Site-adaptation methods for face recognition

Jilin Tu and Xiaoming Liu and Peter Tu

***Abstract*— While the state-of-the-art face recognition algorithms are designed with the goal of reliably recognizing faces under arbitrary illumination background and uncontrolled imaging conditions, the performance of these face recognizers may still varies in the real-world applications, depending on how much the face appearance statistics in the training data matches that in the testing data in the feature space. Assuming the illumination environment and the imaging condition are not subject to frequent changes at each application site where the face recognition systems are deployed, we propose to do site adaptation for the generic face recognizer based on some face images captured by the cameras at the site as an adaptation set. Based on an OSFV[20] face recognizer with Gabor features selected by Adaboost algorithm, we propose several site adaptation methods at the feature level and at the model level. Our experiment results showed that the proposed site adaptation approaches can significantly boost the performance of our generic face recognition algorithm at site with unforeseen illumination background and imaging conditions with a small adaptation set.**

I. INTRODUCTION

For many real-world face recognition applications, the face recognition system will be deployed in many sites with uncontrolled camera setup and unforeseen illumination background. While the state-of-the-art face recognition algorithms are designed with the aim of yielding robust and accurate face recognition performance under arbitrary illumination background and uncontrolled imaging conditions, it is always desirable to develop some model adaptation algorithms that can adapt the generic face recognizer in hand based on a few sample face pictures captured on the sites so that the performance of the generic face recognizer can be further improved. We call the model adaptation in such scenario site adaptation.

The concept of model adaptation have been explored extensively in many speech recognition applications. While the speech recognizer is trained based on a large speech corpus, its performance on the speech of a new subject is usually poor due to the huge variability of human speeches from person to person. A typical solution is to adapt the speech recognition model to the new subjects based on a small adaptation data set collected from the new subjects. In general, the typical adaptation techniques can be classified into three categories: the maximum a posterior (MAP) adaptation[3], parameter transformation based adaptation using maximum likelihood linear regression(MLLR) [11], and speaker clustering-based adaptation approaches [10]. As these adaptation methods are proposed for the speech recognition models formulated in the framework of continuous density Hidden Markov Model(HMM). They can not directly applied to the face recognition scenarios.

There are only a few works in the face recognition domain on model adaptation. In [16], it was observed that brief periods of adaptation may serve to enhance recognition in high-level object processing for human vision systems. In [22], the generic intra-personal subspace for the Bayesian face recognizer [13] is adapted to person-specific intra-personal subspace based on a few adaptation images, so that the performance of the generic face recognizer can be improved on specific subjects in the testing set. In [12], person specific associative memory neural networks are trained based on the wavelet features for the face recognition task. It was shown that the performance of the generic person-specific face recognition model can be improved after adapted to a few adaptation data of these subjects in new environments.

In this paper, a few site adaptation techniques are explored based on a OSFV face recognition algorithm trained a subset of Gabor features selected by Adaboosting algorithm. In section 2, we will first introduce the details of the face recognition algorithm. Based on the algorithm design, we describe a few site adaptation methods in section 3. In section 4, we show the experiment results. We train our generic face recognizer based on the NIST Multiple Biometric Grand Challenge (MBGC) face database[15], and evaluate its performance on a testing set in Face-In-Action (FIA) database[4]. And we show that the performance of the generic face recognizer trained on MBGC face data can be enhanced significantly after our proposed site adaptation approaches when applied to FIA testing set.

II. OSFV FACE RECOGNITION WITH DISCRIMINATIVE GABOR FEATURES

A. Face recognition framework

The framework of our face recognition system is described in Fig. 1. Given a training set, we first compute the Gabor features in five scale spaces and eight orientations, i.e., a image is of size $K \times K$ is converted into Gabor features of size $K \times K \times 5 \times 8 = 40K^2$. As the feature space is of very high dimensions, we first do feature selection using Adaboosting algorithm, so that only those Gabor features with discriminative power are preserved. We then carry out PCA dimension reduction on the training data with

This work was supported by award #2007-DE-BX-K191 from 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 authors and do not necessarily reflect the views of the Department of Justice.

Jilin Tu, Xiaoming Liu and Peter Tu are with Visualization and Computer Vision Laboratory, General Electric Global Research Center, 1 Research Circle, Niskayuna, NY, 12309, {tujilin, liux, tu}@crd.ge.com

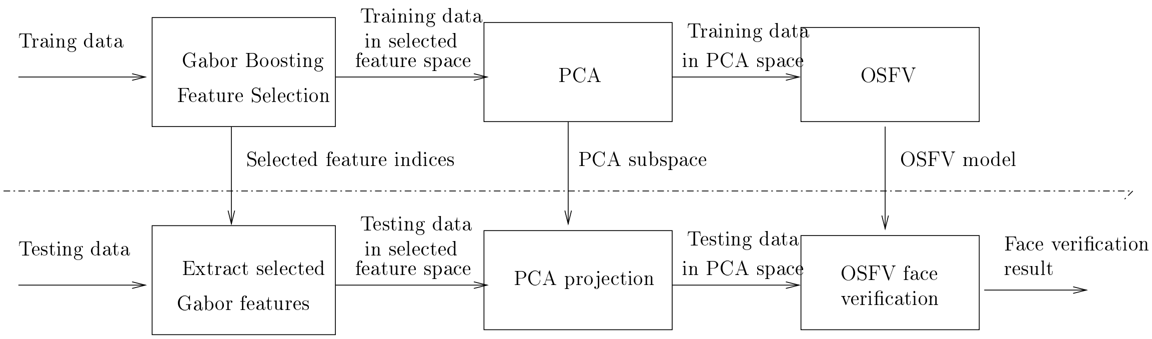

Fig. 1. The face recognition system framework

the selected Gabor features by preserving 99% of the data statistical variations. In the low dimensional PCA subspace, we then train a OSFV subspace[20] as the face recognizer.

At the testing stage, we first extract the corresponding selected Gabor features from the testing image, the selected Gabor features of the testing data are then projected into the PCA subspace, and face verification is then carried out using the OSFV model.

B. Face verification by OSFV

Optimal Subspace for Face Verification(OSFV) was proposed in [20], together with Optimal Subspace for Face Identification(OSFI), with the following motivations:

1) Different face recognition tasks (i.e., face identification and verification) have different performance metrics, which implies that there exist distinguished subspaces that optimize these scores, respectively. Most prior work focused on optimizing various discriminative or locality criteria and neglect such distinctions.
2) As the gallery (target) and the probe (query) data are collected in different settings in many real-world applications, there could exist consistent appearance incoherences between the gallery and the probe data for the same subject. Knowledge regarding these incoherences could be used to guide the algorithm design, resulting in performance gain. Prior efforts have not focused on these facts.

Instead of following the prior efforts that find subspaces to optimize various objective functions for preserving certain distributive, discriminative or locality properties of the data(PCA[9], [21], ICA[1], FDA[2], Bayesian “dual eigenspace”[14], Bayesian Optimal LDA[5], LPP[7], MFA[23], NPE[6], etc.), OSFV/I directly optimizes the face recognition performance score for the face verification and face identification task, respectively. Given the distinction in the face verification error (for the face verification task) and the face recognition rate (for the face identification task), [20] showed that the optimal subspaces for the two face recognition tasks are different.

Specific for the face verification task, considering the training data $\mathbf{X}$ with ground truth person identities as the performance evaluation data, we project the data into a subspace $\mathbf{A}$, denoted as $\mathbf{AX}$, and the evaluated verification error is formulated as

$$PE(\mathbf{A},T|\mathbf{X}) = \frac{FAR(\mathbf{AX},T) + FRR(\mathbf{AX},T)}{2} \tag{1}$$

where FAR and FRR are the false alarm rate and the false rejection rate, respectively, and T is the decision threshold. As FAR and FRR are defined according to the cumulative sum of the penalty function $f(u) = \Pi(u) = \begin{cases} 0, \text{if } u < 0 \\ 1, \text{if } u \geq 0 \end{cases}$ for the face verification error instances(the detailed formulation can be found in [20]), PE is a function of the subspace $\mathbf{A}$ and the decision threshold T. However, PE is not differentiable because $f(u) = \Pi(u)$ is not differentiable. By approximating the penalty function with a sigmoid function, i.e., let $f(u) = \frac{1}{1+e^{-u/\sigma}}$ (we have $f(u) \rightarrow \Pi(u)$, if $\sigma \rightarrow 0$), PE become differentiable, and gradient descent algorithms can be derived to optimize PE with respect to the subspace $\mathbf{A}$ and the decision threshold T, customized for a predefined distance metric that is differentiable (i.e., Euclidean or normalized correlation.). The experiments in [20] showed that OSFV can further improve the performance of the state-of-the-art subspace based face verification algorithms (i.e., FDA, LPP, MFA, NPE) in various databases.

III. The adaptation strategies

As the generic face recognizer achieves good face recognition performance by modeling certain statistical properties of the data in the training set, it would perform well on a testing set that possesses similar statistical characteristics of the training set. Given a testing set collected at a site with illumination background and camera setting different from the training set, the statistical coherence between the testing set and the training set may no longer be satisfied. Given a small set of adaptation data collected from this site, site adaptation can be carried out in two directions:

- **Feature Adaptation**: Find a linear or nonlinear transformation that adapt the site data in the feature space, so that the statistics of the transformed site data matches to that of the training set.
- **Model Adaptation**: Tune the face model parameters by improving the overall system performance on the adaptation data without over-fitting to the adaptation data.

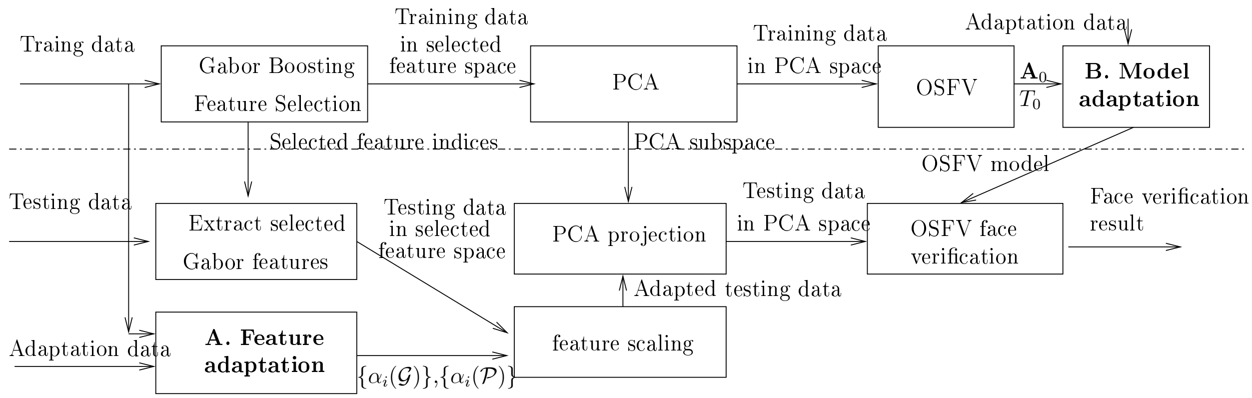

Fig. 2. The face recognition system framework with site adaptation. The Feature adaptation and the Model adaptation components are highlighted by bold font.

A. Feature adaptation for the selected Gabor features

Consider the Gabor features are the responses of the Gabor filters in different scale and orientation in the face images, The same person's frontal face captured in different illumination environment and with different camera configuration would produce Gabor features of different distribution. However, we can make an assumption that the strong edges on the face will yield relatively strong Gabor responses at the same location and orientation, but the magnitude of the Gabor response may vary due to the difference of imaging conditions. If we can model the statistics of the Gabor features in both the training data and the site data, we can compute the scaling factor by matching the statistics of the Gabor features of the training data to that from the site data, assuming the scaling of the corresponding Gabor magnitudes caused by the illumination background and camera setting variations is person-invariant, and is independent with respect to the scales and orientations.

Denote the Gabor magnitude features for the training set as $\mathbf{G}_{NK} = \{\vec{g}_1, \vec{g}_2, ..., \vec{g}_i, ..., \vec{g}_K\}$ where N is the number of samples and K is the number of selected Gabor features, and correspondingly, we denote the Gabor magnitude features for the site data as $\mathbf{H}_{MK} = \{\vec{h}_1, \vec{h}_2, ..., \vec{h}_i, ...\vec{h}_K\}$ (M is the number of samples in site data), we can model the statistics of the Gabor features in two approaches.

1) Modeling the statistics by Rayleigh distribution: In [8], it was reported that the Gabor magnitudes have tendency to satisfy Rayleigh distribution

$$R_\gamma(z) = \frac{z}{\gamma^2} exp(-\frac{z^2}{2\gamma^2}). \tag{2}$$

Let us assume the Gabor feature g_i follows Rayleigh distribution $R_{\gamma_i}(g_i)$, and assume h_i satisfies Rayleigh distribution $R_{\gamma_i'}(h_i)$, the distributions of the corresponding magnitudes can be matched as follows

$$R_{\gamma_i}(g_i) \sim \alpha_i R_{\alpha_i \gamma_i'}(\alpha_i h_i) \tag{3}$$

where

$$\alpha_i = \frac{\gamma_i}{\gamma_i'} \tag{4}$$

$$\gamma_i = \sqrt{\frac{1}{2N}\sum_n g_{ni}^2} \tag{5}$$

$$\gamma_i' = \sqrt{\frac{1}{2M}\sum_m h_{mi}^2}, \tag{6}$$

for $i = 1..K$.

2) Modeling the statistics by histograms: On the other hand, in case the Gabor magnitude feature violates the Rayleigh distribution assumption, we can model the distribution of the Gabor features in the training data and the site data by histograms. Denote $\mathbb{P}(g_i)$ as the histogram of Gabor feature g_i in the training data, and $\mathbb{Q}(h_i)$ as the histogram of Gabor feature h_i in the site data, the optimal scale α_i can be computed as follows

$$\alpha_i = argmin_\alpha KL(\mathbb{Q}(\alpha h_i)||\mathbb{P}(g_i)) \tag{7}$$

As the histogram $\mathbb{Q}(\alpha_i h_i)$ is not differentiable with repect to α_i, we employed discrete line search optimization technique to find the optimal α_i.

3) The algorithm: Consider the gallery and probe sets for the training data and for the site data might be collected in different imaging conditions (as shown in Fig. 3, here we assume the face image pairs for face verification are prepared with target image drawn from the gallery set and the query image drawn from the probe set respectively), the estimation of the scaling factor has to be estimated for the gallery and probe data, respectively.

At the training stage, we compute the sufficient statistics of the selected Gabor features in the training data. If we model the feature statistics by Rayleigh distribution, we compute $\{\gamma_i(\mathcal{G}_{train})|i = 1..N\}$, for the gallery set $\mathcal{G}_{train}$ in the training data and $\{\gamma_i(\mathcal{P}_{train})|i = 1..N\}$, for the probe set $\mathcal{P}_{train}$ in the training data, using Eq. 5. If we model the feature statistics by histograms, we compute the histogram $\{\mathbb{P}(g_i|\mathcal{G}_{train})|i = 1..N\}$, for the gallery set in the training data and $\{\mathbb{P}(g_i|\mathcal{P}_{train})|i = 1..N\}$, for the probe set in the training data.

At the site-adaptation stage, we collect a set of site data that contains the gallery set $\mathcal{G}_{site}$ and the probe set $\mathcal{P}_{site}$ that are captured with the typical illumination background and the typical camera settings. We can then compute the sufficient statistics of the selected Gabor features in the site data, respectively for the gallery set and for the probe set. And the Gabor feature scaling parameters $\{\alpha_i(\mathcal{G})\}$ and $\{\alpha_i(\mathcal{P})\}$ can be computed using Eq. 4 (if the statistics is modeled by Rayleigh distribution) or Eq. 7 (if the statistics is modeled by histogram).

At the system deployment stage, the target(gallery) image and the query(probe) images are acquired on site, and the selected Gabor features are computed, and scaled by $\{\alpha_i(\mathcal{G})\}$ and $\{\alpha_i(\mathcal{P})\}$, respectively, for the images. The scale-adapted Gabor features are then sent to the generic OSFV face verifier for face verification.

B. Model adaptation for OSFV classifier

In [18], prior human knowledge is incorporated into the training of Adaboosting classifier as a means of compensating for a shortage of training data. The prior human knowledge is formulated as a rule set that maps each instance x to an estimated conditional probability distribution $\pi(y|x)$ over the possible label values $y \in \{-1, +1\}$. The training of the Adaboosting classifier then minimizes not only the classification error in the training data, but also the distance between the likelihood distribution of the classification results and the prior model $\pi(y|x)$.

Consider a training set $\{x_i\}$, with a prior model $\pi(y|x)$, and assume the output of the classifier to be trained is $f(x)$, a practical likelihood model for the classifier is $p(+1|f(x)) = \sigma(f(x)) = \frac{1}{1+e^{-f(x)}}$. The distance between the prior model and the likelihood of the classifier can be modeled by KL divergence, $\sum_i RE(\pi(+1|x_i)||\sigma(f(x_i)))$, where $RE(p||q) = pln(p/q) + (1-p)ln((1-p)/(1-q))$.

This formulation can be easily applied to the model adaptation scenario where we can consider the output of the generic face recognizer as the prior "human knowledge" model, and the adaptation set as the training set. Consider $\vec{x}$ represents the face image pair for identity comparison, $f(\mathbf{A}_0\vec{x})$ is the identity similarity measure computed in the generic OSFV subspace $\mathbf{A}_0$, we can define the similarity prior model to be

$$\pi(+1|\vec{x}) = \sigma_{T_0}(f(\mathbf{A}_0\vec{x})) = \frac{1}{1+e^{-(f(\mathbf{A}_0\vec{x})-T_0)}}$$

where T_0 is the generic decision threshold.

We can then retrain the OSFV model on the adaptation set $\mathbf{X}_{adapt}$ by minimizing the verification error evaluated on $\mathbf{X}_{adapt}$, together with the average distance between the decision likelihood of the current classifier (with $\mathbf{A}$ and T) and the prior model, the decision likelihood of the generic classifier (with $\mathbf{A}_0$ and T_0), as follows:

$$\begin{aligned} C(\mathbf{A}, T|\mathbf{X}_{adapt}) &= PE(\mathbf{A}, T|\mathbf{X}_{adapt}) + \\ &\lambda \sum_{\vec{x}\in\mathbf{X}_{adapt}} RE(\pi(+1|\vec{x}))||\sigma_T(f(\mathbf{A}\vec{x}))) \end{aligned} \quad (8)$$

where λ is a weighting factor that determines how much the decision of the adapted face verifier model can deviate away from that of the generic face verifier model on the adaptation set. In summary, our proposed site adaptation approaches can be summarized Fig. 2. The generic face verifier is obtained at the training stage with only the training data. The model adaptation and feature adaptation is carried out with the adaptation data at the site. After site adaptation, the system can be deployed with the adapted model and feature configuration. The feature adaptation and the model adaptation can be utilized jointly if the adaptation data provided to the model adaptation component is processed by feature adaptation beforehand.

IV. Experiment

A. Data preparation

We first train a generic face verification model based on a subset of the NIST Multiple Biometric Grand Challenge (MBGC) database[15]. The training set contains 1395 gallery (target) images and 7663 probe (query) images for 469 subjects. The gallery images are frontal face images captured in controlled illumination, and the probe images are frontal face images captured in uncontrolled indoor illuminations. We split the MBGC data into a training set and a testing set with non-overlapping subject identity. The training set contains 396 subjects and the testing set contains the rest 69 subjects.

We would like to evaluate the performance of the generic face recognizer on a testing face data subset from CMU Face In Action (FIA) database[4]. The FIA database consists of 20-second videos of face database from 206 participants mimicking a passport checking scenario. The gallery videos are collected with indoor illumination background and the probe videos are collected with outdoor illumination backgrounds. For each subject, we extract a frontal face image from the gallery video as the gallery data, and a few frontal face images from the probe videos as the probe data. We then take the data of the first 103 subjects as adaptation set, and the data of the rest 103 subjects as testing set. Figure 3 shows the typical gallery and probe images of subjects from MBGC and FIA database. Due to the fact that they are collected in totally different imaging conditions, there exists strong inconsistency in the appearance statistics between the two databases. We align the faces by the eye locations provided by Pittpatt face detector[17], crop the faces to image size 80 by 80 pixels, and normalize the image to zero mean and standard deviation after DoG filtering for illumination removal [19]. Gabor feature of size 256000 ($80 \times 80 \times 40$) are then generated based on the preprocessed images.

B. The performance of the generic face recognizer

We first apply Adaboosting feature selection algorithm to the MBGC database, and obtained 13888 discriminative Gabor feature indices(about 5% of the total Gabor features). We then compute a PCA subspace of dimension 1133 that preserves 99% of the energy in the training data based on the selected Gabor features. In the PCA subspace, we then apply

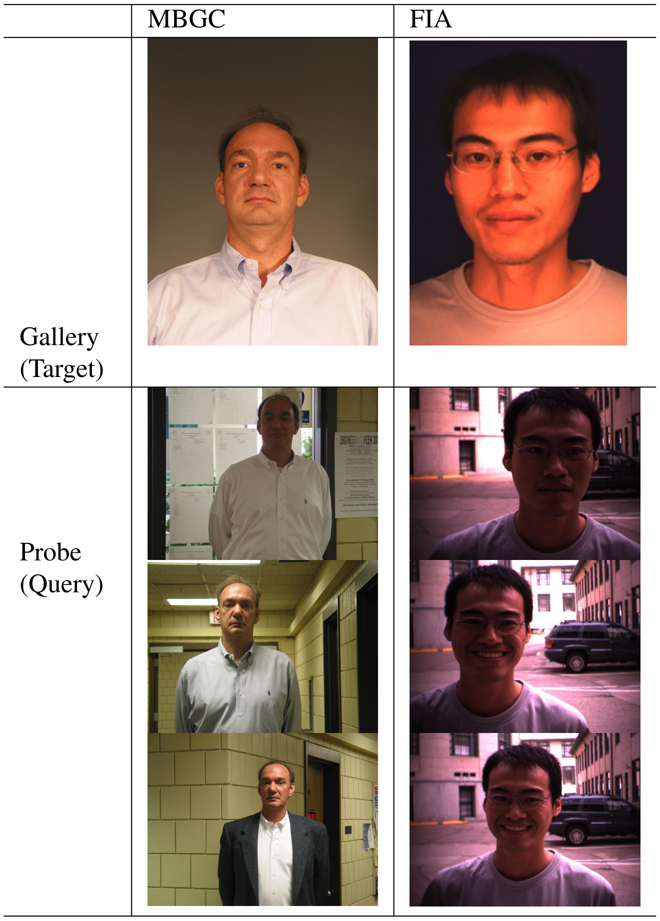

Fig. 3. The gallery and probe face images in MBGC and FIA databases. Notice the universal illumination differences between the MBGC and FIA databases in both Gallery (Target) and probe (Query) set.

the OSFV algorithm we proposed to obtain a discriminative subspace that minimizes the face verification error in the training data. After the training, we apply the trained face model to the MBGC testing data set, and we obtain the testing face verification equal error rate(EER) 8% (as shown in the first column of Tab. I).

We then treat all the FIA adaptation set as a training set for the FIA data set(as it contains face data of 103 subjects), and we train an OSFV face verification model customized for the FIA data using the same training procedure as before. The FIA specific face recognizer achieves EER 8.5% on the FIA testing set (as shown in the second column of Tab. I).

Finally, we apply the OSFV face verification model trained on MBGC data to the FIA testing data and evaluate the cross-database face verification performance. It was shown that the cross database performance achieves EER 14% (as shown in the third column of Tab. I).

We consider the face recognition model trained with MBGC database as our generic face recognizer (denoted as MBGC face recognizer), and the goal of site adaptation is to improve the performance of the generic face recognizer on the FIA testing set, given a subset of the FIA adaptation set. The performance of the face recognizer trained on the whole FIA adaptation set (denoted as FIA face recognzier)

MBGC	FIA	$MBGC \rightarrow FIA$
8%	8.5%	13.9%

TABLE I

THE BASELINE PERFORMANCE FOR THE GENERIC FACE RECOGNIZER TRAINED ON THE MBGC DATABASE AND FIA DATABASE, AND THE CROSS-DATABASE PERFORMANCE EVALUATION.

can serve approximately as the target performance for site adaptation.

C. Feature adaptation

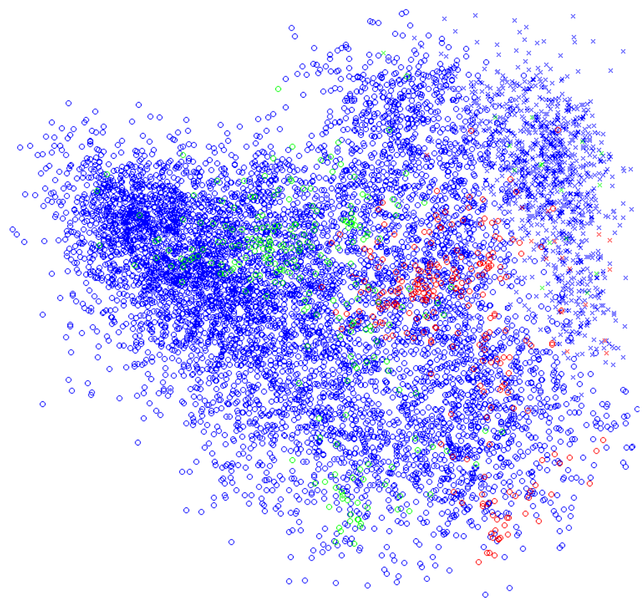

Fig. 4. The distribution of the MBGC training data (blue), the FIA site data before feature adaptation (red) and the site data after feature adaptation (green) in 2 dimensional PCA subspace. The gallery(target) data is labeled with 'x', and the probe(query) data is labeled with 'o'. It shows the distribution of the FIA site data matches the MBGC training data better after feature adaptation.

In Fig. 4, we plot distributions of the MBGC training data (blue), the FIA testing data before (red) and after (green) feature adaptation, after they are projected into the same 2 dimensional PCA subspace. The adaptation parameters are computed based on a FIA adaptation data subset with 16 subjects based on Rayleigh distribution model. The visualization show that feature adaptation can transform the distribution of the FIA testing data so that it become more consistent with that of the MBGC training data.

By increasing the number of subjects in the sequence of 2, 4, 8, 16, 32, 64 in the adaptation subset, we can evaluate how the performance of MBGC trained face recognizer is improved after the feature adaptation with respect to the size of the adaptation data subset. By generating the adaptation subset with different subject number 10 times by randomly drawing subjects from the whole adaptation data, we can compute the statistics of the performance improvement. Fig. 5 shows the mean and standard deviation of the performance

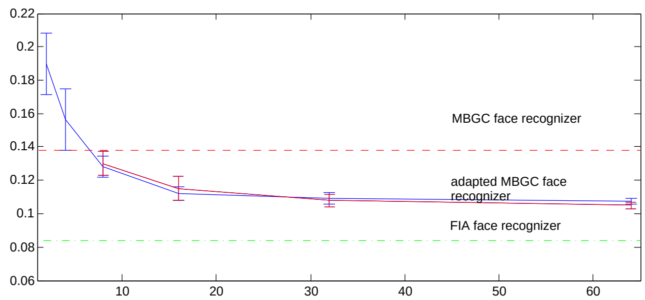

Fig. 5. The performance improvement of the MBGC face recognizer on the FIA testing data improves while the number of subjects in the adaptation set increases. The blue line with error bar shows the performance of feature adaptation by mean match-up. And the red line with error bar shows the performance of feature adaptation by distribution match-up.

of the MBGC face recognizer after adaptation with respect to the number of the subjects in the adaptation subset. The blue line with error bar shows the performance enhancement of the feature adaptation based on Rayleigh distribution model, and the red line with error bar shows the performance enhancement of the feature adaptation based on distribution model. For comparison, we plotted the performance of the MBGC face recognizer evaluated on the FIA testing data with red dotted line, and the performance of the FIA face recognizer on the testing set with green dotted line. It was shown that the performance of MBGC face recognizer after feature adaptation get worse if the number of subjects in the adaptation subset is less than 8, but the performance is stably improved if the number of subject is more than 8. The improvement stabilizes after the number of subject is more 16. The feature adaptation based on histogram model achieves slightly better performance than that based on the Rayleigh distribution model when the number of subject is more than 30.

D. Model adaptation for OSFV classifier

As the training/adaptation of the OSFV model is a gradient descent algorithm that is computationally expensive, we consider a model adaptation scenario where an adaptation subset of 16 subjects is prepared. We first do feature adaptation, and we observe that the EER on the FIA adaptation subset reduces from 14% to 8.5%, and the EER on the testing set reduces from 13.9% to 10.3%. By applying the model adaptation for the OSFV classifier, we observe the EER on the adaptation subset reduces to 1.4% and the EER on the testing set reduces to 9.8%. The results are shown in Tab. II.

%	FIA adaptation subset	FIA testing data
No adaptation	14.0	13.9
Feature adaptation	8.5	10.3
Feature&Model adaptation	1.4	9.8

TABLE II

MODEL ADAPTATION FOR OSFV CLASSIFIER ON FIA DATABASE

V. SUMMARY

In this paper, we proposed two site-adaptation methods for generic face recognizer, with the hope that the performance of a generic face recognizer can further improved when deployed at a specific application site, if a small adaptation data set is provided. Our experiment results showed that the proposed site adaptation approaches can significantly enhance the performance of our face recognizer which is trained on MBGC database, adapted to and evaluated on the FIA database.

REFERENCES

[1] M. Bartlett, H. Lades, and T. Sejnowski. Independent component representations for face recognition. In *SPIE*, volume 3299, pages 528–539, 1998.
[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In *ECCV*, pages 45–58, 1996.
[3] J. L. Gauvain and C. H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Trans. Speech Audio Processing*, 2:291–298, 1994.
[4] R. Goh, L. Liu, X. Liu, and T. Chen. The cmu face in action (fia) database. In *FG*, pages 255–263, 2005.
[5] O. Hamsici and A. Martinez. Bayes optimality in linear discriminant analysis. *IEEE TPAMI*, 30(4):647–657, 2008.
[6] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In *ICCV*, volume 2, pages 1208–1213.
[7] X. He and P. Niyogi. Locality preserving projections. In *NIPS*, 2003.
[8] S. B. Jelena and B. S. Manjunath. On the rayleigh nature of gabor filter outputs. In *in Intl. Conf. on Image Processing (ICIP)*, pages 745–748, 2003.
[9] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE TPAMI*, 12(1):103–108, 1990.
[10] R. Kuhn, P. Nguyen, J.-C. Junqua, L. Goldwasser, N. Niedzielski, S. Fincke, K.Field, and M.Contolini. Eigenvoices for speaker adaptation. In *Int. Conf. Spoken Language Processing*, Dec 1998.
[11] C. J. Leggetter and P. C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models. *Comput. Speech Lang.*, 9:171185, 1995.
[12] J. Lin, J.-P. Li, and M. Ji. Robust face recognition by wavelet features and model adaptation. In *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. International Conference on*, volume 4, pages 1638–1643, Nov. 2007.
[13] B. Moghaddam. Bayesian face recognition. *Pattern Recognition*, 13(11):1771–1782, 2000.
[14] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.
[15] NIST. Multiple biometric grand challenge. http://face.nist .gov/mbgc/, 2008.
[16] I. Oruc and J. Barton. Brief adaptation increases sensitivity of face recognition. *J. Vis.*, 8(6):1143–1143, 5 2008.
[17] P. P. Recognition. Pittpatt face tracking software development kit. http://www.pittpatt.com/solutions/sdk.php.
[18] R. E. Schapire, M. Rochery, M. Rahim, and N. Gupta. Incorporating prior knowledge into boosting. In *Proceedings of the ICML*, pages 538–545, 2002.
[19] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *FG07*, pages 168–182, 2007.
[20] J. Tu, X. Liu, and P. Tu. On optimizing subspaces for face recognition. In *ICCV*, 2009.
[21] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.
[22] L. Wang, X. Wang, and J. Feng. Subspace distance analysis with application to adaptive bayesian algorithm for face recognition. *Pattern Recognition*, 39(3):456–464, 2006.
[23] S. Yan, D. Xu, B. Zhang, and H. Zhang. Graph embedding: A general framework for dimensionality reduction. In *CVPR*, volume 2, pages 830–837, 2005.