier of D_b.

To evaluate whether d_{min} is the only statistics outlier, Q-test [6] is adopted. Though a number of algorithms have been proposed to detect outliers from a set of data [1, 11]

, Q-test is simpler and more suitable. That is because Q-test can handle scenario with only one outlier existing more effectively, while other existing methods, such as Grubb test [9] and t-test [13], consider the extrema as well as other samples as the candidate outliers. Moreover, Q-test is easy to implement and computationally efficient, as it dose not require calculation of the mean and standard deviation in advance.

In this paper, RI of p is defined as the Q-value [6] of the extreme (d_{min}) in Q-test, as follows:

$$
\begin{aligned}
\text{RI} &= \text{Q-value}(d_{min}) \\
&= \frac{\text{Gap}}{\text{Range}} \\
&= \frac{d_{min} - \min_k(\{d_k\} \setminus d_{min})}{\min\{d_k\} - \max\{d_k\}}
\end{aligned}
\tag{8}
$$

Detailed calculation of RI is shown in Algorithm 2.

Algorithm 2 RI Calculation

Input: Reference images $\{x_{kj}\}_{k=1}^{C}$, testing image y_i
Output: RI_i
    for each class k **do**
        $d_k \leftarrow \|y_i - x_{kj}\|$
    end for
    Calculate the Q-value $Q \leftarrow$ by Eq.[8]
    output $\text{RI}_i = \text{Q-value}$

---

3 Experimental Results

This section demonstrates and evaluates the use of DI and/or RI with three public domain face databases. The results are presented into four parts. First, evaluation methodology of the comparative experiments are shown. Second, the use of DI and RI is reported. Third, three databases used in the experiments and the corresponding settings are discussed. Finally the experimental results are given. The results show that the proposed DI/RI can help the recognition system improving the performance.

3.1 Evaluation methodology

To evaluate the effectiveness of DI and RI, comparative experiments are performed based on the existing face recognition system integrating with and without DI and/or RI. As shown in Figure 4, the system consists of two independent modules, namely assessment module and recognition module.

In the assessment module, training images ($\{x_{kj}\}_{k=1}^{C}$) are assessed with proposed DI by Algorithm 1, while testing images ($\{p_i\}_{i=1}^{N_p}$) are assessed with RI by Algorithm 2. After these two assessments, DI of reference images and RI of testing images are obtain and then integrated with the existing face recognition in the way as mentioned in Section 3.2.

In the recognition module, two popular existing face recognition engines (namely Eigenface, and Kernel PCA) with combining classifiers are used to construct the recognition module. The face recognition engines match p_i to the reference images ($\{x_{kj}\}$), and output the similarity s_{ik} which is calculated by Eq.(9):

$$
s_{ik} = \max_j \left(S(p_i, x_{kj}) \right)
\tag{9}
$$

where $S(p_i, x_{kj})$ is the similarity calculated by the face recognition engine. Combining classifiers is used to draw the final result with multiple testing images. In this paper, three typical combining classifiers [14] are used, namely sum rule (SUM), product rule (PROD) and majority voting (MV). We do not make use of max/min rule, because max/min rule actually combine the results based on one maximal/minimal data rather than making use of all data.

3.2 Usage of DI and/or RI

This paper suggests one way to make use of DI and/or RI for face recognition. It is important to mention that there may have other better way to adopt DI/RI.

Based on determined DI/RI of each image, weights are assigned to reference images and testing images. For DI, a reference image x_{kj} is assigned with a weight which is defined as:

$$
w_{kj} = \left(\frac{1 + \text{DI}_{kj}}{2} \right)
\tag{10}
$$

So w_{kj} is normalized to [0 1]. For RI, the testing images with low RI are discarded by assigning a weight of 0, while the ones with high RI are kept by assigning a weight of 1:

$$
w_i' = \begin{cases} 1 & \text{if } \text{RI}_i > th, \\ 0 & \text{otherise.} \end{cases}
\tag{11}
$$

where th is a threshold determined by the Quotient Critical Value [6] of Q-test. In our experiments, we accept the testing images with 90% confidence level as an outlier.

3.3 Databases and experiment settings

Three public domain face databases, namely CMU-PIE [19], YaleB [8] and FRGC [17], are selected for experiments in this paper.
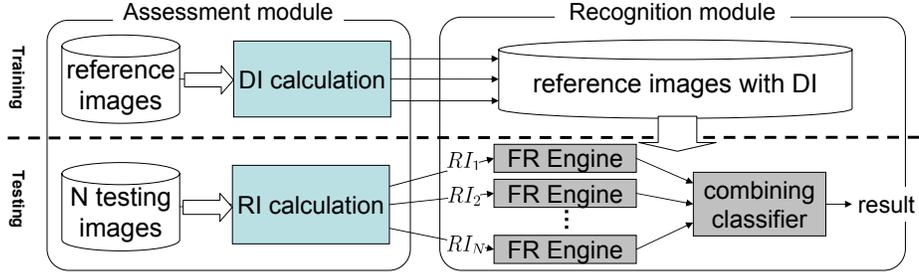
Figure 4. The block diagram of recognition system integrating with DI and RI .

CMU-PIE consists of 68 persons and each persons has 105 images(5 poses times 21 illumination); YaleB database consists of 38 persons with 64 different lighting conditions (totally 38 times 64 images); FRGC database consists of more than 300 persons with number of images ranging from 6 to 220. For FRGC database, we consider the person who has more than 50 images. For each person, 50 images with illumination, expression and wild pose variations are selected. After this selection, there are 311 classes by 50 images per class used in experiments.

By random selection, databases are divide into two separated sets (testing set and training set) in advance. The detailed setting of the experiments is shown in Table 1, where C is the number of individuals in the database, N_t is the number of training images per individual. Not all the images in training set will be used as reference images. N_r images in training set is randomly selected as reference images, and N_p images in testing set are randomly selected as probe.

All of the face images are extracted and aligned manually. Some images from three databases are shown in Figure 5.
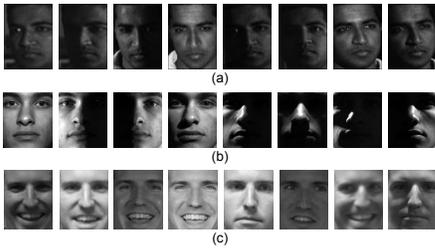
Figure 5. Sample images: (a) CMU-PIE ;(b) YaleB ; (c) FRGC .

3.4 Results and analysis

The results of three databases with different face recognition engines and difference combining classifiers are reported.

For CMU-PIE database, Eigenface representation is used in the recognition module and sum rule combining classifier is used to draw the final result. As shown in Figure 7(a), the results without using RI nor DI are used as benchmark. After using DI, the accuracy is improved from 68.4% to 76.9% (with $N_p = 15$) and from 74.1% to 84.5% (with $N_p = 55$); after using RI, the accuracy is improved to 73.3% (with $N_p = 15$) and 89.2% (with $N_p = 55$); after using both DI and RI, the system achieves the best result, and the accuracy improved to 73.5% (with $N_p = 15$) and 92.1% (with $N_p = 55$). The improvement ranges from 4.9% to 18.0%. Moreover, without DI nor RI, the improvement caused by increasing N_p is 5.7%, while it is 15.9% with both DI and RI. It can be seen from the results that integrating with DI and RI helps to improve the recognition system performance.

The robustness of the recognition system is evaluated by the standard deviation (StdDev) of the accuracy. The standard deviation of accuracy of CMU-PIE database on eigenface recognition engine with sum rule combining classifier is shown in Figure 6. The StdDev of methods using DI and/or RI is less than the one without DI nor RI. This implies that robustness of the recognition system is also improved by making use of DI and RI.

The additional computational time introduced by assessment module is estimated. For the case of CMU-PIE database with Eigenface engine and sum rule, it took 21.0 mins to calculate the DI of training images. That is, the time to calculate DI of one image is less than 3.8 seconds. It took 150 milliseconds to calculate RI of the testing images, specifically 40 microsecond per image, while the whole recognition procedure took 8.19 seconds on average. That means calculating RI in recognition phase only consumed 1.83% the total recognition time.

The experiments are repeated using majority voting and

Table 1. Experiment settings

database	C	N_t	N_r	N_p	variations
CMU-PIE	68	50	10	15∼55	pose, illumination
YaleB	38	32	4	15∼32	illumination
FRGC	311	20	4	15∼30	illumination, expression, mild pose

C is the number of individuals; N_t is the number of training images; N_r is the number of reference images; N_p is the number of testing images. (per individual)
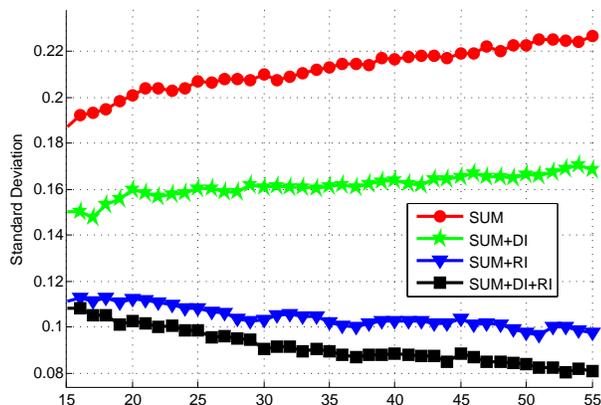

Figure 6. The standard deviation of results amount 100 times on CMU PIE with eigenface representation and sum rule.

product rule. Similar patterns are obtained, shown in Figure 7 (b) and (c), respectively. Moreover, replacing Eigenface representation with Kernel PCA representation, the experiments are repeated. The results are shown in Figure 8. It can be observed that the performances are improved after integrating DI and/or RI with the recognition system.

For YaleB database and FRGC database, the same experiments are conducted and the results are shown from Figure 9-12. It can be seen that same conclusions as that on CMU-PIE are drawn.

4 Conclusions

In this paper, two new measurements, discriminability index (DI) and reliability index (RI), are introduced. Differing from the existing image assessment methods, the proposed method is based on the discriminative power and the matching quality. These two new measurements help the multiple image based face recognition systems improving their performance. The experimental results show that after adopting the proposed DI and RI to the existing recognition algorithm, the recognition accuracy and the robustness of the performance are improved. The improvement of accu-

racy ranging from 4% to 30%.

References

[1] V. Barnett and T. Lewis. *Outliers in Statistical Data.* John Wiley & Sons New York, 1994.

[2] S. Blok, D. Medin, D. Osherson Probability from similarity *AAAI Conference on Commonsense Reasoning* . AAAI Press: Stanford, CA, 2002

[3] K. Bowyer, K. Chang, P. Flynn, and C. X. Face Recognition Using 2-D, 3-D, and Infrared: Is Multimodal Better Than Multisample? . *Proceedings of the IEEE*, 94(11), 2006.

[4] S. Canavan, M. Kozak, Y. Zhang, J. Sullins, M. Shreve, and D. Goldgof. Face Recognition by Multi-Frame Fusion of Rotating Heads in Videos. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2007.

[5] K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2D+ 3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005.

[6] W. Dixon. Analysis of extreme values. *The Annals of Mathematical Statistics*, pages 488–506, 1950.

[7] R.O. Duda, P.E. Hart and D.G. Stork *Pattern Classification*. Wiley New York, 2001

[8] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6): 643–660, 2001.

[9] F.E. Grubbs. Procedures for Detecting Outlying Observations in Samples *Technometrics*, 11(1): 1–21,1969.

[10] A. Hadid and M. Pietikainen. From still image to video-based face recognition: an experimental analysis. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 813–818, 2004.

[11] V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

[12] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, J.L. Wayman Biometrics: A Grand Challenge In *Proceedings of IEEE International Conference on Pattern Recognition (CVPR)*. IEEE Computer Society, 2004

[13] A. John. *Mathematical Statistics and Data Analysis*. Wadsworth & Brooks/Cole, 1988.

[14] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[15] V. Kruger and S. Zhou. Exemplar-based face recognition from video In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 175 – 180, 2002.

[16] X. Liu, J. Rittscher and T. Chen. Optimal Pose for Face Recognition. *Proceedings of IEEE International Conference on Pattern Recognition (CVPR)*.IEEE Computer Society, vol(2), pp:1439-1446, 2006.

[17] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer,*et al.*. Overview of the face recognition grand challenge In *Proceedings of IEEE International Conference on Pattern Recognition (CVPR)*.IEEE Computer Society, vol(1), pp:947-954, 2005.

[18] P.J. Phillips, W.T. Scruggs, A.J. O'Toole,*et al.*. FRVT 2006 and ICE 2006 Large-Scale Experimental Results *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, in press, 2009.

[19] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[20] X. Tan, S. Chen, Z. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.

[21] D. Thomas, K. W. Bowyer, and P. J. Flynn. Multi-frame approaches to improve face recognition. In *Proceedings of IEEE Workshop on Motion and Video Computing*, pages 19–19, 2007.

[22] Y. Zhang and A. Martínez. A weighted probabilistic approach to face recognition from multiple images and video sequences. *Image and Vision Computing*, 24(6):626–638, 2006.

[23] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
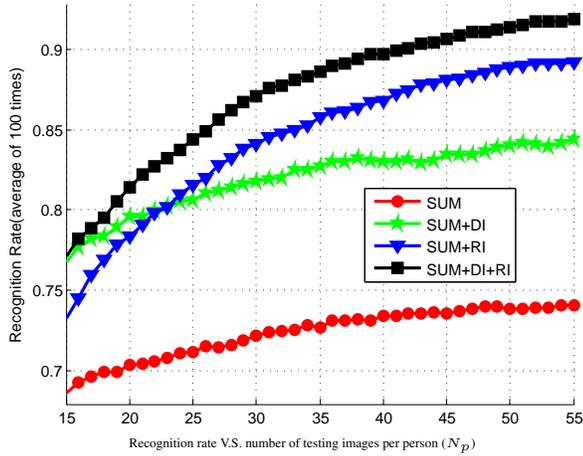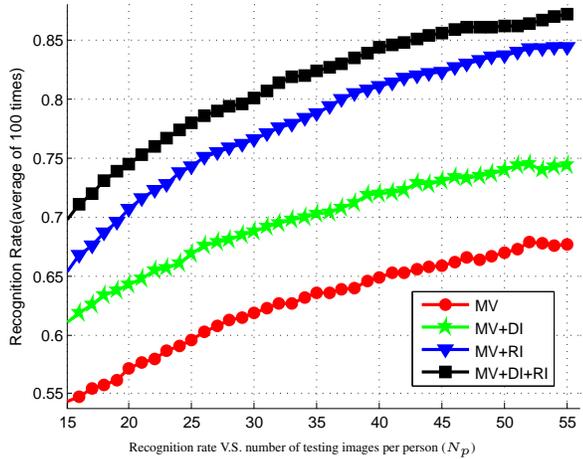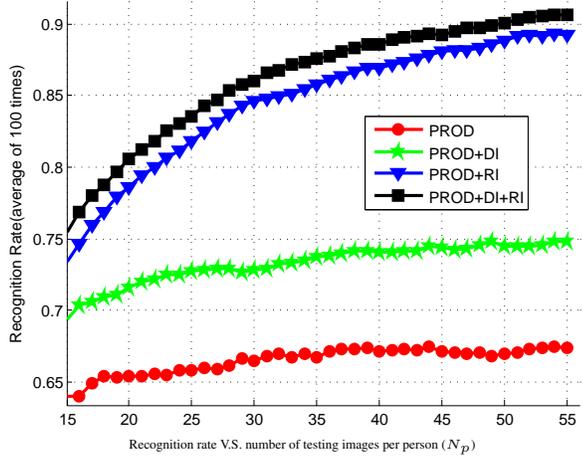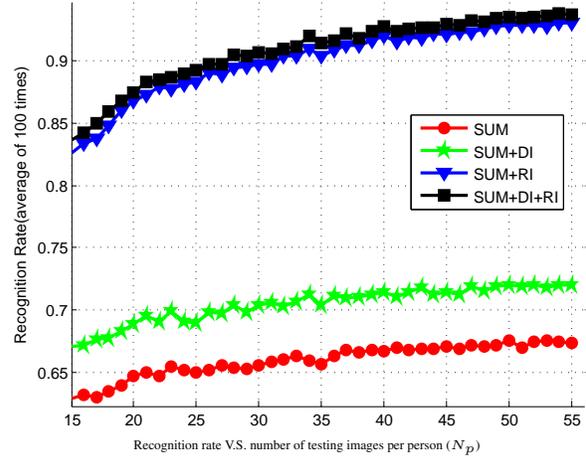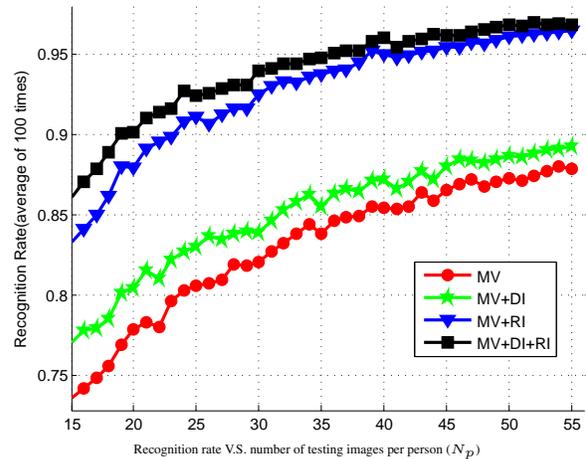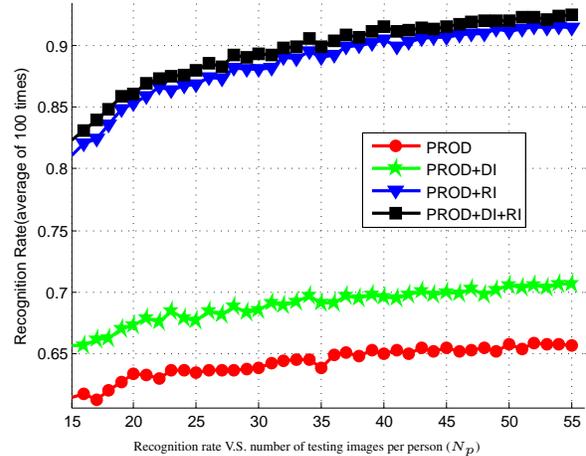
Figure 7. Results on CMU database with Eigenface representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).
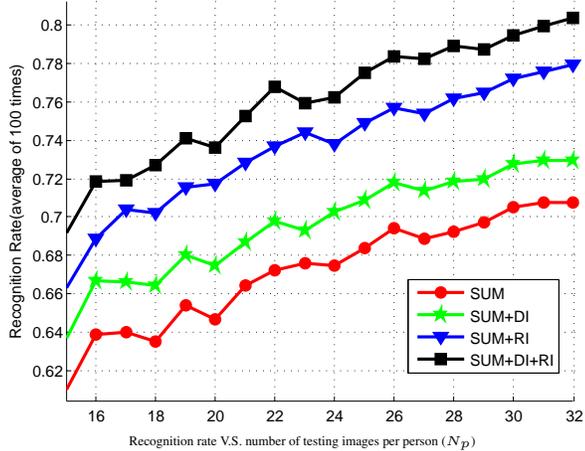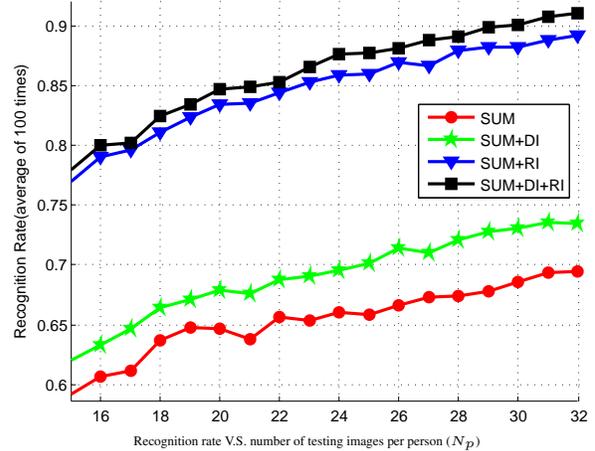
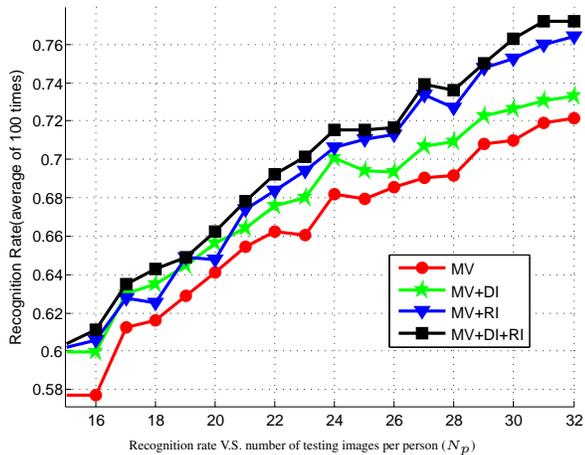**Figure 8. Results on CMU-PIE database with Kernel PCA representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).**
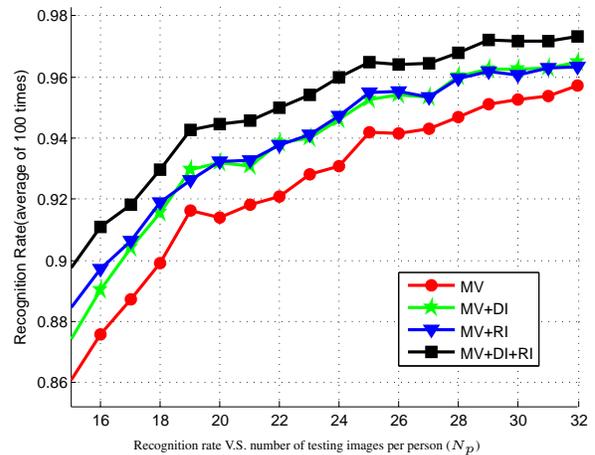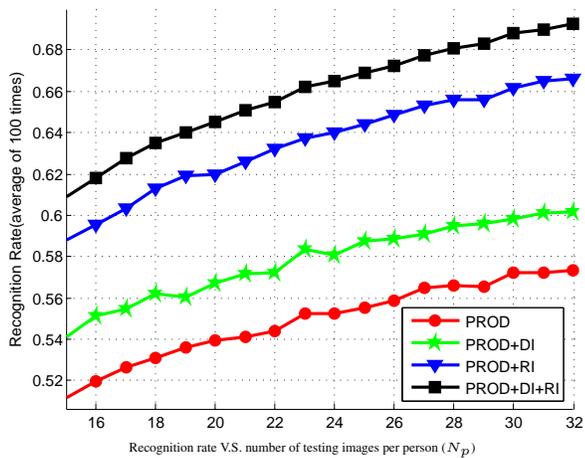
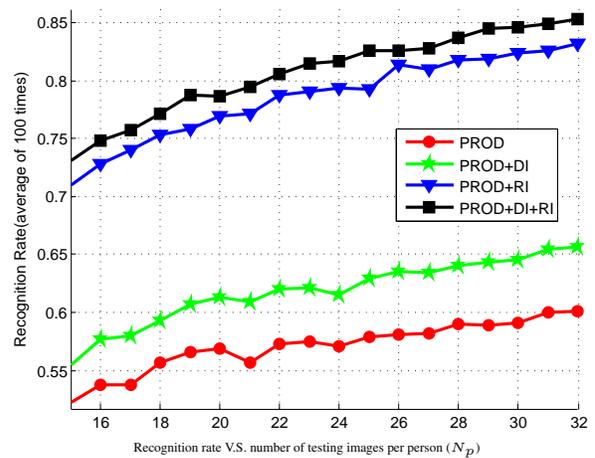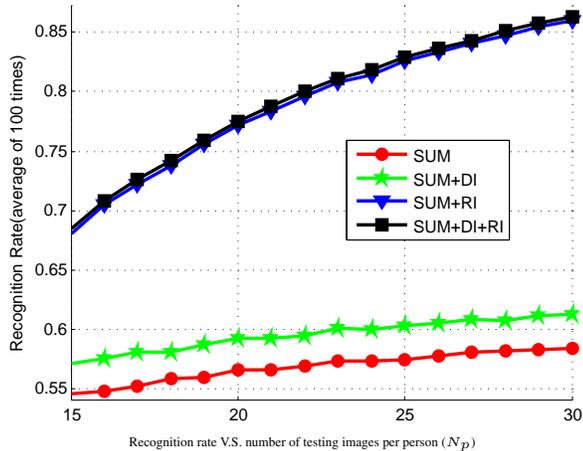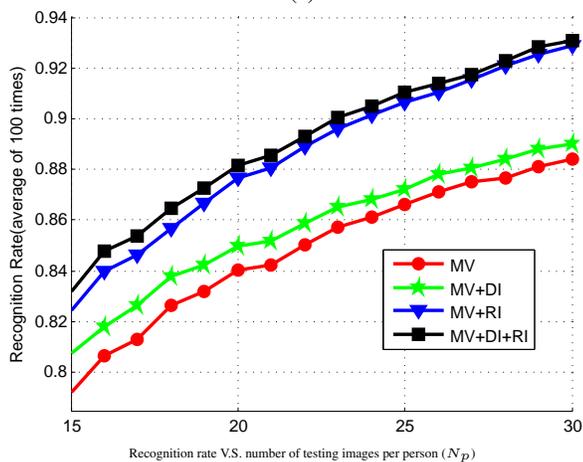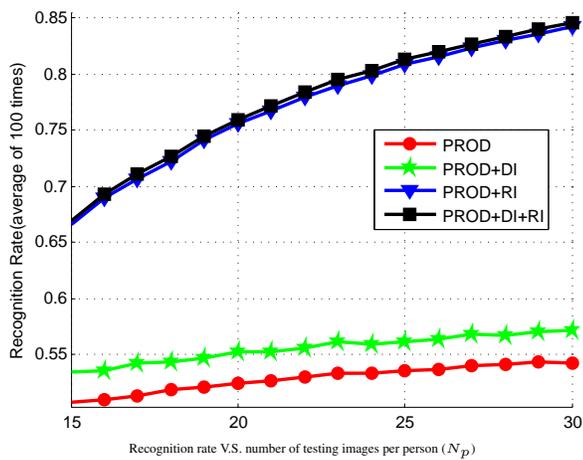**Figure 9. Results on YaleB database with Eigenface representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).**

Figure 10. Results on YaleB database with Kernel PCA representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).
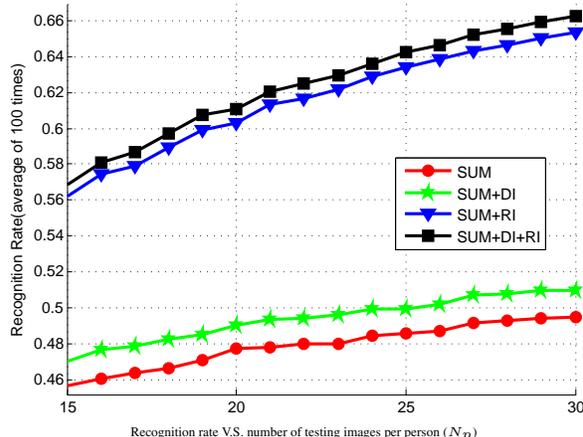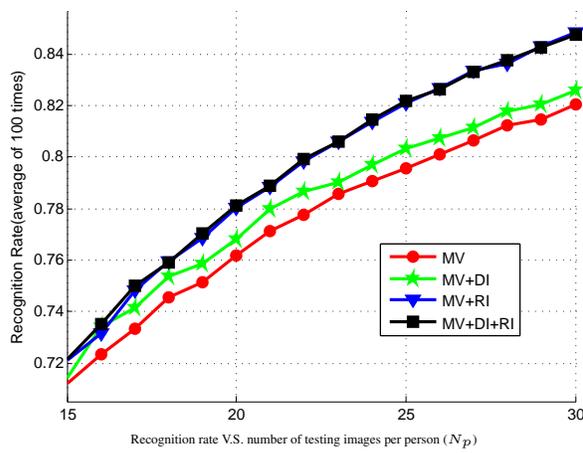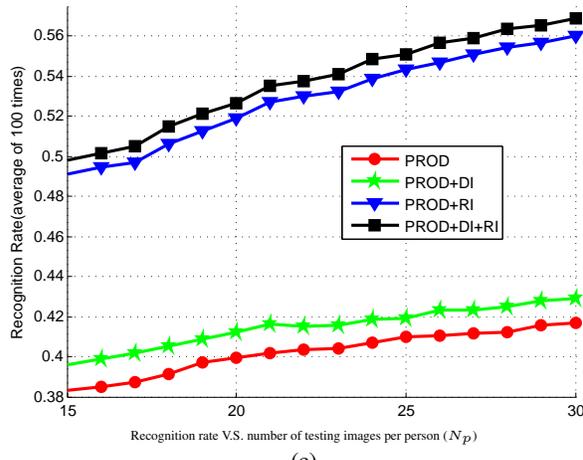
Figure 11. Results on FRGC database with Eigenface representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).

Figure 12. Results on FRGC database with Kernel PCA representation :(a) Sum rule (SUM) (b) Majority voting rule (MV) (c) Product rule (PROD).