CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2012; **24**:1066–1085
Published online 28 July 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cpe.1797

SPECIAL ISSUE PAPER

Mobile device access control: an improved correlation based face authentication scheme and its Java ME application

Kai Xi [1], Jiankun Hu [1,*,†] and Fengling Han [2]

[1] *School of Engineering and Information Technology, UNSW@ADFA, Canberra, Australia*
[2] *School of Computer Science and IT, RMIT University, Melbourne, Australia*

SUMMARY

This paper investigates face authentication based access control solutions for camera-equipped mobile devices. A new hierarchical correlation based face authentication (HCFA) scheme is proposed, which suits resource-constrained mobile devices such as mobile phones and personal digital assistants. The idea of HCFA is conducting a partial correlation output peak analysis (analyze the relationship between each cross-correlation output peak generated from selected sub-regions of a face), in conjunction with conventional direct cross-correlation methods. The experimental results on the public domain database demonstrate that the proposed scheme achieved better performance than that of the conventional direct correlation based schemes. Furthermore, HCFA was implemented on the Nokia S60 CLDC emulator using Java ME (previously J2ME) programming technology in order to test the applicability and implementability. The test results show that the proposed algorithm is implementable on mobile devices. It not only shortens processing time but also reduces resource demand significantly, compared with the direct correlation algorithms. Copyright © 2011 John Wiley & Sons, Ltd.

Received 29 November 2010; Revised 13 May 2011; Accepted 15 May 2011

KEY WORDS: face authentication; access control; mobile biometric; correlation pattern recognition; Java

1. INTRODUCTION

Nowadays, mobile devices have been widely used all around the world. With advances of embedded technologies and strong driving force from the consumer market, the technologies for mobile devices have moved towards an era of convergence. As a result, a mobile phone is no longer a simple electronic device used solely for voice communication. Modern mobile phone may possess much additional functionality, such as camera, radio, MP3 player, web browser, gaming terminal, GPS navigator and TV.

As the necessity and importance of mobile devices are increasing rapidly, there is a trend that mobile devices carry more sensitive information where such information can be contacts' details, schedules, minutes, trading secrets or even classified documents. It is consumer's desire that mobile devices not only support a growing number of applications but also secure to use. When mobile devices are lost or stolen, stored sensitive information may fall into the wrong hands. Such security problems have posed a grave threat to the current mobile environment. It urges stronger protection mechanisms against information leakage and illegitimate use of mobile devices.

Biometric techniques offer a natural and reliable option for recognizing individuals. Biometric-based security systems identify a person by his/her physiological characteristics (like fingerprint,

*Correspondence to: Jiankun Hu, School of Engineering and Information Technology, UNSW@ADFA, Canberra, Australia.
†E-mail: J.Hu@adfa.edu.au

face, iris, DNA etc.) and behavioral characteristics (like gait, speech, keystroke dynamics etc.). Biometrics ensures genuine user presence, thus enhancing the privacy protection [1]. Biometric traits offer three main benefits: (i) Universality - every person possesses the biometric features; (ii) Uniqueness - it is unique from person to person; and (iii) Performance stability - its properties remain stable during a person's lifetime. These characteristics make biometric-based authentication and identification system more secure and convenient than that of the knowledge-based (PIN, password) and token-based (Smartcard, key) systems, because PIN/password can be forgotten or guessed whereas token can be lost or stolen [2].

1.1. Related work in face recognition

As one of the most popular biometric techniques, face recognition has been employed in a wide range of security applications such as secure e-commerce, personal access control, national border control and forensic. Face-based technique attracts interest mainly because facial images can be acquired easily via low-cost camera, rather than specific and expensive scanner. Moreover, face recognition can be done passively without any explicit participation. Various type of medium can become potential source of faces; for instance, surveillance, video call and digital photos.

Extensive research has been done in face recognition, and a variety of face recognition approaches have been proposed. Face recognition techniques can be roughly divided into two categories: feature based and holistic based. Feature-based approaches process the images and extract distinctive facial feature points for example eyes, nose, mouth and so on, and then estimate the geometric relationships among those points. Almost all early face recognition algorithms are based on facial features. In [3], the authors extract 35 facial parameters such as ratios of distances, areas and angles. The recognition decision is determined by a simple Euclidean distance measurement. Another example is elastic bunch graph based matching method proposed in [4]. Holistic approaches use global representation to identify faces. Examples are principal component analysis (PCA) and linear discriminant analysis (LDA). PCA, commonly referred to as the use of eigenfaces, was first proposed by [5]. Latest variation and extension of PCA approaches can be found in [6–8]. LDA is a statistical approach, which maximizes the inter-class difference and intra-class variation. Latest LDA-based approaches can be found in [9, 10].

In [11], the authors proposed a correlation-based recognition approach. The correlation filter, core component of the approach, works in spatial frequency domain, unlike other methods like PCA and the FisherFaces, which work primarily in the spatial domain. Correlation filter's shift invariance nature as well as the capability of handling distortion make it a very promising face recognition approach and attract much further research [12].

Built-in camera has become a standard component of a mobile device, which offers the feasibility and convenience of integrating face authentication function without additional scanner. Therefore, the mobile face recognition is gaining great attention; however, it is still an open research area and very limited research can be found in existing literature. There are three major issues of applying face recognition on mobile devices: (i) Computational efficiency: design a new mobile-oriented recognition algorithm or simplify an existing method to fit the mobile devices, which only possess very limited computing power and memory. Long processing time is not acceptable; (ii) Verification accuracy: the algorithm should output authentication results with a low error rate (FAR/FRR); and (iii) Implementability, universality, and maintainability: the wide deployment will be hindered if the algorithm is relying heavily on underlying hardware for example DSP or some specially designed libraries or mathematical functions that may not be available in every mobile device.

Q. Tao *et al.* [13] proposed a face authentication system on mobile personal devices (Eten M600 Pocket PC) for the personal network. The method verifies the facial feature vectors in a statistically optimal way using the likelihood ratio. Both theoretical and practical works were presented. However, weak points are there. Firstly, the biometric authentication on mobile device has not been resolved because the face verification algorithm hosts on a remote server, meaning that every time, the user has to transmit the facial image to the server and verify it remotely. In such case, mobile phone only played as a biometric feature acquisition device. Note that transmission through a mobile phone will significantly increase the user's network usage charge and drain energy from mobile's

DOI: 10.1002/cpe

battery quickly. Besides, storing a user's facial feature on a third party's server may raise the severe privacy issue. Secondly, the experimental results demonstrated a high error rate. The evaluation on public database Yale B showed FRR = 15% with FAR = 1%, which cannot be considered as a good result.

S. Han *et al.* [14] proposed a face verification algorithm using near-infrared light. The algorithm is on the basis of integer PCA excluding floating-point operation. Evaluation is conducted on Pocket PC (HP iPAQ hx 2750). Note that actually Pocket PCs are special portable computers equipped with powerful processor (ARM core) and large memory (usually more than 64 MB). They are uncommon and high-priced when compared with mobile phone. In addition, the method requires near-infrared light sensors, which is far too rare to be found in most mobile phones. Another major problem is the authentication accuracy. Around 16% equal error rate (EER) is unacceptable.

In consumer market, two Japanese companies Oki Electric Industry Co., Ltd and Omron announced 'Face Sensing Engine (FSE)' in 2005 [15] and 'OKAO Vision Face Recognition Sensor' in 2006 [16], respectively. The FSE and OKAO add face recognition capabilities to camera-equipped mobile phones. A user registers his/her facial images as templates, using the camera. To gain access to the mobile phone, system will verify the query face with the pre-stored template. The FSE and OKAO need both middleware engine and specific sensor. Apparently, manufacturers tend to have their own standards and proprietary technologies [15, 16]. Most of the current commercial solutions are integrations of the hardware and their dedicated technologies. The latest commercial mobile device with face recognition functionality is Apple's IPhone. The face component is called 'Face Match', developed by Polar Bear Farm Ltd (http://itunes.apple.com/app/face-match-face-recognition/id310972844?mt=8). However, similar to Tao's problem [13], the user's facial images are verified in remote servers, so it can hardly be considered as a 'real' mobile face verification method.

1.2. Our contribution

In our research, we consider deploying a user-side face verification function in the application layer of a mobile device, without modifying existing Operating System or adding additional hardware (sensors or circuits). Besides, a better implementability, extendibility, flexibility and portability can be achieved for general mobile devices. Our application is developed on Java Platform, Micro Edition (Java ME). Java ME provides a robust, flexible environment for applications running on mobile and embedded devices for example mobile phones, personal digital assistants (PDAs), TV set-top boxes, smartcards, and so on [17]. It inherits Java Standard Edition's main benefit of being platform independent as well as Object-oriented. Java ME applications can also be emulated on a PC during the development stage and then easily ported to real mobile device, without expensive system-specific development kits and hardware. Figure 1 is an imaginary picture of mobile phone with face verification module. Figure 2 is a snapshot of Java ME emulator for a commercial mobile terminal.

Resource constraint is the major challenge in designing and implementing mobile-based fingerprint/face recognition schemes because mobile devices have limited memory and computational power. Mobile applications should be deployed to consume as little resource as possible. Direct migration of desktop version algorithms to mobile devices will lead to system crash, non-executable (lack mathematical function support), unbearable processing time.

In our previous work [18], the partial correlation output peak analysis (PCOPA) has been originally proposed for fingerprint verification. The PCOPA approach cannot be directly applied to face verification in mobile computing environment because of the differences between fingerprint and facial images. For instance, fingerprint is more stable than face where recognizing face requires handling more factors such as illumination and expression. In this paper, inspired by the work in [18], we proposed a new algorithm for face verification, namely hierarchical correlation face authentication (HCFA) scheme. HCFA consumes less memory resource than that of the conventional direct correlation algorithm with full-size images. The major difference and improvement between the method of [18] and HCFA is that the former is purely based on the PCOPA whereas the latter utilizes PCOPA together with conventional correlation method via a decision fusion. In

DOI: 10.1002/cpe

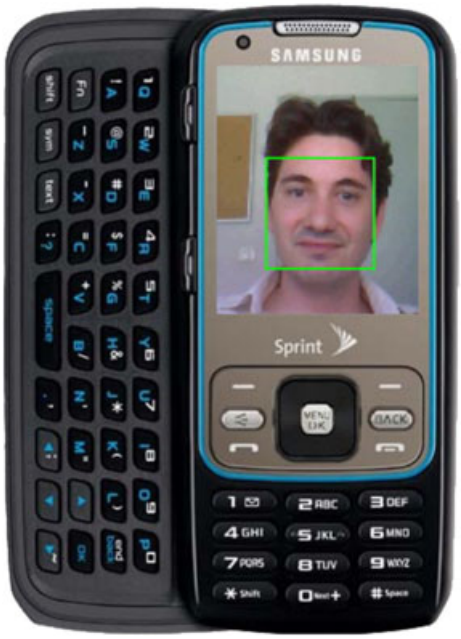

Figure 1. Mobile phone with face verification (Imaginary picture).

Figure 2. Nokia S60 Java ME emulator.

addition, the scheme proposed in [18] is designed for Java CDC (pocket PC) with high computing resource demand whereas HCFA is designed for Java CLDC (mobile phone), which requires only minimum computing resource. To investigate the authentication performance, the proposed algorithm was evaluated through a public domain face database, Yale Face Database [19] and Yale B Face Database (Extended) [20, 21]. Furthermore, the algorithm was implemented on Java ME emulator for the test of memory usage and computational efficiency.

DOI: 10.1002/cpe

The advantages of the proposed HCFA method can be concluded as:

- Computational efficiency
  HCFA is a computationally efficient and optimized algorithm. The hardware requirement meets most recent mobile phones including low-end mobiles on the market. The maximum memory consumption is only around 500 KB. The processing time is acceptable in the case of testing on entry level mobile phone platform.
- Verification accuracy
  Evaluation on large scale public domain database demonstrates a good result, which outperforms recent published mobile face authentication methods.
- Implementability, universality and maintainability
  It is the first time that face verification utilizes Java language without using any specific mathematical functions or extra libraries. The application can be conveniently downloaded over the air. On the contrary, most existing mobile face implementations are C/C++ based that are heavily dependent on the operating system and underlying hardware.

The rest of the paper is organized as follows: Section 2 introduces correlation-based pattern recognition techniques. Section 3 describes the proposed HCFA scheme. Section 4 presents experimental results. Section 5 discusses related issues of the Java ME implementation. Our conclusions and future work are in Section 6.

2. CORRELATION-BASED PATTERN RECOGNITION

Correlation-based pattern recognition was originally designed for automatic target recognition applications [22]. Afterward, it has also been applied in biometric authentication.

2.1. Conventional direct correlation approach

Correlation is a metric for measuring the similarity between a reference pattern $r(x)$ and a test pattern $t(x)$. In the case that there exists relative shifts between $r(x)$ and $t(x)$, it is applicable to compute the cross-correlation $c(x)$ between the two patterns for various possible shifts τ as in Equation (1). Then, the maximum value is selected as the similarity between the two patterns.

$$c(\tau_x) = \int t(x)r(x - \tau_x)dx \tag{1}$$

In biometrics, correlation-based face matching algorithm utilizes overall information provided in a facial image. Often, a synthetic filter is built as a template using a number of training samples [12]. During matching stage, a cross-correlation between a filter (template) and a query image will produce an array of numbers, named the correlation output. They refer to the inner products of different shifted versions of the template with the query image. When a query image perfectly matches with the filter, a well-defined peak will appear in the output correlation plane. Otherwise, a flat correlation output is expected to be observed.

Cross-correlation can be performed in the spatial domain. However, it is more computationally efficient to conduct cross-correlation and its relative calculation in frequency domain. Often, an image is represented as a 2D array (matrix) or a 1D vector of intensity values. The intensity matrix can be transferred from spatial domain to frequency domain using the Fourier transform (FT). In the enrolment stage, multiple training images of his/her face will be collected and utilized for designing a correlation filter in frequency domain. At the verification stage, when a query facial image comes in, it will be transformed into frequency domain using FT and then be multiplied by the correlation filter of the claimed person. The inverse FT of the products is taken to obtain the final correlation output.

The correlation-based face verification process is shown in Figure 3.

DOI: 10.1002/cpe

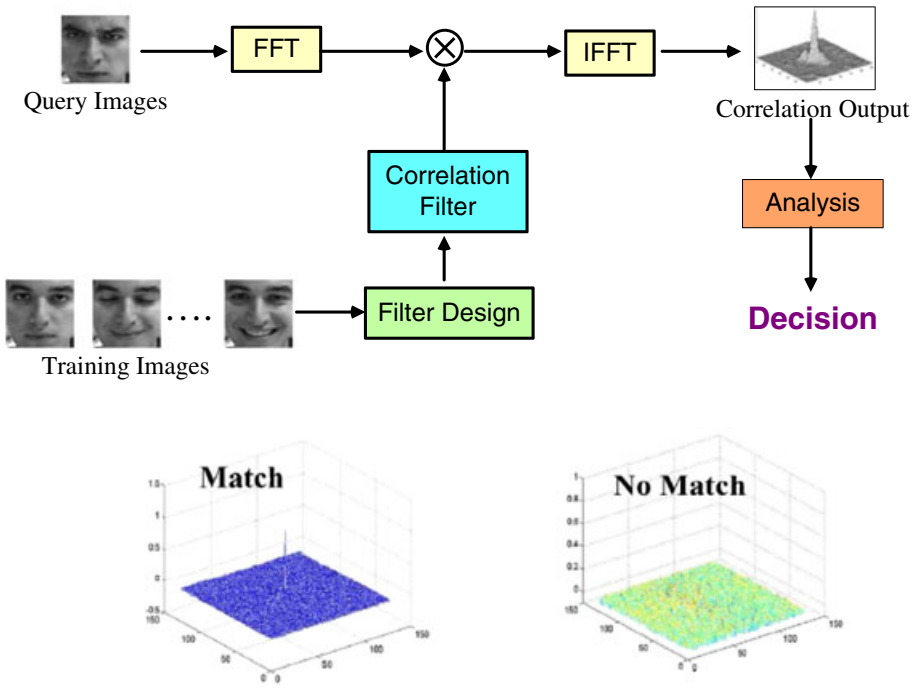

Figure 3. Direct correlation-based face recognition.

2.2. *Correlation filters*

The design of the correlation filter critically affects the verification performance. The simplest correlation filter is matched filters, which provide the maximum signal-to-noise-ratio in detecting a known reference image in the presence of additive white noise. However, matched filters are extremely sensitive to distortion in the reference image [23]. Therefore, each training image should have one corresponding matched filter. During matching stage, the query image needs to be cross-correlated with all matched filters, which is computationally inefficient.

To overcome the drawbacks of matched filter, the synthetic discriminant function filter, a linear combination of multiple training images, was proposed. During matching, query image only needs to be cross-correlated with one filter. However the performance of the synthetic discriminant function was not good because it produced correlation outputs where the sidelobes are larger than the controlled values at the origin [23].

The minimum average correlation energy (MACE) filter [24] is designed to minimize the average correlation plane energy while constraining the certain pre-specified value at the origin. It suppresses the sidelobes of correlation plane such that a sharp correlation peak can be produced. MACE filter is capable of providing a good discrimination without the need for imposter training images. For this reason, MACE filter is employed as an essential component of our scheme.

2.3. *Minimum average correlation energy filter*

Assuming there are N training images of a subject, each image has a total of d pixels. For the i-th training image, the columns of its 2D FT is concatenated to form a column vector x_i containing d elements.

A matrix X from N training images is then defined as

$$X = [x_1, x_2, \ldots x_n]^T \tag{2}$$

The 2D MACE filter obtained in the frequency domain is also ordered in a column vector h. The i-th correlation output at the origin is constrained to a pre-specified value u_i, which can be

DOI: 10.1002/cpe

represented as

$$c(0) = x_i^+ h = h x_i^+ = u_i \tag{3}$$

where the superscript '+' denotes a conjugate transpose. Note that $c(0)$ is also referred to the correlation output peak value.

Based on Parseval's theorem, the average of the correlation plane energy, E_{ave}, can be obtained directly from the frequency domain by

$$E_i = \sum_{p=1}^{d} |c_i(p)|^2 = \sum_{k=1}^{d} |h(k)|^2 |x_i(k)|^2 = h^+ x_i x_i^* h \tag{4}$$

$$E_{ave} = \frac{1}{N} \sum_{i=1}^{N} E_i = h^+ \left[\frac{1}{N} \sum_{i=1}^{N} x_i x_i^* \right] h = h^+ D h \tag{5}$$

where the superscript '$*$' denotes complex conjugation and D is a diagonal matrix of size $d \times d$ whose diagonal elements are the power spectrum of x_i.

Minimizing the average correlation energy E_{ave} subjecting to the constraints placed in Equation (3) leads to the MACE filter solution

$$h = D^{-1} X (X D^{-1} X)^{-1} u \tag{6}$$

where $u = [u_1, u_2, \ldots u_N]^T$.

3. PROPOSED HIERARCHICAL CORRELATION-BASED FACE AUTHENTICATION ALGORITHM

Conventional direct correlation approaches usually use full-size images. However, for mobile applications, large-size images will consume enormous system memory and computational resources. Hence, our HCFA scheme conducts cross-correlation on downsampled or small-size filter and images. Besides, in order to achieve high verification accuracy, a new method, called PCOPA, is proposed as the core component of the HCFA scheme. The preliminary prototype of PCOPA was first mentioned in [18] for fingerprint verification. In this paper, we improve and apply it into face verification. Section 3.1 provides mathematical proof of the PCOPA.

3.1. Proposed partial correlation output peak analysis

First, a continuous system is considered. Suppose $r(x)$ is a reference pattern and $f(x)$ is a test pattern, both are continuous signals. The cross-correlation between r and t is defined as:

$$c_f(t) = \int_{-\infty}^{\infty} f^*(\tau) r(t + \tau) d\tau \tag{7}$$

where f^* denotes the complex conjugate of f.

f is then decomposed into two sub-signals $f_1(x)$ and $f_2(x)$, where

$$f_1(x) = \begin{cases} f(x) & \text{for } x \leqslant a \\ 0 & \text{for } x > a \end{cases} \tag{8}$$

$$f_2(x) = \begin{cases} 0 & \text{for } x \leqslant a \\ f(x) & \text{for } x > a \end{cases} \tag{9}$$

Then cross-correlation of $f_1(x)$ and $f_2(x)$ are:

$$c_{f_1}(t) = \int_{-\infty}^{\infty} f_1^*(\tau) r(t + \tau) d\tau = \int_{-\infty}^{a} f^*(\tau) r(t + \tau) d\tau \tag{10}$$

DOI: 10.1002/cpe

$$c_{f_2}(t) = \int_{-\infty}^{\infty} f_2^*(\tau) r(t+\tau) d\tau = \int_{a}^{\infty} f^*(\tau) r(t+\tau) d\tau \tag{11}$$

The summation of $c_{f_1}(t)$ and $c_{f_2}(t)$ is:

$$c_{f_1}(t) + c_{f_2}(t) = \int_{-\infty}^{a} f^*(\tau) r(t+\tau) d\tau + \int_{a}^{\infty} f^*(\tau) r(t+\tau) d\tau = \int_{-\infty}^{\infty} f^*(\tau) r(t+\tau) d\tau = c_f(t) \tag{12}$$

The Equation (12) indicates that at a certain point t, the correlation output of $f(x)$ can be expressed as the summation of its sub-signals' correlation outputs.

In correlation-based biometric pattern recognition, a 2D image is converted to a 1D discrete function by concatenating each column vector to a single vector. In the spatial domain, correlation of a test pattern $f[x]$ with a reference pattern $r[x]$ leads to the following output:

$$c_f[t] = \sum_{p=1}^{N} f[p] \cdot r[p+t] \tag{13}$$

where Equation (13) indicates that the correlation output is a summation of inner products within the range $[1, N]$. Suppose $f[x]$ can be decomposed into two sub-signals $f_1[x]$ and $f_2[x]$, where

$$f_1[x] = \begin{cases} f[x] & \text{for } 1 \leqslant x \leqslant a \\ 0 & \text{for } a < x < N \end{cases} \tag{14}$$

$$f_2[x] = \begin{cases} 0 & \text{for } 1 \leqslant x \leqslant a \\ f[x] & \text{for } a < x < N \end{cases} \tag{15}$$

Then cross-correlation of $f_1[x]$ and $f_2[x]$ are:

$$c_{f_1}[t] = \sum_{p=1}^{N} f_1[p] \cdot r[p+t] = \sum_{p=1}^{a} f[p] \cdot r[p+t] \tag{16}$$

$$c_{f_2}[t] = \sum_{p=1}^{N} f_2[p] \cdot r[p+t] = \sum_{p=a+1}^{a} f[p] \cdot r[p+t] \tag{17}$$

The summation of $c_{f_1}(t)$ and $c_{f_2}(t)$ is:

$$c_{f_1}[t] + c_{f_2}[t] = \sum_{p=1}^{a} f[p] \cdot r[p+t] + \sum_{p=a+1}^{a} f[p] \cdot r[p+t] = \sum_{p=1}^{N} f[p] \cdot r[p+t] = c_f[t] \tag{18}$$

In Equations (16)–(17), $c_{f_1}(t)$ and $c_{f_2}(t)$ are the correlation outputs of two sub-signals $f_1[x]$ and $f_2[x]$, respectively.

If $f[x]$ is a linear transformed version of $r[x]$, their cross-correlation of them is an autocorrelation and $c_f[t]$ is equal to the summation of $c_{f_1}[t]$ and $c_{f_2}[t]$, that is Equation (18) holds. In this case, the correlation output peak value should be equal to the summation of peak values obtained from the corresponding fractions of the original segment. On the contrary, it is of very low probability that $f[x]$ and $r[x]$ are same patterns if $c_f[t] \neq c_{f_1}[t] + c_{f_2}[t]$.

3.2. Hierarchical correlation-based face authentication algorithm

Based on the mathematical characteristics of partial correlation, a new correlation-based hierarchical face verification scheme is proposed, consisting of two stages: enrolment stage and verification stage.

DOI: 10.1002/cpe

3.2.1. Enrolment stage. This stage is for a user to register his facial images for the purpose of template design. For each training image, the effective facial region will be detected and extracted using the modified Viola–Jones rapid face detector [25]. The face detection process is done offline in our experiment; however, Viola detector has been proved to be implementable on mobile platform [26]. Necessary scaling needs to be conducted on each facial region in order to ensure that all faces have the same size. Templates are constructed from a set of registered images using the MACE filter as described in section 2.3.

Conventional direct correlation verification schemes generate only one filter (template) from a set of large-sized training images that is the size of the filters are exactly the same as the original images. However, it does not suit resource-constrained mobile devices because images with full size will consume much memory, storage space and computational resource.

Unlike the conventional direct correlation verification schemes, HCFA utilizes either partial or downsized templates and query images. During enrolment, two types of templates will be generated. One is global filter, denoted by T_{global}, generated from several downsized training images, using a conventional correlation approach. See Figure 4.

The other templates are partial correlation filters that are generated from selected sub-region M_i of the original training images. M_i should be small and discriminating, which could be the area of eyes, nose, mouth and so on. These distinctive sub-regions can be extracted using feature detection algorithm (e.g. some recent face detection algorithm already possess eye, mouth, nose detection functionality) or from fixed location (in our experiment). An extractor has been developed to generate three fragments from M_i. The shapes of the three fragments are a small rectangle R_1, a hollow rectangle R_2 and a big rectangle R_3, as shown in Figure 5. The area of R_3 is the union of R_1 and R_2. For each face, local area is first selected and passed into the extractor, and then three corresponding parts M_{i_r1}, M_{i_r2} and M_{i_r3} will be extracted. Shapes of M_{i_r1}, M_{i_r2} and M_{i_r3} correspond

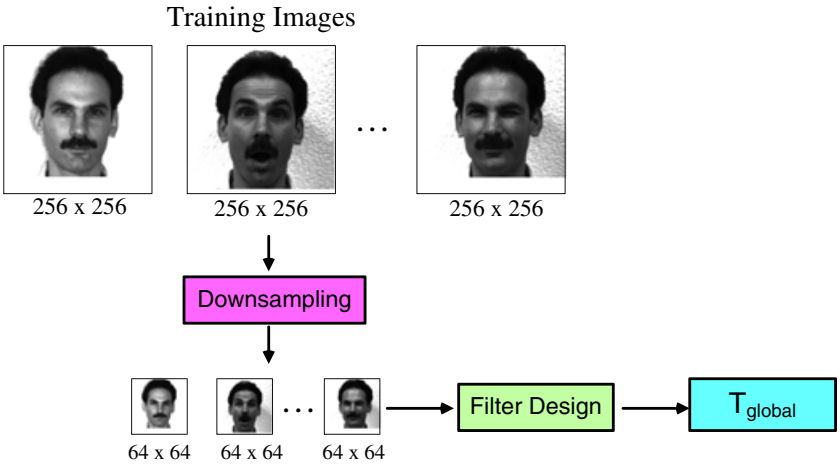

Figure 4. Global filter design.

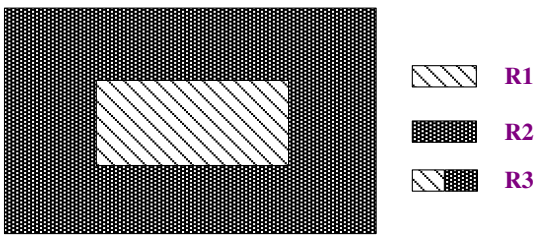

Figure 5. Fragment extractor.

DOI: 10.1002/cpe

to R_1, R_2 and R_3, respectively. For one local area T_i, three MACE filter will be generated at last, denoted as T_{i_r1}, T_{i_r2} and T_{i_r3}. T_{i_r1}, T_{i_r2} and T_{i_r3} are partial correlation filters that will serve for the PCOPA during verification stage. The process of generating partial correlation filter is depicted in Figure 6.

Both the global template T_{global} and partial template T_{i_rj} are represented as a discrete signal in frequency domain. The overall resource demand for T_{global} and partial correlation filters are much less than a full-size filter.

The number of genuine training samples (k) is suggested to be between 4 and 15. In our experiment, k of Yale database and Yale B (Extended) database are set to 4 and 10, respectively.

3.2.2. Verification stage. The verification stage consists of three tasks: (i) global matching; (ii) local matching; and (iii) decision fusion. Before global and local matching, effective facial region will be detected and extracted from the query image in advance.

Global matching

At global matching stage, the query image is firstly downsized to a small version Q_{global}, whose size should be the same as T_{global}. Then cross correlate Q_{global} with T_{global} to get the correlation output. A performance metric peak-to-correlation energy (PCE) is calculated to measure the sharpness of a correlation peak. PCE is defined as the ratio of the correlation peak and the energy in the correlation plane, as shown in

$$\mathrm{PCE} = \frac{|c(0,0)|^2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|c(x,y)|^2\,dxdy} \tag{19}$$

PCE is typically large for authentic and small for imposters. PCE obtained in this stage is recorded as PCE_{global}. Details of global matching process are shown in Figure 7.

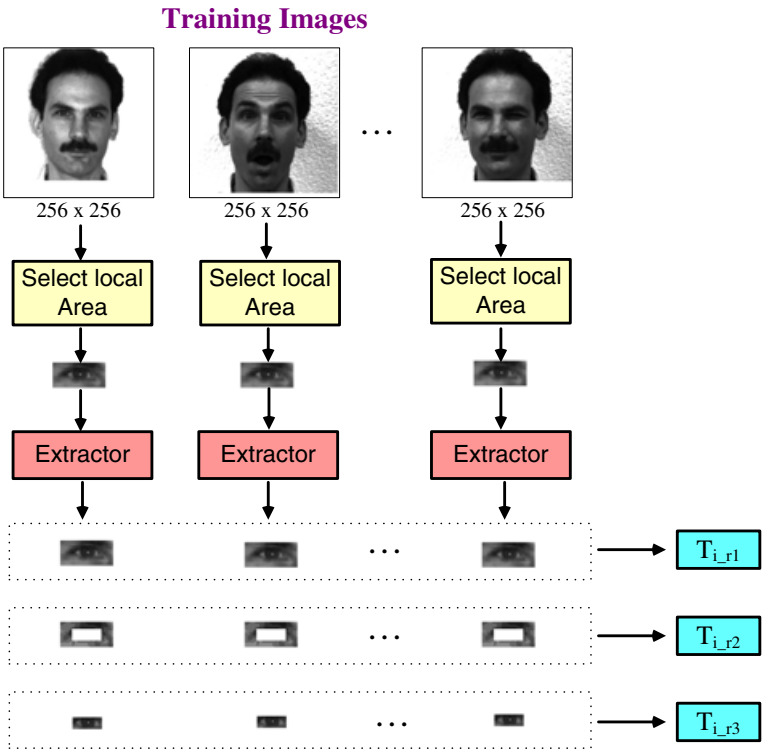

Figure 6. Generating partial correlation filters.

DOI: 10.1002/cpe

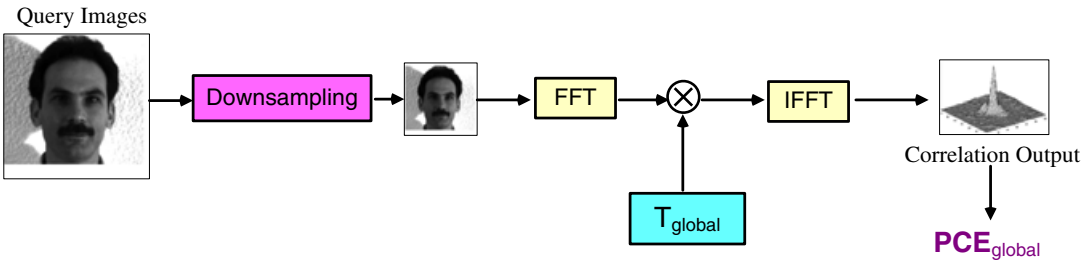

Figure 7. Global matching of HFVC scheme.

Local matching (PCOPA)

Firstly, a few partial sub-regions Q_i in the full-size query image are detected and selected. It corresponds to M_i in filter design stage. Then Q_i are passed into the extractor to get three corresponding parts Q_{i_r1}, Q_{i_r2} and Q_{i_r3}. The sizes and shapes of Q_{i_r1}, Q_{i_r2} and Q_{i_r3} are the same as M_{i_r1}, M_{i_r2} and M_{i_r3}. For a certain area Q_i, the Q_{i_r1}, Q_{i_r2} and Q_{i_r3} will be cross-correlated with its corresponding filters T_{i_r1}, T_{i_r2} and T_{i_r3}, respectively, yielding three correlation peak values p_{i_r1}, p_{i_r2} and p_{i_r3}. Let

$$d_i' = p_{i\_r3} - (p_{i\_r1} + p_{i\_r2}) \tag{20}$$

The difference can also be normalized to

$$d_i = d_1'/p_{i\_r3} = 1 - (p_{i\_r1} + p_{i\_r2})/p_{i\_r3} \tag{21}$$

Now we analyse the relations among partial correlation output peaks. Ideally, d_i will be zero when the query face is from the same source as the template and be large otherwise. However in practice, the value of d will never be perfectly equal to zero as a result of intra-class variations presenting in facial images such as variable illumination conditions or expressions. Usually, a genuine test d_i can only be close to zero and vice versa. We combine a set of d to a single metric D through calculating Harmonic mean of d_i, as follows:

$$D = N \cdot \left(\sum_{i=1}^{N} \frac{1}{d_i} \right)^{-1} \tag{22}$$

where N is the size of d. N is set to 3 in the experiment such that three partial areas, left eye, mouth and nose, are compared. Similar to d_i, D should be close to zero if query image originates from the same person. Details of PCOPA process is shown in Figure 8.

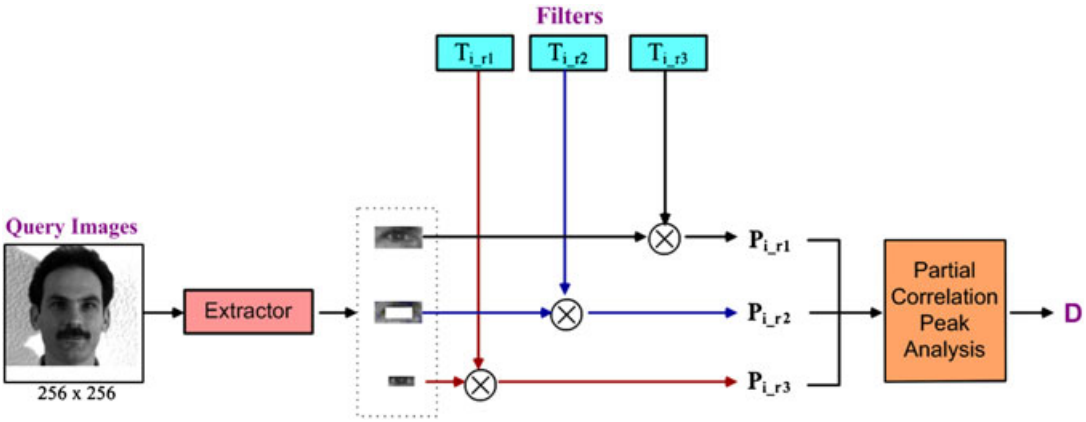

Figure 8. Partial correlation output peak analysis of HFVC scheme.

DOI: 10.1002/cpe

Decision making

Decision making is based on both PCE_{global} and D. It is necessary to combine D and PCE to find a single scalar score. When treating D and PCE as a vector $x = (D, PCE)$, the LDA method can be applied for score level decision fusion.

A genuine set G with n_1 samples and an imposter set F with n_2 samples are selected for training. Both G and F are matched with the template, yielding n_1 genuine matching scores and n_2 imposter matching scores. Based on our experience, the values of n_1 and n_2 are suggested to be set between 5 and 10. To seek a linear boundary, under the Gaussian assumption, the best separate of G and F could be computed as $W = S^{-1}(m_1 - m_2)$. The group covariance matrix S is defined as:

$$S = \frac{1}{(n_1 + n_2)} \left(\sum_{x \in G} (x - m_1)(x - m_1)^T + \sum_{x \in F} (x - m_2)(x - m_2)^T \right) \tag{23}$$

And m_1 and m_2 the mean vector of G and F, respectively.

For a test x_t, Equation (24) is employed to make a decision:

$$f = W^T x_t^T - ln\left(\frac{n_1}{n_2}\right) - \frac{1}{2}(m_1 + m_2)W \tag{24}$$

If $f > 0$, x_t is genuine. Otherwise x_t is imposter.

In practice, a user's template is stored in the device memory. Both correlation filter training and LDA decision threshold training process can utilize either live-captured snapshot (real-time mode) or pre-stored photos (offline mode).

4. PERFORMANCE EVALUATION

The HCFA scheme was implemented using Matlab and tested on desktop PC with some facial images. The main purpose is to statistically estimate the accuracy and effectiveness, regardless of memory restriction as well as computational power of mobile devices. Java ME implementation and its related issues will be discussed in Section 5.

We consider the worst case for HCFA scheme in which the selected test samples present various facial expressions or under different illuminations. Public domain databases.

Yale face database [19] and Yale B (Extended) database [20, 21] are used in our experiment. Yale database contains 165 grayscale images of 15 subjects (person), 11 images per subject. The size of database images is 320×243. The faces include center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised and wink. See Figure 9.

Figure 9. Yale face database.

DOI: 10.1002/cpe

Yale B (Extended) database is much larger, which consists of 38 subjects each with 64 impressions in conditions of different illuminations and expressions. The situation such as Figure 10.1 and Figure 10.2 were included in our experiment. Some extreme cases such as Figure 10.3 were not employed in our experiments. In real commercial product, the case in Figure 10.3 can be avoided. The light/exposure compensation is the basic functionality of almost all recent cameras, and it is usually implemented at the hardware level.

The reason of selecting such databases is because the evaluation of correlation-based method requires multiple images (usually over 10) with same pose. Some databases such as FERET do not have sufficient same-pose facial images for each subject.

At the beginning of our experiment, necessary image pre-processing is done. Then we conducted face detection algorithm to extract effective facial region from images. Each extracted facial area was cropped and scaled to a new size 256×256. All of them are taken negative and normalized so that the energy of each image was equal to 1.

In the experiment on Yale database, four sample images of each subject are selected to design the correlation filter. Facial samples from same subject will be regarded as genuine tests whereas the ones from all other subjects will be imposter tests. Therefore, a total $7 \times 15 = 105$ genuine tests and $11 \times 15 \times 14 = 2310$ imposter tests are performed.

For a fair comparison, we conducted five groups of experiments. In the first group, we use conventional direct correlation approach in which the correlation filters are generated from full-size 256×256 images. Test images are in full size as well. The 2nd, 3rd, 4th group of experiment still tested the conventional direct correlation approach however downsized facial images ($64 \times 64, 32 \times 32, 16 \times 16$) were used for both template design and verification test. These tests were exactly the global matching process of the HCFA scheme. It allows us to investigate the verification performance of using global matching with downsized samples. In the 5th group of experiment, we applied the HCFA scheme that consists of global and local matching. The sizes of templates and test images are 64×64. In terms of local sub-region selection, R_1 is set to 20×12 and R_2 and R_3 set to 42×21. Three local areas, left eye, nose and mouth, are extracted. The segmentation size is chosen on the basis of empirical research. Without degrading the overall verification performance, we expect the sizes of sub-regions are as small as possible, in order to reduce the resource demand.

The system's verification performance can be assessed by several different metrics. Three commonly used ones are false acceptance rate (FAR), false rejection rate (FRR) and genuine acceptance rate (GAR). The FAR is the error rate of imposters being falsely accepted by the system as genuine users, whereas the FRR is the error rate of genuine users being falsely rejected as imposters. GAR is the measurement of how many genuine users successfully pass the system test. They are defined as:

$$\text{FAR} = \frac{\text{Number of incorrectly accepted imposter tests}}{\text{Total number of imposter tests}} \tag{25}$$

$$\text{FRR} = \frac{\text{Number of incorrectly rejected genuine tests}}{\text{Total number of genuine tests}} \tag{26}$$

Figure. 10.1 Figure. 10.2 Figure. 10.3

Yale B (Extend) Facial Images in Different Illumination Conditions

Figure 10. Yale B (Extended) facial images in different illumination conditions.

DOI: 10.1002/cpe

$$\text{GAR} = \frac{\text{Number of accepted genuine tests}}{\text{Total number of genuine tests}} = 1 - \text{FRR} \tag{27}$$

Figure 11 plots the FAR and GAR produced by the five groups of experiments mentioned earlier. Apparently, with the same FAR, the proposed HCFA algorithm obtains a higher GAR than the other two approaches. On the other hand, direct correlation approaches with neither full-size images nor downsized images are able to achieve 100% GAR.

There is a strict tradeoff between FAR and FRR in all biometric systems [1]. Normally there will never be zero FAR accompanied with zero FRR as no biometric system is 100% accurate. When the FAR and FRR are equal, the common value is referred to as the EER. The low EER obtained, the better performance achieved.

Figure 12 shows EER for the three approaches. EER of HCFA scheme is only 1.2%, much lower than those of the others.

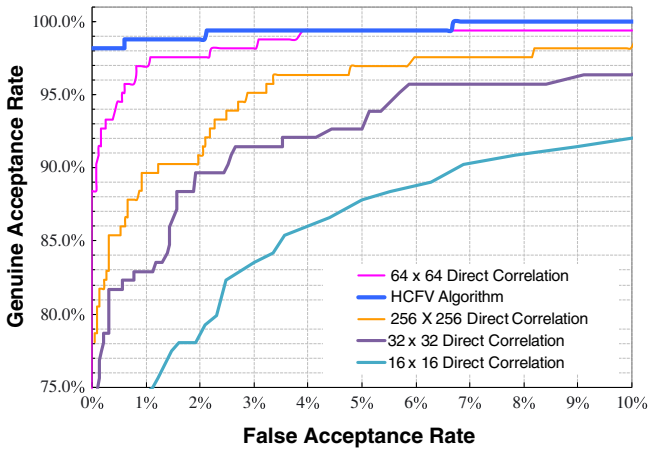

Figure 11. Performance: GAR/FAR curve (Yale database).

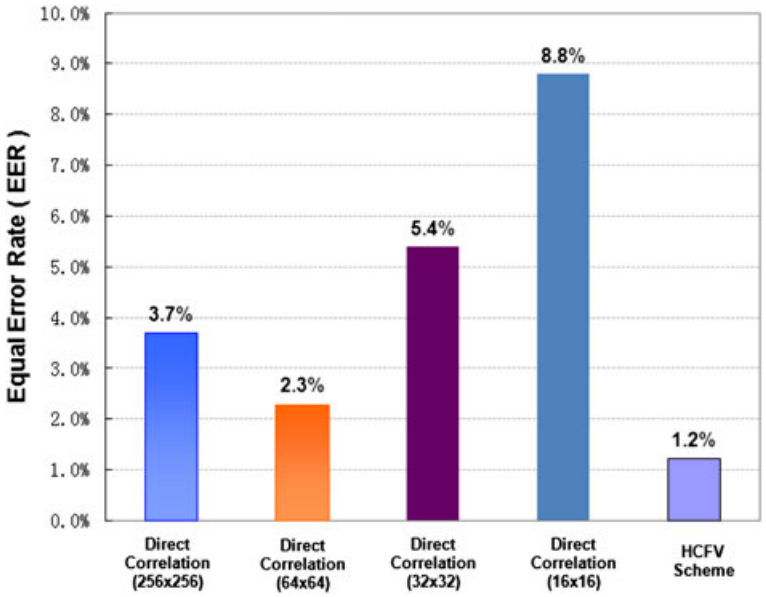

Figure 12. Equal error rate (EER) (Yale database).

DOI: 10.1002/cpe

For a verification application on a mobile device, a low FAR is more important because the primary objective is to stop imposter. However, a high FRR may irritate the users by requiring multiple logon attempts. Therefore, we consider the FRR at zero FAR as our performance measurement.

Figure 13 shows the FRR at zero FAR for HCFA scheme and conventional correlations with different image size. Compared with conventional direct correlation (64×64), HCFA scheme reduced the FRR dramatically, from 11.6% to 1.8%.

From the experimental results shown in the figures, it is clear that HCFA scheme outperforms the other four approaches. Conventional direct correlation method with small image size (64×64) is better than direct cross-correlation with large image (256×256) but worse than the HCFA scheme. HCFA can provide more accuracy than direct correlation (64×64) method, because the partial correlation peak analysis seeks a new way that differs the conventional correlation PCE analysis. When PCE is not able to distinguish between genuine users and imposters, correlation peak analysis method may still work, via observing the relationship between each correlation peak value.

Normally, it is expected that the larger images are used, the higher accuracy can be achieved. However, the performance rank is $64 > 256 > 32 > 16$. The result indicates 64×64 is the best choice in terms of authentication accuracy. The reason is that, unlike feature-based methods, correlation is a content-based method that takes all pixel information. A 64×64 image possesses sufficient information. 32×32 and 16×16 image losses some critical information, yielding a worse performance. However, increasing the image size from 64×64 to 256×256 may only add very limited useful information; on the contrary, more noise will be involved, which may dramatically degrade the performance. In other words, downsizing from 256×256 to 64×64 can be considered as an image smoothing process, which will remove some noises and enhance the performance. In [27], the direct correlation method was applied on fingerprint and tested using NIST 24 database. Similar conclusion was exhibited. The experiment result was $256 > 128 > 512 > 64 > 32$. 256×256 is the best image size for fingerprint, rather than 512×512.

For Yale B (Extended) database, 10 training images with different illumination and expression are used for template training, yielding around 1,500 genuine tests and 70,000 imposter tests.

Figure 14 is the performance curve for three methods: HCFA, full-sized direct correlation and downsampled direct correlation. HCFA shows higher GAR than those of the other two methods when FAR is less than 8% and exhibits close GAR performance when FAR is greater than 8%. The GAR performance of the downsampled correlation method outperforms that of the full-sized

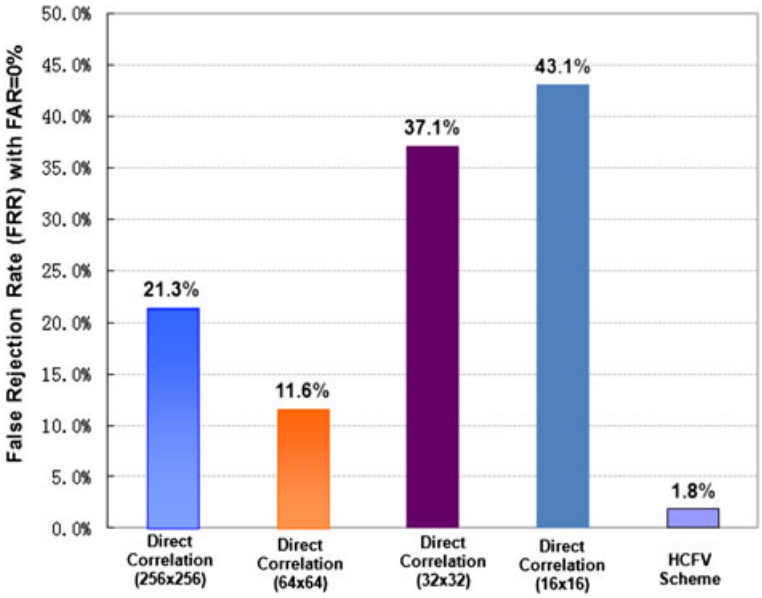

Figure 13. FRR with zero FAR (Yale database).

DOI: 10.1002/cpe

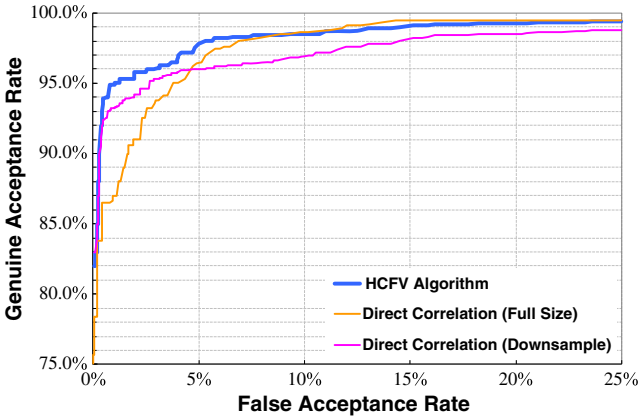

Figure 14. Performance: GAR/FAR curve (Yale B extended).

correlation method when FAR is less than 4.8%. However, it falls below that of the full-sized correlation when FAR is greater than 4.8%. FAR = 4.8% is the turning point for the two direct correlation methods.

Figures 15 and 16 illustrate the EER and FRR (FAR = 1%) of the three methods. HCFA enhanced the EER and FRR to 3.58 and 5.3%, respectively, more accurate than two direct correlation methods.

Also, we compare our algorithm with two most recent mobile face verification algorithms, Tao's method [13] and Han's method [14], as shown in Table I. The test data set of Tao's method and Han's method are Yale B database and self-collected data, respectively. It is clear that for both EER and FRR metrics, HCFA demonstrated higher verification accuracy.

In conclusion, the HCFA scheme that combines PCE (global matching) and partial correlation peak analysis (local matching), was proven to be reasonable and effective from not only the theoretical analysis but also the experimental evaluation.

5. JAVA PLATFORM, MICRO EDITION IMPLEMENTATION

Java ME is a Java platform designed for mobile devices and embedded systems such as mobile phones, PDAs and set-top boxes. Java Me is one of the most popular mobile application development technologies. A major advantage of Java ME development is its cross-platform nature that

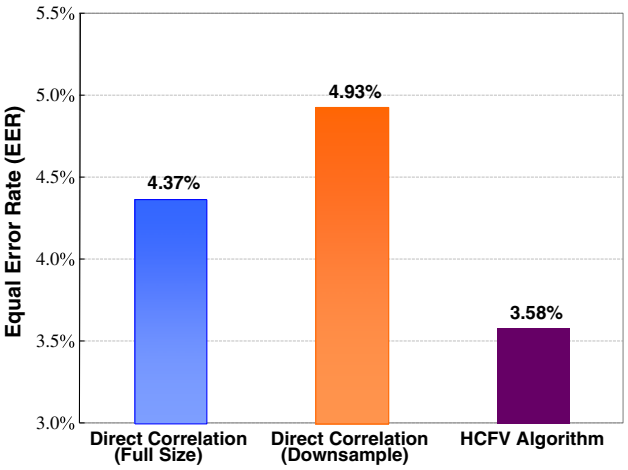

Figure 15. Equal error rate (EER) (Yale B extended).

DOI: 10.1002/cpe

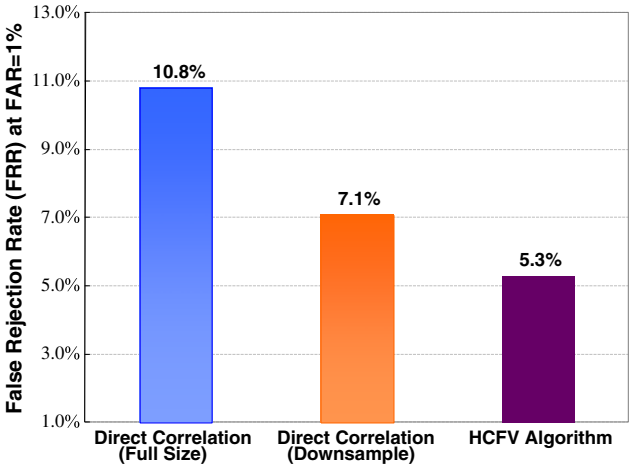

Figure 16. FRR with FAR = 1% (Yale B extended).

Table I. Performance comparison between HCFA and two most recent algorithms.

Method	EER(%)	FRR at 1% FAR
HCFA	3.58	5.3
Tao [14]	7.00	15.00
Han [20]	14.81	>14.81

means the same source code can be executed on all platforms and the whole Java ME architecture is independent of the underlying mobile Operating System.

A crucial component of Java ME is the Java Virtual Machine. It consists of Java compiler and Java interpreter where the former translates source code into an intermediate language called bytecode, and the later reads the bytecode and executes it. Basic functionalities of Java ME are summarized in configurations. The Java ME platform defines two configurations [17], one for general mobile devices, called the Connected Limited Device Configuration (CLDC), and the other targets at mobile devices with more capabilities like smart-phones and set-top boxes, called the Connected Device Profile. Our application is deployed based on CLDC as it is much more popular than Connected Device Profile and is widely supported by almost all Java-enabled mobile phones.

The Nokia Series 60 (S60) emulator is utilized as the development and test platform. As one of the leading mobile platforms in the world, S60 provides a perfect support to the latest Java ME CLDC technology. Besides CLDC, the Mobile Information Device Profile (MIDP 2.0) was also employed in our implementation. The device profile of the CLDC emulator is Nokia N73, which is the lowest-end mobile phone in recent consumer market. The purpose is to simulate and test the worst case. High-end smartphone platform, for instance Apple Iphone (with 1-GHz processor and 512-MB RAM), is not suitable for developing and testing. Because even if the algorithm can pass a test on high-end device, there is still no guarantee it can run well on low-end mobiles.

Complex number operations are needed when implementing Spatialfrequency domain signal processing. However, the CLDC 1.1 only supports floating point in real-number field. Hence, a class is developed on our own to handle complex numbers and the related mathematical operations. In addition, the Radix-2 Fast FT (Radix-2 FFT) algorithm was implemented from scratch. The FFT algorithms provide computational advantages of spatial frequency domain transformation where the complexity of N point DFT is whereas Radix-2 FFT is only. Computational cost of FFT is complex multiplications and complex additions. The computational improvement of FFT is roughly proportional to an efficiency ratio. As N increases, this FFT efficiency ratio increases significantly and so does the resource usage.

DOI: 10.1002/cpe

Original template size and query image size are pixels. Performing Radix-2 FFT/IFFT on template, the required complex operations for query image and final correlation output is

$$3 \times 3/2 \times 65536(log_2 65536) = 4718592 \tag{28}$$

In the HCFA scheme, the FFT is performed only on downsized or fractional part of template and query image. In terms of the global matching process, downsized filter and image are $64 \times 64 = 4096$ pixels and FFT/IFFT needs $3 \times 3/2 \times 4096(log_2 4096) = 221184$ operations. For partial correlation peak analysis process, the selected three partial areas contain $20 \times 12 = 240$ pixels, $42 \times 21 = 882$ pixels, $42 \times 21 = 882$ pixels, respectively, yielding

$$3 \times 3/2 \times 256(log_2 256) + 2 \times (3 \times 3/2 \times 1024(log_2 1024)) = 101376 \tag{29}$$

HCFA scheme conducts peak analysis three times. Hence, total complex operations of both global matching and local matching are:

$$221184 + 3 \times 101376 = 525312 \tag{30}$$

It is clear that HCFA requires much less complex operations than conventional direct correlation with full-size images.

On the other hand, we also compared the resource consumption of conventional direct correlation methods and HCFA scheme. Table I summarizes the memory usage of HCFA, storage usage and verification time of HCFA on the Java ME S60 CLDC emulator, compared with the conventional direct correlation approach with image. The device profile is Nokia N73 which is an entry level mobile phone recently. Performance, especially processing speed, will be improved if middle-high end mobile phones are employed.

In Table II, it is clear that HCFA consumes approximately 1/10 memory and 1/6 storage space as the Direct Correlation. The entire calculating process of HCFA is only 14.596 second, much shorter than Conventional Direct Correlation.

The computational efficiency comparison of HCFA with [13, 14] is shown in Table III. It is noted that in Tao's method [13] facial images are verified on desktop PC rather than mobile devices. Mobile devices only play a role as image acquisition component. Han's method [14] seems to be faster than HCFA. However, it is because of the high efficiency of the programming language (C/C++) and high computing power of the testbed rather than the algorithm itself. The fact is C/C++ is much faster than Java. Implementation using Java takes around twenty times longer time than using C/C++. Besides, unlike low-end mobile phones (e.g. N73), Han's testbed, HP Pocket PC, is equipped with high performance ARM processor and large memory, which significantly accelerate the computing process. In reality, HCFA and Han's method should be at the same level in terms of

Table II. Resource consumption and computational time of conventional correlation and HCFA.

	Conventional correlation (256×256)	HCFA
Max memory usage	6374.375 KB	587.997 KB
Storage size	1152.425 KB	183.842 KB
Computational time	419.331 s	14.596 s

Table III. Resource consumption and computational time of HCFA and two other algorithms.

Method	Processing time	Testbed	Testbed computing power	Testbed language	Language execution speed
HCFA	14 s	Nokia N73 Mobile Phone	Low	Java	Slow
Tao [14]	8 s (on desktop)	Eten M600 Pocket PC and desktop PC	High	C/C++	Fast
Han [20]	80 ms	HP hx2750 Pocket PC	High	C/C++	Fast

DOI: 10.1002/cpe

computational efficiency; nevertheless, it should be noted that HCFA outperforms Han's method in terms of verification accuracy.

Hence, HCFA scheme is proven to be implementable and computational efficient.

6. CONCLUSION AND FUTURE WORK

In this paper, we present a new correlation-based face authentication scheme for controlling the access of camera-equipped mobile devices. The proposed HCFA scheme consists of global matching on downsized images and local sub-region matching on the basis of PCOPA, making it differ from conventional direct correlation approaches. The fusion of global matching score and local matching score determines the final matching result. Performance test on the Yale face database that indicates a high accuracy has been achieved. The implementation on Java ME emulator further proves our scheme suits resource-constrained mobile computing environment as a result of its computational efficiency as well as low memory and storage demand. Future work includes adopting advanced fusion strategies for decision making. Moreover, implementation and test of the HCFA scheme on real mobile devices also worth some further research.

ACKNOWLEDGEMENTS

This research is supported by ARC (Australia Research Council) Projects LP110100602, LP100200538, LP100100404 and DP0985838.

REFERENCES

1. Wang Y, Hu J, *et al*. A fingerprint orientation model based on 2D Fourier Expansion (FOMFE) and its application to singular-point detection and fingerprint indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 2007; **29**(4):573–585.
2. Maltoni D, Maio D, Jain A, Prabhakar S. *Handbook of Fingerprint Recognition*. Springer-Verlag: New York, 2003.
3. Brunelli R, Poggio T. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 1993; **15**(10):1042–1052.
4. Wiskott L, Fellous J-M, *et al*. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 1997; **19**(7):775–779.
5. Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A: Optics, Image Science, and Vision* 1987; **4**(3):519–524.
6. Yang J, Zhang D, *et al*. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 2004; **26**(1):131–137.
7. Zhang D, Zhou Z, *et al*. Diagonal principal component analysis for face recognition. *Pattern Recognition* 2006; **39**(1):140–142.
8. Sun N, Wang H, *et al*. An efficient algorithm for kernel two-dimensional principal component analysis. *Neural Computing and Applications* 2008; **17**(1):59–64.
9. Jing X, Tang YY, *et al*. A Fourier LDA approach for image recognition. *Pattern Recognition* 2005; **38**(3):453–457.
10. Zhao H, Yuen PC. Incremental linear discriminant analysis for face recognition. *IEEE Transactions on System, Man & Cybernetics: Part B* 2008; **38**(1):210–221.
11. Savvides M, Vijayakumar BVK, *et al*. Face verification using correlation filters. *Proceedings Of the Third IEEE Automatic Identification Advanced Technologies*, 2002; 56–61.
12. Venkataramani K, Qidwai S, *et al*. Face authentication from cell phone camera images with illumination and temporal variations. *IEEE Trans. on Systems, Man and Cybernetics – Part C: Applications and Reviews* 2005; **35**:411–418.
13. Tao Q, Veldhuis R. Biometric authentication system on mobile personal devices. *IEEE Transactions on Instrumentation and Measurement* 2010; **59**(4):763–773.
14. Han S, Park HA, *et al*. Face recognition based on near-infrared light using mobile phone. *8th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA)*, 2007; 11–14.
15. Digital Trends. Cell phones to gain face recognition. (Available from: http://news.digitaltrends.com/news-article/8580/cell-phones-to-gain-face-recognition) [accessed on Oct 2005].
16. Ijiri Y, Sakuragi M, Lao S. Security management for mobile devices by face recognition. *Proceedings of the 7th International Conference on Mobile Data Management*, 2006.
17. Sun Website. (Available from: http://java.sun.com/).
18. Wang Y, Hu J, *et al*. Investigating correlation-based fingerprint authentication schemes for mobile devices using the J2ME technology. *IEEE workshop on Automatic Identification Advanced Technologies (AutoID'07)*, Alghero, Italy, 2007; 35–40.
19. Yale Face Database, 1997. (Available from: http://cvc.yale.edu/projects/yalefaces/yalefaces.html).

DOI: 10.1002/cpe

20. Georghiades AS, Belhumeur PN, *et al*. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2001; **23**(6):643–660.
21. Lee K-C, Ho J, *et al*. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2005; **27**(5):684–698.
22. Vijayakumar BVK. Tutorial survey of composite filter designs for optical correlators. *Applied Optics* 1992; **31**:4773–4801.
23. Vijayakumar BVK, Mahalanobis A, Juday RD. *Correlation Pattern Recognition*. Cambridge University Press: UK, 2005.
24. Mahalanobis A, Vijayakumar BVK, *et al*. Minimum average correlation energy filters. *Applied Optics* 1987; **26**:3633–3630.
25. Viola P, Jones M. Robust Real-time Object Detection, 2001. Cambridge Research Laboratories, CRL.
26. Ren J, Kehtarnavaz N, *et al*. Real-time optimization of Viola Jones face detection for mobile platforms. *Circuits and Systems Workshop: System-on-Chip Design, Applications, Integration, and Software* 2008.
27. Venkataramani K, Vijaya Kumar BVK. Performance of composite correlation filters in fingerprint verification. *Optical Engineering* 2004; **43**(8):1820–1827.
28. Engadgetmobile. Hands-on with ASUS touchscreen M536 QWERTY. (Available from: http://www.engadgetmobile.com/2008/03/03/hands-on-with-asus-touchscreen-m536-qwerty/).
29. Qidwai S, Venkataramani K, Vijayakumar BVK. Face Authentication from Cell Phone Camera images with Illumination and Temporal Variations. In *Biometric Authentication. LNCS 3072*. Springer: Heidelberg, 2004.
30. Softpedia. MWC 2008: The First Fingerprint Enabled World-phone From Fujitsu and AuthenTec. (Available from: http://news.softpedia.com/news/MWC-2008-The-First-Fingerprint-Enabled-World-phone-From-Fujitsu-and-AuthenTec-78683.shtml).
31. Vijayakumar BVK, Savvides M, *et al*. Correlation pattern recognition for face recognition. *Proceedings of the IEEE* 2006; **94**:1963–1976.

DOI: 10.1002/cpe