Biometric hash: High-confidence face recognition

Article *in* IEEE Transactions on Circuits and Systems for Video Technology · June 2006

DOI: 10.1109/TCSVT.2006.873780 · Source: DBLP

David Ngo
Sunway University
64 PUBLICATIONS **3,556** CITATIONS

Andrew Beng Jin Teoh
Yonsei University
376 PUBLICATIONS **8,859** CITATIONS

Alwyn Goh
Malaysian Institute of Microelectronic Systems
46 PUBLICATIONS **1,978** CITATIONS

Transactions Letters

Biometric Hash: High-Confidence Face Recognition

David C. L. Ngo, *Member, IEEE*, Andrew B. J. Teoh, and Alwyn Goh

***Abstract*—In this paper, we describe a biometric hash algorithm for robust extraction of bits from face images. While a face-recognition system has high acceptability, its accuracy is low. The problem arises because of insufficient capability of representing features and variations in data. Thus, we use dimensionality reduction to improve the capability to represent features, error correction to improve robustness with respect to within-class variations, and random projection and orthogonalization to improve discrimination among classes. Specifically, we describe several dimensionality-reduction techniques with biometric hashing enhancement for various numbers of bits extracted. The theoretical results are evaluated on the FERET face database showing that the enhanced methods significantly outperform the corresponding raw methods when the number of extracted bits reaches 100. The improvements of the postprocessing stage for principal component analysis (PCA), Wavelet Transform with PCA, Fisher linear discriminant, Wavelet Transform, and Wavelet Transform with Fourier–Mellin Transform are 98.02%, 95.83%, 99.46%, 99.16%, and 100%, respectively. The proposed technique is quite general, and can be applied to other biometric templates. We anticipate that this algorithm will find applications in cryptographically secure biometric authentication schemes.**

***Index Terms*—Biometric cryptography, face recognition, random projection.**

I. Biometric Key Cryptography

BIOMETRIC ergonomics and cryptographic security are highly complementary, hence the motivation for their integrated application. Known methods for generating cryptographic keys from biometric measurements can be characterized as having two stages. In the first stage, features of raw input are used to compute a bit string. The second stage develops a cryptographic key from the bit string. If two bit strings are sufficiently similar, then the same cryptographic key will be generated from them.

Several techniques fitting this two-stage structure have been proposed for generating cryptographic keys from biometrics. Davida *et al.* [1] describe a second-stage strategy using majority decoding and error-correcting codes, and how it could be conjoined with first-stage approaches for generating bit strings from iris scans [2]. Soutar *et al.* [3] describe a different approach for generating cryptographic keys from fingerprints using optical computing techniques. Monrose *et al.* [4] describe a mechanism for developing a cryptographic key, called a hardened password, from a text password and the keystroke dynamics of the user while typing the password.

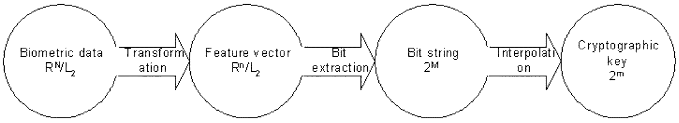


Fig. 1. Biometric hashing scheme.

In this paper, we describe the first scheme (to our knowledge) for generating biometric hash values from a face image. Hash values are of great interest for applications that require key management strategies, as cryptographic keys can be derived directly from the hash value. The current techniques, however, are inappropriate for our goals because they either require a biometric with high accuracy (as in the iris scheme of Davida *et al.* [1] and the fingerprint scheme of Soutar *et al.* [3]) or depend on the presentation of a one-dimensional (1-D) signal (as in the keystroke dynamics scheme of Monrose *et al.* [4]).

A face-recognition system does not handle poor-quality input as well as other physical biometric systems like fingerprint, palmprint, and iris systems [5]. For this reason, even though it has high acceptability, its accuracy is low [6]. The problem arises mainly from three factors: insufficient capability of representing features in the feature space and within-class and between-class variations. Despite these challenges, in this paper, we apply dimensionality reduction to improve the capability to represent features, error correction to improve robustness with respect to within-class variations, and random projection (RP) and orthogonalization to improve discrimination among classes to the task of generating a biometric hash from a face image. Our goal is to design a procedure for extracting bits from face images so that all similar-looking images will produce almost the same bit sequence. Experimental results show high-confidence recognition of persons.

The second stage is not the focus in this paper, as it was previously covered in [7]. While our techniques could be conjoined with any of several second-stage techniques to generate cryptographic keys, we evaluate our scheme mainly with a new technique using Shamir's secret sharing scheme [8].

II. Biometric Hashing Scheme

The process is illustrated in Fig. 1. Our approach begins with the steps for normalizing a face image. Next is to construct a bit string from this normalized face image. The resulting bit string

Manuscript received January 6, 2005. This paper was recommended by Associate Editor K. Aizawa.

D. C. L. Ngo and A. J. Teoh are with the Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia (e-mail: david.ngo@mmu.edu.my; bjteoh@mmu.edu.my).

A. Goh is with Corentix Technologies, 47301 Petaling Jaya, Malaysia (e-mail: alwyn@corentix.com).

Digital Object Identifier 10.1109/TCSVT.2006.873780

will be used to reconstruct the derived cryptographic key, provided that the bit string is sufficiently close to the enrolled one. The bit-extraction technique extends the previous technique [9] by projecting the transformed image onto random patterns and thresholding the result. This technique extends to biometric data which is unstable in nature.

III. Linear Orthogonal Projection

The Johnson–Lindenstrauss (JL) result [10] states that Euclidean distances are retained well in RP; RP is beneficial in applications where the distances of the original high-dimensional data points are meaningful. This study is complicated by the fact that image data is typically present in a raw, uncompressed form. Thus, by applying appropriate dimensionality reduction first and then running RP on the reduced space, we can minimize distortions in the data set.

Linear projection is one of the approaches to cope with the problem of excessive dimensionality. Linear methods have the advantages in that the computation is simple and analytically tractable. The methods used in this paper are based on orthogonal transforms including principal component analysis (PCA) [11], Fisher linear discriminant (FLD) [12], Wavelet Transform (WT) [13], Wavelet Transform with PCA (WT_PCA) [14], and Wavelet Transform with Fourier–Mellin Transform (WT_FMT) [15]. (See the references for more details.)

IV. Random Thresholding Projection

In RP, the original high-dimensional data is projected onto a lower dimensional subspace using a random matrix whose columns have unit lengths. The key idea of random mapping arises from the JL lemma [10]: if points in a vector space are projected onto a randomly selected subspace of suitably high dimension, then the distances between the points are approximately preserved.

RP is used as a further dimensionality-reduction tool for projecting random, key-dependent vectors onto principal directions of images, using the robustness of those principal directions for bit extraction. The bits are extracted by thresholding projections onto random vectors generated from a user-specified key. The method presented here is an extension and improvement of the robust bit-extraction method proposed in [9]. We base our method on the observation that the absolute value of a transform coefficient cannot be altered without causing visible change to the image. To make the procedure key-dependent, we combine transform coefficients with random vectors generated from a secret key. These projections are then thresholded and quantized in order to remove subjective redundancy in the image. This will improve robustness with respect to within-class variations. To maximize the information content of the extracted bits, we calculate a threshold μ, so that, on average, half of the bits are zeros and half are ones.

All known constructions of JL embeddings involve projecting a set of n points in d-dimensional Euclidean space onto a random k-dimensional Euclidean space, where k is logarithmic in n and independent of d. Our construction of such embeddings involves projecting different sets of n points onto the same number of random k-dimensional subspaces. The main result, in the next section, asserts that Hamming distance comparisons between points of different subspaces are binomially distributed, with parameters mean $p = 0.5$ and $N \approx m$ degrees of freedom (DOFs). Therefore, we expect the chance of agreement between two bits for different faces is 50/50. The main idea is to fix the mean of all distributions at 0.5. Theoretically, two codes are orthogonal if exactly one half of their bits are different. Furthermore, two subspaces are orthogonal if all the codes in one are orthogonal to all the codes in the other. This implies that the randomly selected subspaces are orthogonal to each other.

Since the random vectors used in the bit-extraction technique are not orthogonal to each other, the extracted bits are correlated. Because of correlations within codes, the number of DOFs is considerably smaller than the number of bits extracted. One solution would be to use the Gram–Schmidt orthogonalization procedure and make the random vectors orthogonal.

Let α be a projected vector (for example, a PCA-based representation of the biometric input whose entries are floating-point numbers) in R^{n_c} (respectively, $n \times n$ matrix). Let k_t be a number uniquely associated with a token.

1) Use k_t to compute n_c random vectors in n_c-space (respectively, $n \times n$ matrices) with real number entries normally distributed in the interval $[-1, 1]$

$$\{\beta_1, \ldots, \beta_{n_c}\}.$$

Such vectors may be generated by means of a secret password or serial number (associated with a physical token) as a cryptographic key or initial condition. There are various pseudorandom generators, including the ANSI X9.17 standard, in the public domain which can be used for this purpose.

2) Apply the Gram–Schmidt process to transform the basis $\{\beta_1, \ldots, \beta_{n_c}\}$ into an orthonormal set of vectors (respectively, matrices) $\{\chi_1, \ldots, \chi_{n_c}\}$.

3) Compute $(\langle \alpha, \chi_1 \rangle, \langle \alpha, \chi_2 \rangle, \ldots, \langle \alpha, \chi_{n_c} \rangle)$ using an inner product on R^{n_c} (respectively, M_{nn}).

4) Compute n_c bits b_i. Let μ be a threshold such that on average, half of the bits are zeros and half are ones.

$$b_i = \begin{cases} 0, & \text{if } \langle \alpha, \chi_i \rangle \leqslant \mu \\ 1, & \text{if } \langle \alpha, \chi_i \rangle > \mu. \end{cases}$$

V. Recognition Experiment

A. Experiment Setup

Our version of the FERET database [16] contained 1000 frontal face images (5 shots of 200 individuals) taken over a period of a few years under varying lighting conditions. The frontal face images include the whole head, parts of the shoulder and neck, and background. Instead of training on the whole images, which contain much irrelevant information, we trained on face images that were preprocessed by masking them in windows of 100×200 pixels, placing the several face features in the same relative places. Here, a feature-location method that progressively reduces the search area is developed. We first locate the face region, and then find possible locations of the features including eyes and mouth (Haar-like features).

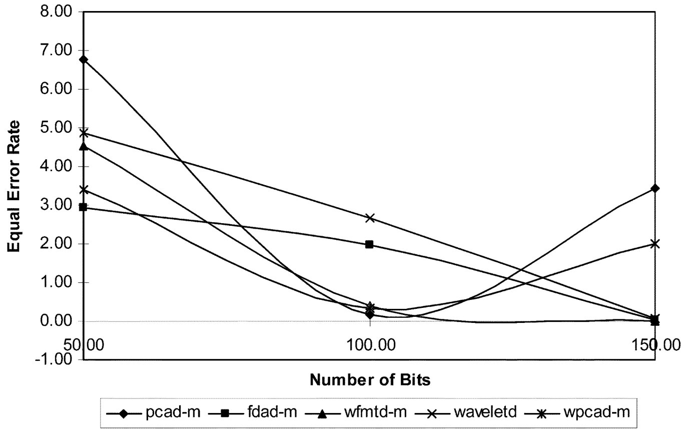


Fig. 2. EER as a function of the number of bits extracted.

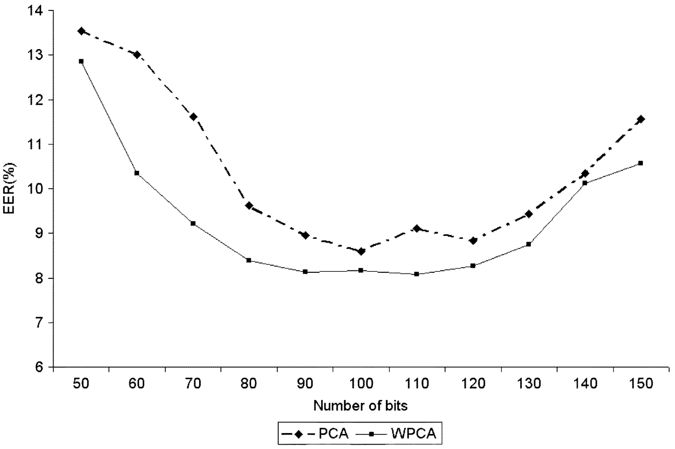


Fig. 3. EER as a function of the number of retained eigenvalues.

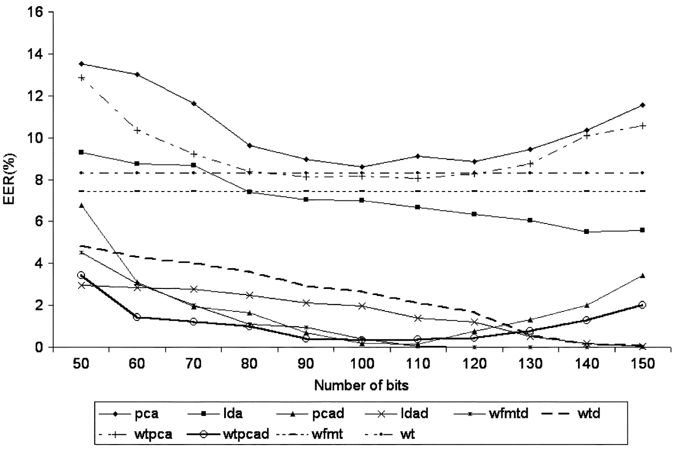


Fig. 4. Line plot of EERs of the BH, compared with other methods.

Finally, a local search is used to locate the exact positions (integral projection). Each face image is normalized using these three points by an affine deformation.

The experiments were conducted following a left-one-out principle [17]. Alternatively, each person is labeled as an impostor, while the other 199 are considered as clients. For each combination, 4 shots of the 199 clients build the training set, while the fifth shot is used as a testing set. That is, each client attempts to access under its own identity and the impostor attempts to access under the identity of the 199 clients, adding up to 199 authentic tests and 199 impostor tests. This procedure is repeated five times, by successively considering each shot for testing. In total, the client and impostor verifications amount to $200 \times 5 \times (199 + 199) = 398\,000$.

In this paper, Spline Bior 5.5 is adopted for image decomposition. In fact, in order to select a suitable wavelet, the error rates are computed by applying different wavelets (including Haar, Db4, Db8, Sym4, Sym8, Spline Bior 1.1, and Spline Bior 5.5) on face-image decomposition. Spline Bior 5.5 performs better than others, and hence, it is adopted for image decomposition in our system.

B. Method Evaluation

Fig. 2 shows the equal error rate (EER) for the biometric hashing (BH) for five popular face-recognition techniques: PCA [11], FLD [12], WT [13], WT_PCA [14], and WT_FMT [15].

The performance of the system increases with the number of bits extracted, until a minimum is reached at about 100 bits for PCA_BH, WT_PCA_BH, and WT_FMT_BH, and at about 150 bits for FLD_BH and WT_BH.

The performance of PCA_BH and WT_PCA_BH increases until a minimum, at which the curves diverge. This pattern is also observed in the equal error curves as a function of the number of eigenvectors retained for PCA and WT_PCA (Fig. 3). In each case, the minimal equal error value occurs at 100 eigenvectors and is associated in the corresponding BH equal error curve with 100 bits. Note that PCA (respectively, WT_PCA) is equivalent to PCA_BH (respectively, WT_PCA_BH) when the number of eigenvectors equals the size of the bit string, and since performance increases with the dimension of the eigenspace, PCA (respectively, WT_PCA) should do no better than PCA_BH (respectively, WT_PCA_BH).

For WT_BH and FLD_BH, the equal error decrease is much less than that observed in PCA_BH, WT_PCA_BH and WT_FMT_BH.

Fig. 4 and Table I present the EERs of the individual techniques and their extension. It can be observed that:

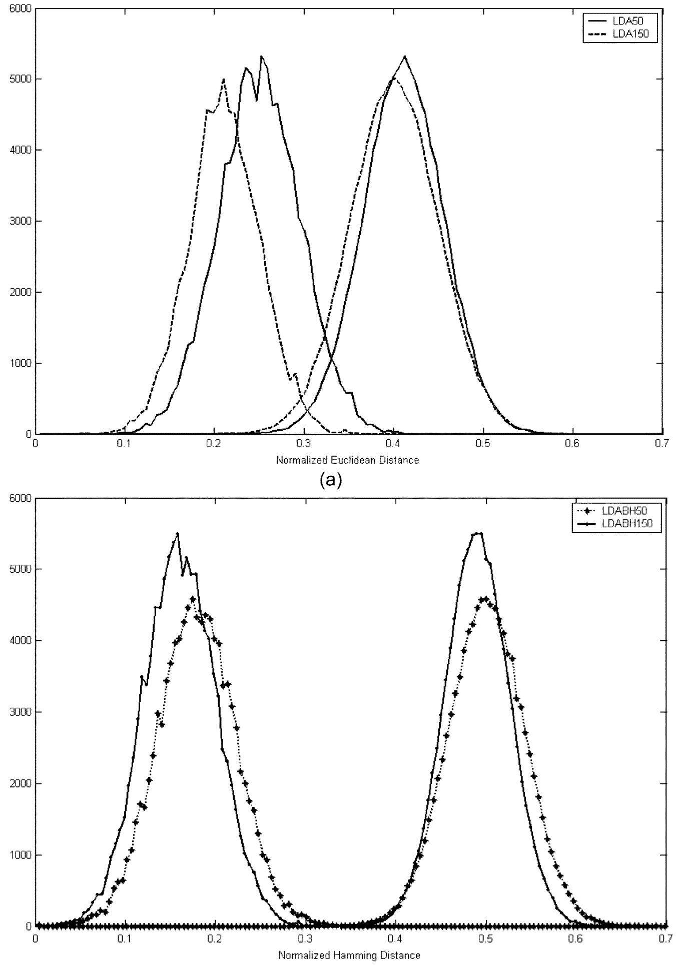
1) All the EERs are reduced. The improvements for PCA_BH, WT_PCA_BH, FLD_BH, WT_BH, and WT_FMT_BH are 98.02%, 95.83%, 99.46%, 99.16%, and 100%, respectively. The best results are obtained with FLD_BH, where the EER is reduced from 5.59% to 0.03%.
2) The postprocessing stage significantly improves the performance of the raw methods. Table I shows that the average EER is improved from 7.614% down to 0.122%. However, combining the worst transform with bit extraction is better than the best transform without bit extraction. Not surprisingly, combining the best transform with bit extraction yielded the best results.
3) A small face code of about 150 bits is sufficient in achieving high recognition rates for all the cases. This is much smaller than previous reports [1]–[3]. While the proposed method pursues a small representation of the faces, it is not necessarily with poor discrimination capability between different faces. On the contrary, the bit extraction is able to provide almost 100% verification performance for $b > 100$ for PCA and WT_PCA and $b > 150$ for WT, WT_FMT, and FLD.

TABLE I
EERs of the BH, Compared with Other Methods

	P/K/m	FAR	FRR	EER	FRR(FAR=0)
PCA(P)	50.00	16.80	10.26	13.53	64.10
	100.00	9.51	7.69	8.60	53.85
	150.00	12.89	10.26	11.57	57.42
FDA(K)	50.00	8.30	10.26	9.28	66.67
	100.00	7.11	6.91	7.01	53.85
	150.00	5.91	5.27	5.59	38.46
PCA_BH(m)	50.00	5.87	7.69	6.78	43.59
	100.00	0.34	0.00	0.17	7.69
	150.00	3.31	3.56	3.43	17.95
FDA_BH(m)	50.00	3.40	2.50	2.95	9.23
	100.00	1.41	2.50	1.96	3.53
	150.00	0.06	0.00	0.03	0.06
WT_FMT	-	8.12	6.72	7.42	51.28
WT_FMT_BH(m)	50.00	3.91	5.13	4.52	9.23
	100.00	0.81	0.01	0.41	2.56
	150.00	0.00	0.00	0.00	0.00
WT	-	8.13	8.47	8.30	43.59
WT_BH(m)	50.00	4.59	5.13	4.86	28.21
	100.00	2.77	2.56	2.67	12.82
	150.00	0.14	0.00	0.07	7.69
WT_PCA(P)	50.00	12.89	12.82	12.85	66.67
	100.00	8.64	7.69	8.16	46.15
	150.00	10.86	10.26	10.56	51.28
WT_PCA_BH(m)	50.00	4.25	2.56	3.41	30.77
	100.00	0.67	0.00	0.34	2.56
	150.00	1.42	2.56	1.99	7.69

4) Among the raw methods, PCA provides the poorest results. Theoretically, the PCA provides the best orthogonal description of normalized face images in the sense that the reconstruction from PCA coefficients yields, on average, the smallest sum of squared errors [17]. This does not assure that PCA is the best discrimination approach. Similar to previous findings [13]–[15], WT, WT_PCA, and WT_FMT provide significantly better performance in our tests. Still, the best results are obtained with FLD. A similar effect has been reported by Swets and Weng [12], where FLD showed improved performance over PCA.

However, error rates do not convey the full story, and it is instructive to view the distribution of Hamming distances and examine the margins of separation. Fig. 5(a) shows a histogram of the within-class and between-class Euclidean distances for FDA(K), and (b) shows the Hamming results for FDA_BH(m). Table II gives the corresponding statistical data of these methods. Ideally, these two distributions should be distinctly separated, since any overlap between them causes recognition errors. The larger the margin of separation, the more confidence we have in the system to handle larger variations of the true class faces beyond that seen in the training set.



(a)

(b)

Fig. 5. Histograms of the in-class and between-class Euclidean/Hamming distances for (a) FDA(K) and (b) FDA_BH(m).

TABLE II
Decidability for the FDA Methods

Technique	K/m	Genuine		Imposter		d'
		Mean	Variance	Mean	Variance	
FDA(K)	50	0.24751	0.01907	0.40998	0.020014	1.1622
	150	0.21129	0.016745	0.4	0.024432	1.3151
FDA_BH(m)	50	0.17913	0.015713	0.49935	0.008	2.97
	150	0.15965	0.016281	0.49025	0.0026895	3.39

Two opposing patterns can be observed. First, Fig. 5(a) shows that an increase in the number of the most significant eigenvectors retained (K) causes the two distributions to move leftward as a single coherent surface. It can be seen that the distributions have different means and little overlap. Second, an increase in the number of bits extracted (m) leads to a larger separation between the two distributions in Fig. 5(b). The left distribution shifts leftward while the mean of the right distribution is fixed at 0.5, which supports our conjecture. Note that if the two distributions are completely separated, face recognition can be successfully performed to the degree that one can confidently decide whether a captured sample belongs to the left or the right distribution.

Daugman [2] defines a metric for "decidability" (d') as the separation between the means of the two distributions, divided by the square-root of their average variance. According to our tests, the decidability for the proposed technique is $\mathrm{d}' = 3.39$, which is slightly lower than for Daugman iris-recognition algorithms. The left distribution shows the result when images of the same face are compared; about 15% of the bits differ, whereas the right distribution is the result when images of different faces are compared; the fraction of disagreeing bits is packed around 50%. The different-face distribution on the right corresponds to a binomial having mean $p = 0.490$, standard deviation $\sigma = 0.003$, and DOF $N = p(1-p)/\sigma^2 = 92.92$. This implies that it is possible to make recognition decisions with high confidence levels.

The left distributions in Fig. 5(b) look more Gaussian after RP. This agrees with Diaconis and Freedman's results [18], which show that various high-dimensional distributions look more Gaussian when randomly projected onto a low-dimensional subspace. This effect is of major importance, because raw high-dimensional data can be expected to form very eccentric (that is, far from spherical) clusters, presenting an algorithmic challenge [19].

The standard deviation of this distribution, 0.003, reveals the effective number of independent bits when codes are compared. Because of correlations within codes, the number of DOFs ($N = 92.92$) is considerably smaller than the number of bits extracted ($m = 150$). To some degree, this is different from our assumption that the correlation between the bits is removed using the Gram–Schmidt orthogonalization procedure. An explanation is that for this projection space, there are meaningless variables (outliners) with a high noise level. To improve the dimensional reduction, Navarrete and Ruiz del Solar [20] recommends applying a given criterion for neglecting the components with small projection variance. A good criterion would be to choose only m components, obtained by the normalized residual mean-square error.

VI. Conclusion

In this paper, we study a secure, robust approach to generating biometric hash values for use with any biometric template. We use dimensionality reduction to improve the capability to represent features, and error correction to improve robustness with respect to variations in biometric data. This is empirically demonstrated for face here, and in previous studies, other image-based biometric templates, including fingerprint [21] and palmprint [22]. To the best of our knowledge, such a study has not been undertaken before. Due to the reliance of Monrose *et al.*'s. technique on timing information, this approach is not readily applicable to generating cryptographic keys from 2-D surfaces. Davida *et al.*'s. techniques are inappropriate for our goals as the stored error-correcting parameters, if captured, can be computed for the user's biometric, which is insecure. Also, since the performance of both this scheme and the Soutar *et al.* scheme on actual biometric data was not explored, it is unknown whether these techniques are applicable in our setting. Future work for the proposed framework and modules will involve its application to other biometrics and detailed exploration of each module in the framework, specifically feature-extraction strategies.

References

[1] G. I. Davida, Y. Frankel, and B. J. Matt, "On enabling secure applications through offline biometric identification," in *Proc. IEEE Symp. Security and Privacy*, 1998, pp. 148–157.

[2] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.

[3] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. K. V. Kumar, "Biometric Encryption," in *ICSA Guide to Cryptography*, R. K. Nichols, Ed. New York: McGraw-Hill, 1999, pp. 649–675.

[4] F. Monrose, M. K. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamics," in *Proc. 6th ACM Conf. Comput. Commun. Security*, 1999, pp. 73–82.

[5] The Biology of Security, CDW Canada, Inc., Jul. 2002 [Online]. Available: http://www.cdw.ca/webcontent/editorial/technologies/073102_TheBiologyOfSecurity.asp.

[6] V. Matyàs and Z. Rìha, "Biometric authentication - Security and usability," in *Proc. 6th IFIP TC6/TC11 Conf. Commun. Multimedia Security*, 2002, pp. 227–239.

[7] A. Goh and D. Ngo, "Computation of cryptographic keys from face biometrics," in *Proc. 7th IFIP TC6/TC11 Conf. Commun. Multimedia Security*, 2003, pp. 1–13.

[8] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, pp. 612–613, 1979.

[9] J. Fridrich, "Robust bit extraction from images," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, 1999, vol. 2, pp. 536–540.

[10] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proc. 7th ACM SIGKDD Int. Conf. Knowledge Discovery, Data Mining*, 2001, pp. 245–250.

[11] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[12] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, 1996.

[13] J. Tang, R. Nakatsu, S. Kawato, and J. Ohya, "A wavelet-transform-based asker identification system for smart multipoint teleconference," *J. Visualiz. Soc. Jpn.*, vol. 20, no. 1, pp. 303–306, 2000.

[14] G. C. Feng, P. C. Yuen, and D. Q. Dai, "Human face recognition using PCA on wavelet subband," *J. Electron. Imaging*, vol. 9, no. 2, pp. 226–233, 2000.

[15] X. Luo and G. Mirchandani, "An integrated framework for image classification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2000, vol. 1, pp. 620–623.

[16] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput.*, vol. 15, no. 5, pp. 295–306, 1998.

[17] B. Duc, S. Fischer, and J. Bigün, "Face authentication with Gabor information on deformable graphs," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 504–516, Apr. 1999.

[18] P. Diaconis and D. Freedman, "Asymptotics of graphical projection pursuit," *Ann. Statist.*, vol. 12, pp. 793, 815, 1984.

[19] S. Dasgupta, "Experiments with random projection," in *Proc. 16th Conf. Uncertainty Artif. Intell.*, 2000, pp. 143–151.

[20] P. Navarrete and J. Ruiz-del-Solar, "Eigenspace-based recognition of faces: Comparisons and a new approach," in *Proc. 11th Int. Conf. Image Anal. Process.*, 2001, pp. 42–47.

[21] A. B. J. Teoh, D. C. L. Ngo, and A. Goh, "Biohashing: Two factor authentication featuring fingerprint data and tokenized random number," *Pattern Recog.*, accepted for publication.

[22] C. Tee, M. Goh, A. B. J. Teoh, and D. C. L. Ngo, "A novel approach for dual factor authentication," *Pattern Anal. Appl.*, accepted for publication.