[7] B. V. K. V. Kumar, D. W. Carlson, and A. Mahalanobis, "Optimal trade-off synthetic discriminant function filters for arbitrary devices," *Opt. Lett.*, vol. 19, no. 19, pp. 1556–1558, 1994.
[8] M. Savvides, K. Venkataramani, and B. V. K. Vijaya Kumar, "Incremental updating of advanced correlation filters for biometric authentication systems," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 3, Jul. 2003, pp. 229–232.
[9] B. V. K. V. Kumar, M. Savvides, C. Xie, K. Venkataramani, and J. Thornton, "Using composite correlation filters for biometric verification," in *Proc. SPIE*, vol. 5106, Apr. 2003, pp. 13–21.
[10] K. Venkataramani and B. V. K. Vijaya Kumar, "Performance of composite correlation filters in fingerprint verification," *Opt. Eng.*, vol. 43, no. 8, pp. 1820–1827, Aug. 2004.
[11] X. Liu, T. Chen, and B. V. K. Vijaya Kumar, "Face authentication for multiple subjects using eigenflow," *Pattern Recognit.*, vol. 36, no. 2, pp. 313–328, Feb. 2003.
[12] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

Target Dependent Score Normalization Techniques and Their Application to Signature Verification

Julian Fierrez-Aguilar, Javier Ortega-Garcia, and Joaquin Gonzalez-Rodriguez

***Abstract*—Score normalization methods in biometric verification, which encompass the more traditional user-dependent decision thresholding techniques, are reviewed from a test hypotheses point of view. These are classified into test dependent and target dependent methods. The focus of the paper is on target dependent score normalization techniques, which are further classified into impostor-centric, target-centric, and target-impostor methods. These are applied to an on-line signature verification system on signature data from the First International Signature Verification Competition (SVC 2004). In particular, a target-centric technique based on the cross-validation procedure provides the best relative performance improvement testing both with skilled (19%) and random forgeries (53%) as compared to the raw verification performance without score normalization (7.14% and 1.06% Equal Error Rate for skilled and random forgeries, respectively).**

***Index Terms*—Biometrics, decision threshold, score normalization, signature verification.**

I. Introduction

Automatic extraction of identity cues from personal traits (e.g., fingerprints, speech, or face images) has given rise to a particular area of pattern recognition (*biometrics*) where the goal is to infer identity of people from personal data [1], [2]. The increasing interest in biometrics is related to the number of important applications where a correct assessment of identity is crucial. Biometrics provides a way to establish an identity based on "*who you are*," rather than by "*what you possess*" or "*what you know*." This concept not only ensures enhanced security but also avoids the need to remember and maintain multiple passwords.

Manuscript received December 15, 2003; revised May 19, 2004. This work was supported by the Spanish Ministry of Science and Technology under projects TIC2003-09068-C02-01 and TIC2003-08382-C05-01. This paper was recommended by Guest Editor D. Zhang.

The authors are with the Biometrics Research Lab.-ATVS, EUIT Telecomunicacion, Universidad Politecnica de Madrid, 28031 Madrid, Spain (e-mail: jfierrez@diac.upm.es; jortega@diac.upm.es; jgonzalez@diac.upm.es).

Digital Object Identifier 10.1109/TSMCC.2005.848181

Previous studies have shown that the performance of a number of biometric verification systems, especially those based on behavioral traits such as written signature [3]–[5] and voice [6], [7], can be improved with user-dependent decision thresholds. Even greater verification performance improvement can be expected through the use of score normalization techniques [8], [9]. These methods (which include the user-dependent decision thresholding as a particular case) account not only for user specificities but also for intersession and environment changes [10].

The objectives of this work are: 1) to provide a framework for score normalization collecting previous work in related areas; 2) to provide some guidelines for the application of these techniques in real world scenarios; and 3) to provide an example of a successful application of the proposed normalization methods regarding the First International Signature Verification Competition (SVC 2004) [11], where the system proposed by the authors [12] was ranked first and second for random and skilled forgeries, respectively.

The paper is structured as follows: The Introduction includes some definitions, the system model of biometric verification with score normalization, and the description of a preliminary experiment which corroborates the motivation of this work.[1] In Section III, the subset of score normalization methods we focus on is detailed. Some experiments on the development corpus of SVC 2004 extended task are reported in Section IV. Conclusions are given in Section V.

A. Definitions and System Model

In *authentication* (also known as *verification*) applications, the *clients* or *targets* are known to the system (through an *enrollment* or *training* process) whereas the impostors can potentially be the world population. In such applications, the users provide a biometric sample X (e.g., a written signature) and their claimed identities $\mathcal{T}$ and a one-to-one matching is performed. The result of the comparison s (*similarity score*) can be further normalized to s_n before comparing it to a *decision threshold*. If the score is higher than the decision threshold, then the claim is accepted; otherwise, the claim is rejected. The system model of biometric authentication with score normalization is provided in Fig. 1 for an on-line signature verification application.

Depending on the biometric verification system at hand, impostors may know information about the client that lowers verification performance when it is exploited (e.g., signature shape in signature verification). As a result, two kinds of impostors are usually considered, namely: 1) *casual impostors* producing *random forgeries*, when no information about target user is known and 2) *real impostors* producing *skilled forgeries*, when some information regarding the biometric trait being forged is used.

B. Experimental Motivation

As pointed out above, it has been observed in a number of biometric verification systems that using user-dependent thresholds improves verification performance [3]–[7]. This occurs because the client and impostor score distributions are not aligned for the different targets involved (mainly due to target specificities). The following preliminary experiment by using the on-line signature verification system described in [12] on the development corpus of the SVC 2004 extended task [11] corroborates this fact.

Target-dependent client and impostor score distributions (Gaussian fit) are plotted in Fig. 2 and show testing either with skilled (left) or

[1] In Section II, the framework for score normalization and some background on error estimation methods is described.

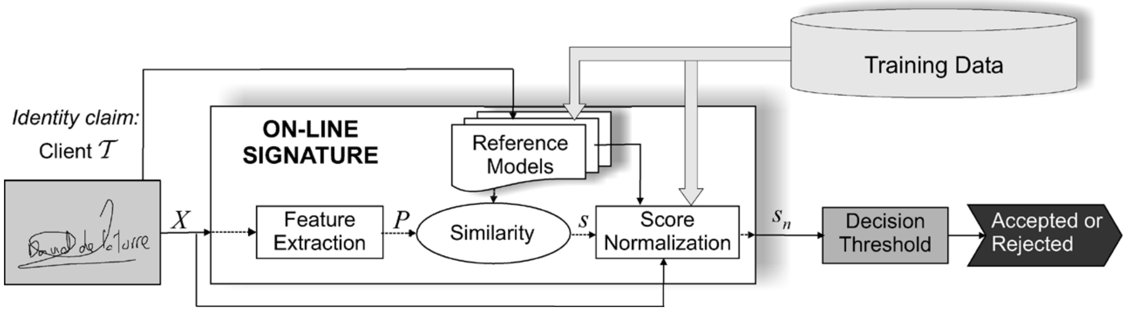

Fig. 1. System model of biometric authentication with score normalization for an on-line signature verification application.

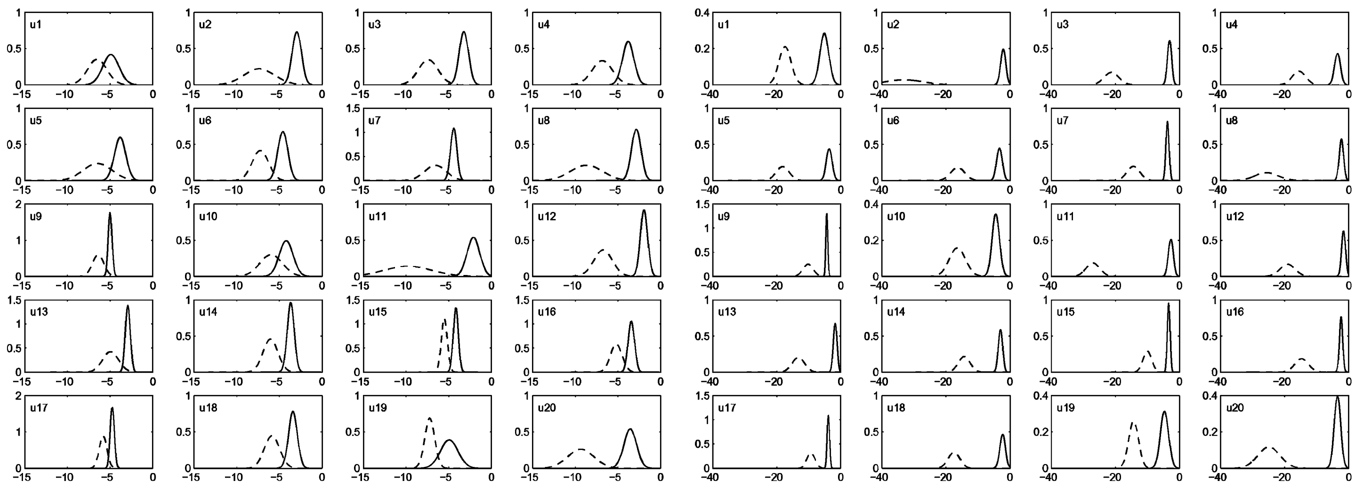

Fig. 2. Gaussian fit of client (solid) and impostor (dashed) score distributions for targets u1 to u20 of SVC 2004 development corpus considering either skilled (left) or random (right) forgeries from respectively real and casual impostors.

random forgeries (right). In this experiment, it can be observed that individual verification performance between some of the targets is quite different (e.g., u1 and u8 when testing with skilled forgeries). Big differences can also be observed in the client-impostor scoring regions (e.g., u1 and u9 when testing with random forgeries). The main objective of using user-dependent decision thresholds or, more generally, applying target dependent score normalization techniques with a unique user-independent decision threshold [8], [9], is to prevent such misalignments.

II. Theoretical Background

A. Score Normalization

Given a test sample X the problem of biometric authentication can be stated as a basic hypotheses test between two hypotheses:

- $H0$: X is from hypothesized client $\mathcal{T}$.
- $H1$: X is *not* from hypothesized client $\mathcal{T}$.

The optimum test to decide between these two hypotheses is a likelihood ratio test given by [13]

$$\frac{p(X|H0)}{p(X|H1)} \begin{cases} > \theta & \text{Accept } H0 \\ < \theta & \text{Accept } H1 \end{cases} \tag{1}$$

where $p(X|H0)$ and $p(X|H1)$ are respectively the probability density functions for the hypotheses $H0$ and $H1$ evaluated for the observed biometric sample X. The decision threshold for accepting or rejecting $H0$ is θ. An equivalent log-likelihood ratio test is obtained by transforming (1) into the log domain

$$\log p(X|H0) - \log p(X|H1) \begin{cases} > \log\theta & \text{Accept } H0 \\ < \log\theta & \text{Accept } H1 \end{cases}. \tag{2}$$

A common practice in biometric verification (e.g., GMM in the case of speaker recognition [14], HMM in the case of signature recognition [12], etc.) consists in characterizing each client $\mathcal{T}$ by a statistical model $\lambda^{\mathcal{T}}$ (i.e., the reference model in Fig. 1). In this case, the similarity s is computed as

$$s = \log p(X|\lambda^{\mathcal{T}}) \tag{3}$$

which is an estimation of $\log p(X|H0)$. As a result, the optimal score normalization strategy for an authentication system based on statistical modeling is given by

$$s_n = s - \log p(X|H1). \tag{4}$$

Note that, the normalizing term $\log p(X|H1)$ is affected, in general, by the following.

- Input information: the input biometric sample X.
- Information from client: scores $s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}$ from the hypothesized target $\mathcal{T}$ claiming its model $\lambda^{\mathcal{T}}$.
- Information from impostors: both models $\{\lambda_1^{\bar{\mathcal{T}}}, \ldots, \lambda_{N_I}^{\bar{\mathcal{T}}}\}$ and scores $\{s_1^{\bar{\mathcal{T}}}, \ldots, s_{N_{\bar{\mathcal{T}}}}^{\bar{\mathcal{T}}}\}$ from N_I possible impostors (either real or casual) of the hypothesized client $\mathcal{T}$ claiming the model $\lambda^{\mathcal{T}}$.

Estimation of $\log p(X|H1)$ based on the different information involved is not a straightforward task. Thus, operational procedures of *score normalization* (also known as *likelihood normalization*) are usually employed. Much effort has gone in to deriving such procedures based on the statistical formalism described above, mainly by the speaker recognition community [8], [9]. These operational procedures aim at designing a function

$$s_n = f\left(s, X, \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \left\{\lambda_1^{\bar{\mathcal{T}}}, \ldots, \lambda_{N_I}^{\bar{\mathcal{T}}}\right\}, \left\{s_1^{\bar{\mathcal{T}}}, \ldots, s_{N_{\bar{\mathcal{T}}}}^{\bar{\mathcal{T}}}\right\}\right) \tag{5}$$

so as to minimize the error rate of the verification task. The use of linear functions of various statistics of the information involved in (5) is the prevailing strategy for deriving normalization schemes. This is the case of [9]: 1) z-norm, which considers only scores from impostors; 2) t-norm, based on the input biometric sample and models from impostors; and 3) UBM-norm, which considers the input biometric sample and a universal background model characterizing the average target. Other examples can also be found regarding face [15] or signature recognition [16].

In order to simplify the discussion while providing a powerful framework for score alignment, neither input test information nor models from impostors are considered in this work

$$s_n = f\left(s, \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \left\{s_1^{\bar{\mathcal{T}}}, \ldots, s_{N_{\bar{\mathcal{T}}}}^{\bar{\mathcal{T}}}\right\}\right). \tag{6}$$

This family of score normalization methods will be referred to as *target dependent score normalization techniques*. Other normalization methods using the input biometric sample and models from impostors will be referred to as *test dependent normalization techniques*.

B. Error Estimation Methods

Biometric verification involves a tradeoff between two types of errors: 1) *False Rejection* (FR), occurring when a client user is rejected by the system and 2) *False Acceptance* (FA), taking place when an impostor is accepted as being a true user. A specific point is attained when FA and FR rates coincide, the so-called *Equal Error Rate* (EER). For the estimation of these performance measures on a specific system, data from benchmark corpora are usually divided into training and testing sets. Training data are used for computing the reference models in Fig. 1 and testing data are used to generate scores from which error rates are computed. This is the so-called *hold-out method* [17]. Other approaches that may be useful in small sample size situations have also been proposed [18]. For our discussion, the following error estimation methods are of interest.

- Resubstitution: all the available data is used for training as well as testing.
- Rotation: this is a version of *cross-validation* [18]. Regarding the available data for each target, the reference model is designed by choosing k consecutive samples as the design set, and the remaining samples constitute the test set; this is repeated for all distinct choices of k consecutive observations. When k is chosen to be equal to the number of samples minus one, the *leave-one-out* procedure is obtained [17].

III. Target Dependent Score Normalization Techniques

In the following sections, target dependent score normalization techniques are classified according to [8].

A. Impostor-Centric Methods

In Impostor-Centric methods (IC) no information about client score intra-variability is used. Therefore

$$s_{\mathrm{IC}} = f\left(s, \mathcal{I} = \left\{s_1^{\bar{\mathcal{T}}}, \ldots, s_{N_{\bar{\mathcal{T}}}}^{\bar{\mathcal{T}}}\right\}\right). \tag{7}$$

The following IC methods are considered in this work:

- $\mathrm{IC}-1$: $s_{\mathrm{IC}-1} = s - \mu_{\mathcal{I}}$
- $\mathrm{IC}-2$: $s_{\mathrm{IC}-2} = s - (\mu_{\mathcal{I}} + \sigma_{\mathcal{I}})$
- $\mathrm{IC}-3$: $s_{\mathrm{IC}-3} = (s - \mu_{\mathcal{I}})/\sigma_{\mathcal{I}}$

where $\mu_{\mathcal{I}}$ and $\sigma_{\mathcal{I}}$ are respectively the mean and standard deviation of the impostor scores $\mathcal{I}$. $\mathrm{IC}-1$ is proposed here as a robust technique for small sample size normalization problems [18], $\mathrm{IC}-2$ is equivalent to the *a priori* decision threshold setting described in [6], and $\mathrm{IC}-3$ is the well known z-norm technique [8].

Note that the impostor scores $\mathcal{I}$ can be, in general, from either casual impostors (in this case leading to a casual-Impostor-Centric method, $c\mathrm{IC}$) or from real impostors (similarly, leading to $r\mathrm{IC}$).

B. Target-Centric Methods

In Target-Centric methods (TC) no information about impostor score variability is used. Therefore

$$s_{\mathrm{TC}} = f\left(s, \mathcal{C} = \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}\right). \tag{8}$$

Similarly to the impostor-centric case, the following methods are obtained:

- $\mathrm{TC}-1$: $s_{\mathrm{TC}-1} = s - \mu_{\mathcal{C}}$
- $\mathrm{TC}-2$: $s_{\mathrm{TC}-2} = s - (\mu_{\mathcal{C}} - \sigma_{\mathcal{C}})$
- $\mathrm{TC}-3$: $s_{\mathrm{TC}-3} = (s - \mu_{\mathcal{C}})/\sigma_{\mathcal{C}}$

where $\mu_{\mathcal{C}}$ and $\sigma_{\mathcal{C}}$ are, respectively, the mean and standard deviation of the client scores $\mathcal{C}$. $\mathrm{TC}-1$ is based on the running average normalization strategy proposed in [19], $\mathrm{TC}-2$ is a form of the *a priori* decision thresholding technique proposed in [20], and $\mathrm{TC}-3$ is the normalization scheme proposed in [21].

Client scores $\mathcal{C}$ should be obtained from the available training set. In this work, we propose to generate $\mathcal{C}$ by using one of the sampling methods described in Section II-B, namely resubstitution (in this case leading to a resubstitution-Target-Centric method, resTC) or rotation (similarly, leading to rotTC). The former strategy leads to optimistically biased estimates whereas the later one gives unbiased estimates with larger computational requirements. These two resampling techniques have been considered for convenience regarding the experiments reported later on. Other techniques such as cross-validation or bootstrap should be also considered regarding other verification problems.

C. Target-Impostor Methods

In Target-Impostor methods (TI) information from both client score intra-variability and impostor score variability is used. Therefore

$$s_{\mathrm{TI}} = f\left(s, \mathcal{C} = \left\{s_1^{\mathcal{T}}, \ldots, s_{N_{\mathcal{T}}}^{\mathcal{T}}\right\}, \mathcal{I} = \left\{s_1^{\bar{\mathcal{T}}}, \ldots, s_{N_{\bar{\mathcal{T}}}}^{\bar{\mathcal{T}}}\right\}\right). \tag{9}$$

Based on the decision thresholding techniques in [7] and [22], we obtain the two following target-impostor normalization methods:

- $\mathrm{TI}-1$: $s_{\mathrm{TI}-1} = s - s_{\mathrm{EER}}(\mathcal{C}, \mathcal{I})$
- $\mathrm{TI}-2$: $s_{\mathrm{TI}-2} = s - (\mu_{\mathcal{I}}\sigma_{\mathcal{C}} + \mu_{\mathcal{C}}\sigma_{\mathcal{I}})/(\sigma_{\mathcal{I}} + \sigma_{\mathcal{C}})$

where $s_{\mathrm{EER}}(\mathcal{C}, \mathcal{I})$ is the target-dependent decision threshold at the empirical EER obtained from the histograms of $\mathcal{C}$ and $\mathcal{I}$. For the experiments reported below, the operational procedure for computing the EER proposed in [23] has been followed.

IV. Experiments

Various practical aspects of the normalization methods described above are explored in this section. Experiments are carried out by using the on-line signature verification system described in [12], [16] on SVC 2004 signature data [11].

A. On-Line Signature Verification System

For the experiments reported in this paper, the HMM-based on-line signature verification system from Universidad Politecnica de Madrid competing in the First International Signature Verification Competition (SVC 2004) has been used. Below we briefly describe the basics of the

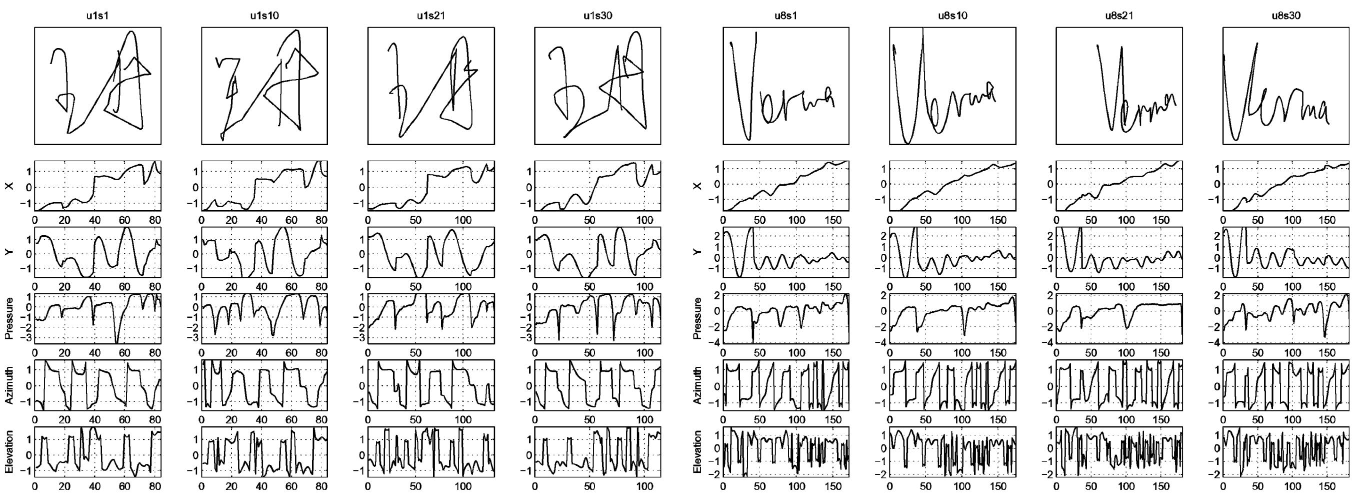

Fig. 3. Signature examples from SVC 2004 corpus. For each one of the targets u1 (left) and u8 (right), two genuine signatures (left columns) and two skilled forgeries (right columns) are given.

system. For further details we refer the reader to [12], [16] and the references therein.

1) Feature Extraction: Only coordinate trajectories $(x[n], y[n])$, $n = 1, \ldots, N_s$ and pressure signal $p[n], n = 1, \ldots, N_s$ are considered in the feature extraction process, where N_s is the duration of the signature in time samples. Signature trajectories are first preprocessed by subtracting the center of mass followed by a rotation alignment based on the average path tangent angle. An extended set of discrete-time functions are derived from the preprocessed trajectories consisting of sample by sample estimations of various dynamic properties. As a result, the signature is parameterized as the following set of seven discrete-time functions $\{x[n], y[n], p[n], \theta[n], v[n], \rho[n], a[n]\}, n = 1, \ldots, N_s$, and first order time derivatives of all of them (θ, v, ρ, and a stand, respectively, for path tangent angle, path velocity magnitude, log curvature radius, and total acceleration magnitude). A whitening linear transformation is finally applied to each discrete-time function so as to obtain zero mean and unit standard deviation function values.

2) Similarity Computation: Given the parameterized enrollment set of signatures of a client $\mathcal{T}$, a left-to-right Hidden Markov Model $\lambda^{\mathcal{T}}$ is estimated by using the Baum-Welch iterative algorithm [24]. No transition skips between states are allowed and multivariate Gaussian mixture density observations are used (2 states and 32 mixtures per state). On the other hand, given a test signature X parameterized as P (with a duration of N_s time samples) and a claimed identity $\mathcal{T}$ modeled as $\lambda^{\mathcal{T}}$, the similarity matching score

$$s = \frac{1}{N_s} \log p(P|\lambda^{\mathcal{T}}) \tag{10}$$

is computed by using the Viterbi algorithm [24].

B. Database Description

There are not many signature databases publicly available at the moment for research purposes [25]. As a result, the common practice in on-line signature recognition research is to evaluate the proposed recognition strategies on small data sets acquired at the different research laboratories [5]. In this environment, the First International Signature Verification Competition (SVC 2004) has been organized [11] providing a common reference for system comparison on the same signature data and evaluation protocol.

Development corpus of the extended task (including coordinate and timing information, pen orientation and pressure) is used in the experiments that follow. This corpus consists of 40 sets of signatures. Each set contains 20 genuine signatures from one contributor (acquired in two separate sessions) and 20 skilled forgeries from five other contributors. The signatures are mostly in either English or Chinese. Some examples are depicted in Fig. 3 for two different targets of the data set. Plots of the coordinate trajectories, pressure signal, and pen orientation functions are also given. The highly skilled nature of the signature forgeries is remarkable.

C. Experimental Procedure

Signature data from the two acquisition sessions in SVC 2004 are used both for training and testing. In case of training, five random genuine signatures from both sessions are used for each target. In case of testing, the remaining 15 genuine signatures are used. For a specific target user, casual impostor test scores are computed by using signatures from all the remaining targets and real impostor test scores are computed by using the 20 skilled forgeries of each target. Impostor scores for error estimation are computed in a leave-one-out fashion; i.e., testing signatures from one impostor with a normalization scheme estimated with statistics from the remaining impostors and rotating the scheme. Score normalization results are provided using statistics either from casual or real impostors for the computation of the normalization functions.

A priori score normalization methods are compared in the experiments. This means that only the information from the training set is used both for the enrollment of the targets and for the estimation of the parameters of the normalization functions (by using the resampling techniques described in Section II-B). In order to have an indication of the level of performance with an ideal score alignment between targets, results using the target dependent score normalization $\mathrm{TI}-1$ (see Section III-C) *a posteriori* are also given. Only in this case, test information is used both for error estimation and for the computation of the normalization functions.

For the comparison of competing algorithms in SVC 2004, FR and FA rates for different threshold values, EER, and ROC curves are used [11]. A variant of these, the so-called Detection Error Tradeoff (DET) plot [26], is used in the experiments reported below; in this case, the use of a normal deviate scale makes the comparison of competing systems easier. For the computation of EERs, the operational procedure described in [23] has been used.

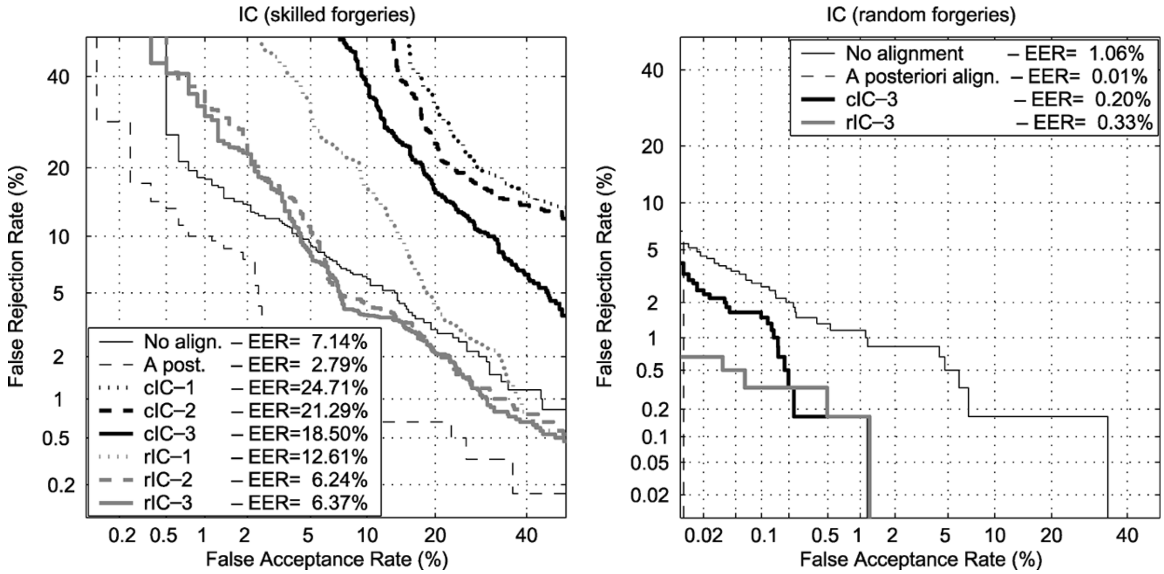

Fig. 4. Verification performance for various impostor-centric normalization methods.

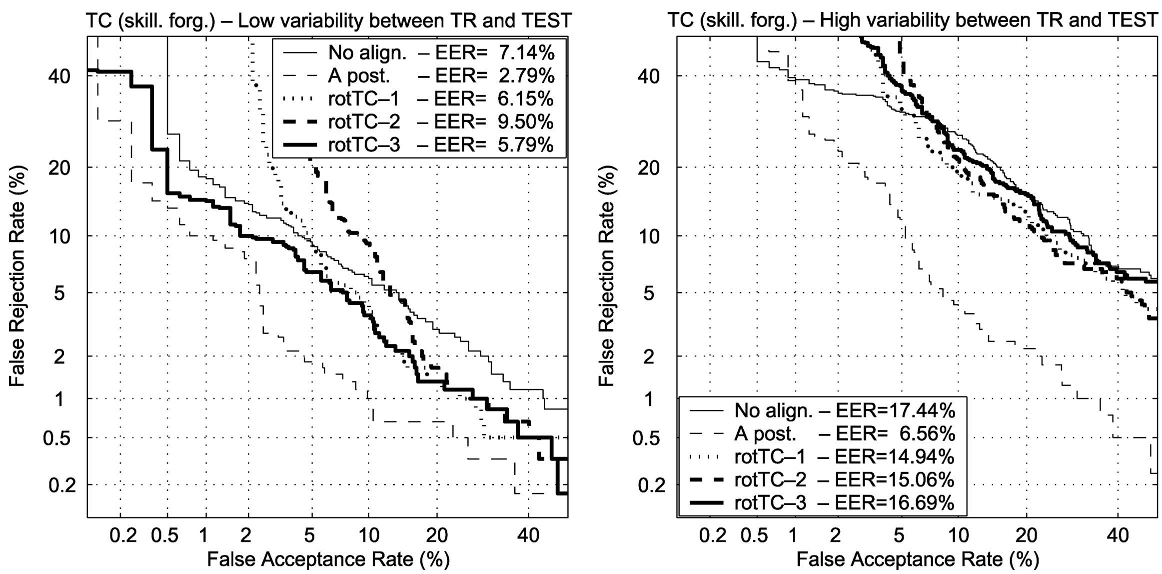

Fig. 5. Verification performance for various target-centric normalization methods. Effects of variability between training and testing on verification performance.

D. Results

1) Impostor-Centric Methods: In the first experiment summarized in Fig. 4 the different impostor-centric methods described in Section III-A are compared, testing either with skilled (left) or random (right) forgeries.

Raw verification performance with no normalization (7.14% and 1.06% EER for skilled and random forgeries, respectively) is significantly improved by the *a posteriori* normalization scheme (2.79% and 0.01%, respectively). This corroborates the preliminary experiment in Fig. 2 and will be used as the reference for further comparisons.

Regarding the test with skilled forgeries, *a priori* method $\mathrm{IC}-3$ outperforms $\mathrm{IC}-1$ and $\mathrm{IC}-2$. Raw performance is only improved in this case by considering statistics from real impostors $(r\mathrm{IC})$. This means that testing with skilled forgeries, while using statistics from random impostors for estimating the normalization functions, leads to degraded performance as compared to not using score normalization at all.

Regarding the test with random forgeries, significant improvements are obtained considering statistics either from casual $(c\mathrm{IC})$ or from real impostors $(r\mathrm{IC})$.

2) Target-Centric Methods: The effects of using different training strategies on the verification performance are summarized in Fig. 5. We consider two training strategies: a) training data from both acquisition sessions and testing data from both sessions (left) and b) training data from the first session and testing data from the second session (right), as in SVC 2004 protocol. With regard to the second strategy, raw verification performance drops significantly and the robust score normalization method $\mathrm{TC}-1$ outperforms the others. The multisession training and testing strategy is used in all other experiments.

Results using different resampling techniques for the estimation of target score variability are summarized in Fig. 6 for three different verification systems of decreasing verification performance (from left to right). For the rotation scheme (rotTC), $k = 3$ is used (see Section II-B), so 2×5 target scores are considered for the computation of

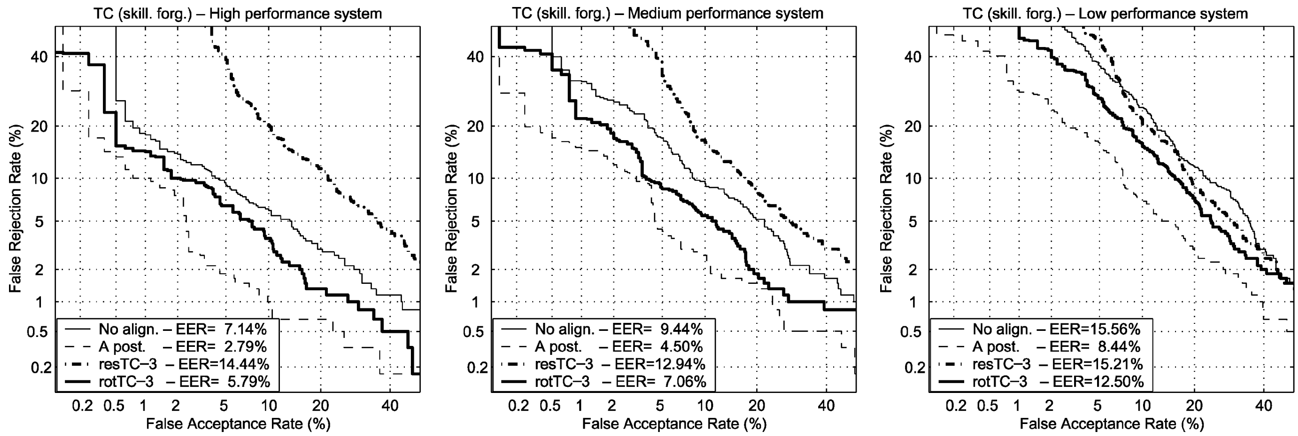

Fig. 6. Verification performance for various target-centric normalization methods. Effects of target score variability estimation on verification performance.

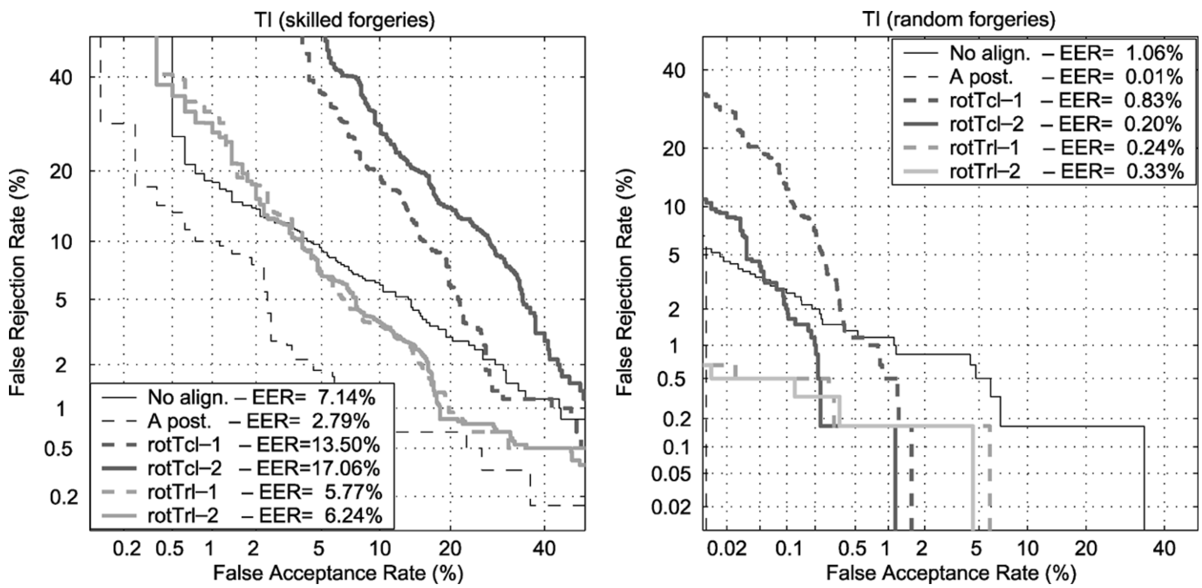

Fig. 7. Verification performance for various target-impostor normalization methods.

the statistics involved in the computation of the normalization function. In the case of resubstitution, 5 target scores are used. As can be observed, the rotation scheme always leads to verification improvements whereas the resubstitution strategy only leads to improvements in the low performance system. This result penalizes the biased estimation provided by the resubstitution scheme in favor of the unbiased rotation procedure.

3) Target-Impostor Methods: Verification performance for the target-impostor methods described in Section III-C is shown in Fig. 7. In the case of tests with skilled forgeries, only the use of target-impostor normalization schemes based on real impostor statistics (rotTrI) leads to performance improvements as compared to not using score normalization. With regard to the test with random forgeries, verification performance improvements are obtained considering either casual or real impostor statistics for the computation of the normalization functions.

4) Summary: Results using a selection of practical *a priori* normalization methods following SVC 2004 guidelines (i.e., using real impostor statistics for the computation of the normalization functions is not permitted) are summarized in Fig. 8. In this case, only the target-centric method is capable of performance improvements testing both with skilled and random forgeries.

V. Conclusion

Score normalization methods in biometric verification, which encompass the more traditional user-dependent decision thresholding techniques, have been reviewed from a test hypotheses point of view. These have been classified into test dependent and target dependent methods. The focus of the paper has been on target dependent score normalization techniques, further classifying them into impostor-centric, target-centric, and target-impostor methods.

The techniques described have been applied to the on-line signature verification system ranked first and second in SVC 2004 for random and skilled forgeries, respectively (1.06% EER and 7.14% EER, respectively, on SVC 2004 development data with multisession training). Various experimental findings have been obtained. Most remarkably, target-centric techniques based on a variation of the cross-validation procedure provided the best performance improvement testing both with random (0.50% EER) and skilled (5.79% EER) forgeries.

Other experimental findings worth noting are that: 1) the use of casual impostor statistics for estimating the normalization functions leads to the highest performance improvement when testing with random forgeries but lowers verification performance in case of testing against skilled forgeries; 2) the use of real impostor statistics leads to verification performance improvements when testing either with random or

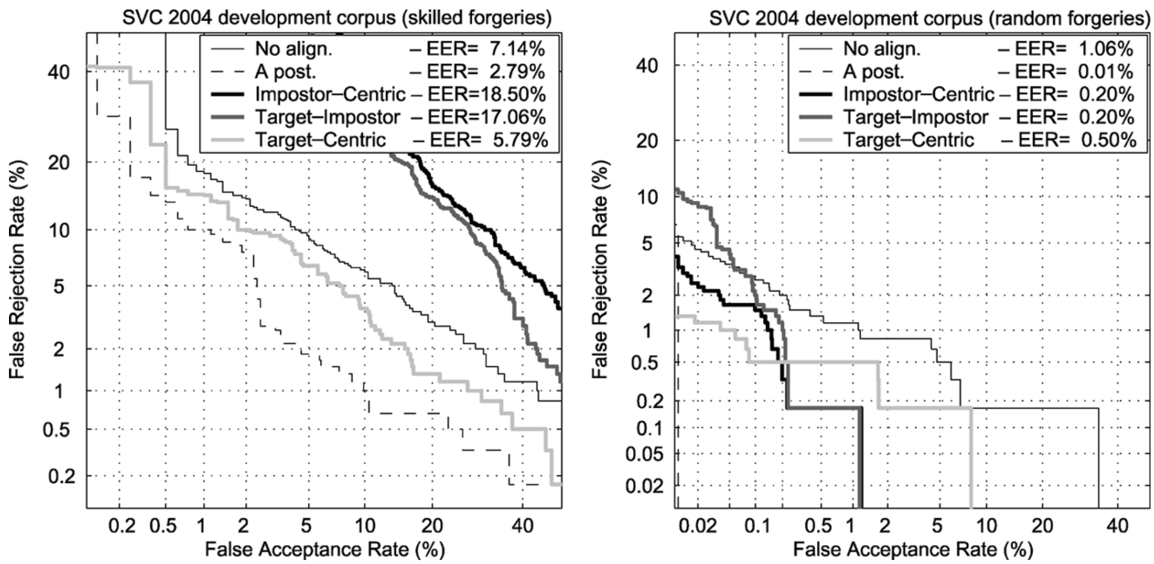

Fig. 8. Verification performance for various target dependent normalization methods on SVC 2004 development corpus.

skilled forgeries; 3) when high variability between sessions is encountered, robust score normalization methods (as those based only on mean scores) should be used; and 4) resampling methods for the estimation of target score intra-variability should be unbiased.

Future work includes the review of test dependent score normalization techniques [8] and their application to the described on-line signature verification system.

ACKNOWLEDGMENT

The authors would like to thank SVC 2004 organizer D.-Y. Yeung for providing signature data for the experiments. J. Fierrez-Aguilar also thanks Consejeria de Educacion de la Comunidad de Madrid and Fondo Social Europeo for supporting his doctoral research.

REFERENCES

[1] A. Jain, R. Bolle, and S. Pankanti, *Biometrics—Personal Identification in a Networked Society.*. Norwell, MA: Kluwer, 1999.

[2] A. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 14, no. 1, pp. 4–20, 2004.

[3] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification—The state of the art," *Pattern Recognit.*, vol. 22, no. 2, pp. 107–131, 1989.

[4] M. Paulik, N. Mohankrishnan, and M. Nikiforuk, "A time varying vector autoregressive model for signature verification," in *Proc. 37th Midwest Symp. Circuits Syst.*, vol. 2, 1994, pp. 1395–1398.

[5] A. Jain, F. Griess, and S. Connell, "On-line signature verification," *Pattern Recognit.*, vol. 35, no. 12, pp. 2963–2972, 2002.

[6] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 29, no. 2, pp. 254–272, 1981.

[7] T. Matsui, T. Nishitani, and S. Furui, "Robust methods of updating model and *a priori* threshold in speaker verification," in *Proc. IEEE Intl. Conf. Acoustics, Speech Signal Process., ICASSP*, 1996, pp. 97–100.

[8] R. Auckenthaler, M. Carey, and H. Lloyd-Tomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Process.*, vol. 10, pp. 42–54, 2000.

[9] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D.-A. Reynolds, "A tutorial on text-independent speaker verification," *J. Appl. Signal Process.*, vol. 2004, no. 4, pp. 430–451, 2004.

[10] J. Navratil and G. Ramaswamy, "The awe and mystery of t-norm," in *Proc. ESCA Eur. Conf. Speech Comm. Tech., EuroSpeech*, 2003, pp. 2009–2012.

[11] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. , and G. Rigoll, "SVC2004: First international signature verification competition," in *Proc. Intl. Conf. Biometric Authentication, ICBA*, 2004, pp. 16–22.

[12] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, A function-based on-line signature verification system exploiting statistical signal modeling, in Intl. J. Pattern Recognit. Artif. Intell., 2005. submitted.

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

[14] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Process.*, vol. 10, pp. 19–41, 2000.

[15] C. Sanderson and K. K. Paliwal, "Likelihood normalization for face authentication in variable recording conditions," in *Proc. IEEE Intl. Conf. Image Process., ICIP*, vol. 1, 2002, pp. 301–304.

[16] J. Ortega-Garcia, J. Fierrez-Aguilar, J. Martin-Rello, and J. Gonzalez-Rodriguez, "Complete signal modeling and score normalization for function-based dynamic signature verification," in *Proc. IAPR Intl. Conf. Audio- and Video-Based Person Authentication, AVBPA*, 2003, pp. 658–667.

[17] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, 2000.

[18] S. Raudys and A. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 252–264, 1991.

[19] J. Naik and G. Doddington, "High performance speaker verification using principal spectral components," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Process., ICASSP*, 1986, pp. 881–884.

[20] J. Saeta and J. Hernando, "Automatic estimation of a priori speaker dependent thresholds in speaker verification," in *Proc. IAPR Intl. Conf. Audio- and Video-Based Person Authentication, AVBPA*, 2003, pp. 70–77.

[21] D. Garcia-Romero, J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, and J. Ortega-Garcia, "U-norm likelihood normalization in PIN-based speaker verification systems," in *Proc. IAPR Intl. Conf. Audio- and Video-Based Person Authentication, AVBPA*, 2003, pp. 208–213.

[22] D. Burton, "Text-dependent speaker verification using vector quantization source coding," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 35, no. 2, pp. 133–143, 1987.

[23] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2000: Fingerprint verification competition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 402–412, 2002.

[24] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[25] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "MCYT baseline corpus: A bimodal biometric database," in *IEE Proc. Vision, Image Signal Process.*, vol. 150, 2003, pp. 395–401.

[26] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of decision task performance," in *Proc. ESCA Eur. Conf. Speech Comm. Tech., EuroSpeech*, 1997, pp. 1895–1898.

Hallucinating Face by Eigentransformation

Xiaogang Wang and Xiaoou Tang

***Abstract*—In video surveillance, the faces of interest are often of small size. Image resolution is an important factor affecting face recognition by human and computer. In this paper, we propose a new face hallucination method using eigentransformation. Different from most of the proposed methods based on probabilistic models, this method views hallucination as a transformation between different image styles. We use Principal Component Analysis (PCA) to fit the input face image as a linear combination of the low-resolution face images in the training set. The high-resolution image is rendered by replacing the low-resolution training images with high-resolution ones, while retaining the same combination coefficients. Experiments show that the hallucinated face images are not only very helpful for recognition by humans, but also make the automatic recognition procedure easier, since they emphasize the face difference by adding more high-frequency details.**

***Index Terms*—Eigentransformation, face hallucination, face recognition, principal component analysis, super-resolution.**

I. Introduction

In video surveillance, the faces of interest are often of small size because of the large distance between the camera and the objects. Image resolution becomes an important factor affecting face recognition performance. Since many detailed facial features are lost in the low-resolution face images, the faces are often indiscernible. For identification, especially by humans, it is useful to render a high-resolution face image from the low-resolution one. This technique is called face hallucination or face super-resolution [1].

The simplest way to increase image resolution is a direct interpolation of input images with such algorithms as nearest neighbor or cubic spline. However, the performance of direct interpolation is usually poor since no new information is added in the process. A number of super-resolution techniques have been proposed in recent years [2]–[13]. Most try to produce a super-resolution image from a sequence of low-resolution images [2], [3]. Some other approaches [5], [6], [8]–[10], [12], [14], [15] are based on learning from the training set containing high- and low-resolution image pairs, with the assumption that high-resolution images are Markov random fields (MRFs) [5], [9], [13]. These methods are more suitable for synthesizing local texture, and are usually applied to generic images without special consideration of the property of face images.

Baker and Kanade [1], [11], [16] developed a hallucination method based on the property of face images. Abandoning the MRF assumption, it infers the high-frequency components from a parent structure by recognizing the local features from the training set. Liu*et al.* [4] developed a two-step statistical modeling approach integrating global and local parameter models. Both of the two methods use complicated probabilistic models and are based on an explicit resolution reduction function, which is sometimes difficult to obtain in practice.

Since face images are well structured and have similar appearances, they span a small subset in the high dimensional image space [17], [18]. In the study by Penev and Sirovich [19], face images are shown to be well reconstructed by Principal Component Analysis (PCA) representation with 300–500 dimensions. Zhao*et al.* [20] showed that the dimensionality of face space is insensitive to image size. Moghaddam [21] downsampled face images to 12 by 21 pixels and still achieved 95% recognition accuracy on $1800+$ face images from the FERET database. These studies imply that facial components are highly correlated and the high-frequency details of face images may be inferred from the low-frequency components, utilizing the face structural similarities.

Instead of using a probabilistic model, we propose a face hallucination method using PCA to represent the structural similarity of face images. The algorithm treats the hallucination problem as the transformation between two different image styles. This method is closely related to the work in [22], [23], in which a style transformation approach was applied to transform a photo into a sketch. In a similar way, we could transform face images from low-resolution to high-resolution based on mapping between two groups of training samples without deriving the transformation function [24]. The hallucinated face image is rendered from the linear combination of training samples. Using a small training set, the method can produce satisfactory results.

Hallucination can effectively improve the resolution of a face image, thus making it much easier for a human being to recognize a face. However, how much information has been extracted from the low-resolution image by the hallucination process and its contribution to automatic face recognition has not been studied before. In our method, PCA is applied to the low-resolution face image. In the PCA representation, different frequency components are uncorrelated. By selecting the number of eigenfaces, we could extract the maximum amount of facial information from the low-resolution face image and remove the noise. We also study the face recognition performance using different image resolutions. For automatic recognition, a low resolution bound is found through experimentation. We find that hallucination may help the automatic recognition process, since it emphasizes the face difference by adding some high-frequency details.

II. Hallucination by Eigentransformation

A. Multiresolution Analysis

Viewing a two-dimensional (2-D) image as a vector, the process of getting a low-resolution face image from the high-resolution one can be formulated as

$$\vec{I}_l = H\vec{I}_h + \vec{n}. \tag{1}$$

Here, $\vec{I}_h$ is the high-resolution face image vector to be rendered, with length N as the total pixel number. $\vec{I}_l$ is the observed low-resolution face image vector with length s^2N, where s is the downsampling factor

Manuscript received March 15, 2003; revised July 13, 2004. This work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK 4190/01E, CUHK 4224/03E, AOE/E01/99, and N_CUHK409/03). This paper was recommended by Guest Editor D. Zhang.

The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: xgwang1@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

Digital Object Identifier 10.1109/TSMCC.2005.848171