Biometric Authentication System on Mobile Personal Devices

Qian Tao and Raymond Veldhuis

Abstract—We propose a secure, robust, and low-cost biometric authentication system on the mobile personal device for the personal network. The system consists of the following five key modules: 1) face detection; 2) face registration; 3) illumination normalization; 4) face verification; and 5) information fusion. For the complicated face authentication task on the devices with limited resources, the emphasis is largely on the reliability and applicability of the system. Both theoretical and practical considerations are taken. The final system is able to achieve an equal error rate of 2% under challenging testing protocols. The low hardware and software cost makes the system well adaptable to a large range of security applications.

Index Terms—Biometric authentication, detection, face, fusion, illumination, registration, verification.

I. INTRODUCTION

IN a modern world, there are more and more occasions in which our identity must be reliably proved. But what is our identity? Most often, it is a password, a passport, or a social security number. The link between such measures and a person, however, can be weak as they are constantly under the risk of being lost, stolen, or forged. Biometrics, the unique biological or behavioral characteristics of a person, e.g., face, fingerprint, iris, speech, etc., is one of the most popular and promising alternatives to solve this problem. Biometrics is convenient as people naturally carry it and is reliable as it is virtually the only form of authentication that ensures the physical presence of the user.

In this paper, we study the biometric authentication problem on a personal mobile device (MPD) in the context of secure communication in a personal network (PN). A PN is a user centric ambient communication environment [19] for unlimited communication between the user and the personal electronic devices. An illustration of the PN is shown in Fig. 1. The biometric authentication system is envisaged as a secure link between the user and the user's PN, providing secure access of the user to the PN. Such an application puts forward the following three requirements of the biometric authentication system: 1) security; 2) convenience; and 3) complexity.

Manuscript received March 7, 2009; revised August 14, 2009. Current version published March 20, 2010. The Associate Editor coordinating the review process for this paper was Dr. David Zhang.

The authors are with the Signals and Systems Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands (e-mail: q.tao@lumc.nl; r.n.j.veldhuis@el.utwente.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2009.2037873

Security is the primary reason for introducing biometric authentication into the PN. There are two types of authentication in the MPD scenarios: 1) authentication at logon time and 2) authentication at run time. In addition to logon time authentication, run time authentication is also important because it can prevent unauthorized users from taking an MPD in operation and accessing confidential user information from the PN. To quantify the biometric authentication performance with respect to security, the false acceptance rate (FAR) is used, specifying the probability that an impostor can use the device. It is required that FAR be low for security concerns.

The false rejection rate (FRR), which specifies the probability that the authentic user is rejected, is closely related to user convenience. A false rejection will force the user to reenter biometric data, which causes considerable inconvenience. This leads to the requirement of a low FRR of the biometric authentication system. Furthermore, in terms of convenience, a higher degree of user-friendliness can be achieved if the biometric authentication is *transparent*, which means that the authentication can be done without explicit user actions. Transparency should also be considered as a prerequisite for the authentication at run time, as regularly requiring a user who may be concentrating on a task to present the biometric data is neither practical nor convenient.

Generally speaking, a mobile device has limited resources of computation. Because the MPD operates in the PN, it offers the possibility that biometric templates be stored in a central database and that the authentication is done in the network. Although the constraints on the algorithmic complexity become less stringent, the option brings a higher security risk. First, when biometric data have to be transmitted over the network it is vulnerable to eavesdropping [5]. Second, the biometric templates need to be stored in a database and are vulnerable to attacks [30]. Conceptually, it is also preferable to make the MPD authentication more independent of other parts of the PN. Therefore, it is required that the biometric authentication be done locally on the MPD. More specifically, the hardware (i.e., biometric sensor) should be inexpensive, and the software (i.e., algorithm) should have low computational complexity.

In this paper, we developed a secure, convenient, and low-cost biometric authentication system on the MPD for the PN. The biometric that we chose is the 2-D face image, taken by the low-end camera on the MPD. Two-dimensional face image is by far one of the best biometrics that compromises well among accuracy, transparency, and cost [37]. The user authentication, therefore, is done by analyzing the face images of the person who intends to logon into the PN or who is operating the MPD. The only requirement in using this system is that the user

Fig. 1. User in the PN.

Fig. 2. Diagram of the biometric authentication system on the MPD.

presents his or her face in a more or less frontal way within the capture range of the camera.

Although the subproblems of face recognition (like registration, illumination, classification, etc.) have been addressed in many publications [22], [52], [53], there have been very few publications describing a complete face recognition system, particularly for small mobile devices with limited computational resources. We propose a biometric authentication system consisting of the following five key important modules, as illustrated in Fig. 2: 1) face detection; 2) face registration; 3) illumination normalization; 4) face authentication; and 5) information fusion. Following the same order, this paper is organized as follows. Sections II and III present the real-time and robust face detection and registration algorithms, Section IV presents the illumination normalization method, Section V describes the face verification method for the processed face patterns, and Section VI describes the information fusion between different time frames. Section VII presents the experimental setup and result, and Section VIII gives the conclusion.

II. FACE DETECTION

Face detection is the initial step for face authentication. Although the detection of the face is an easy visual task for a human, it is a complicated problem for computer vision due to the fact that face is a dynamic object, subject to a high

degree of variability, originating from both external and internal changes. An extensive literature exists on automatic face detection [24], [52].

The known face detection methods can be categorized into the following two large groups: 1) heuristic-based methods and 2) classification-based methods. Examples of the first category include skin color methods and facial geometry methods [21], [25], [33], [38]. The heuristic methods are often simple to implement but are not reliable as the heuristics are often vulnerable to exterior changes. In comparison, classification-based methods are able to deal with much more complex scenarios, because they treat face detection as a pattern classification problem and, thus, benefit largely from the existing pattern classification resources. The biggest disadvantage of the classification-based methods, however, is their high computational load as the patterns to be classified must cover the exhaustive set of image patches at any location and scale of the input image, as shown in Fig. 3.

The Viola-Jones face detector is one of the most successful face detection methods [50]. There are three characteristics in this method, which are listed as follows: 1) Haar-like features that can be rapidly calculated across all scales; 2) Adaboost training to select and weight the features [16], [39], [50]; and 3) cascaded classifier structure to speed up the detection. In comparison to other advanced features like Gabor wavelets in the elastic bunch graph matching (EBGM) [51] and Gabor wavelet network [18], the Haar-like features are extremely fast to compute, and most importantly, they form key points for scalable computation and, thus, rapid search through scales and locations. For details, see [50]. The face detector only needs to be trained once and then stored for any general-purpose face detection. These characteristics enable real-time robust face detection.

For the application of face detection on the MPD in particular, we propose strategies that can further improve the detection speed. The specificity of the face images in the MPD application is related to the distribution of face sizes in the normal self-taken photos from a handheld device. This information provides

Fig. 3. Exhaustive set of candidates at any location and scale of the input image, where x is the basic classification unit to be classified.

Fig. 4. (Left) Typical face images taken from ordinary handheld PDA (Eten M600), with the size 320×240 . (Right) Downscaled face images with the size 100×75 . Equally good face detection results can be obtained in both the original and the downscaled images.

useful constraints on the searching and significantly speeds up the implementation. In Fig. 4 (left), some typical face images taken from ordinary handheld personal digital assistant (PDA; Eten M600) are shown.

Suppose the detected face size s lies in a scope between $s_{\min}$ and $s_{\max}$, we propose two steps to reduce the computational efforts for face detection: First, downscale the original image before detection. The downscaling factor is set around $s_{\min}/s_{\text{face}}$, where s_{template} is the size of the training template, i.e., minimally detectable size. In the trained detectors, $s_{\text{template}} = 24$. Second, in the reduced image, restrict the scanning window to be from the minimal size 24 to the maximal size $24(s_{\max}/s_{\min})$.

Referring to Fig. 3, it can easily be seen that the number of candidates for classification increases exponentially with the size of the input image. The first step, therefore, radically reduces the number of possible classification units. In addition, the second step avoids the unnecessary search for faces of too small or too large sizes. This further reduces the number of classification units to a large extent. Fig. 4 shows detection results both in the original and in the reduced image. We observed that in the latter, almost equally good results are obtained, but with far less computational load. One drawback of downscaling, however, is that the detected face scale is coarser than that in the original image, since much fewer scales have been searched through. This nevertheless does not affect the final face verification performance, as is shown in the following section,

where we will further registrate the detected faces to a finer scale.

III. FACE REGISTRATION

Generally speaking, the location of the detected face is not precise enough for further analysis of the face content. It has been emphasized in the literature that face registration is an essential step after face detection for subsequent face interpretation task [3], [4], [36]. The following two popular ways of doing the face registration can be found in the literature: 1) holistic methods and 2) local methods.

Holistic registration methods take advantage of both global face texture information and the local facial feature information (i.e., locations of eyes, nose, mouth, etc.). Examples are the active shape model (ASM) [10], active appearance model (AAM) [9], and their variants. Fitting such models onto the input face image is often formulated as an iterative optimization problem. As common to such complex optimization problems, however, two potential drawbacks of the holistic registration are the possibility to be trapped into local minimal and the relatively high computation load, particularly for an MPD. In contrast, the local registration method is more direct and faster as it only takes the locations of local facial features to calculate the transformation, not involving any global optimization process. The disadvantage of the local method, on the other hand, is that the facial features are very difficult to be reliably

Fig. 5. Comparison of the false acceptance models. (a) Original detections. (b) Conventional probabilistic model. (c) Proposed model with type I and type II errors.

detected [6] due to their high variability and insufficient shape content.¹

We propose a real-time robust facial feature detector based on the Viola–Jones method, incorporating a novel error model that is precise and concise. Given that it is, in reality, impossible to build a reliable facial feature detector with both low FAR and low FRR, we guarantee the detection of the facial features in the first place, at the cost of a high number of false detections. The facial feature detection problem is therefore converted into a postselection problem of the multiple detected facial features.

A conventional way to do the postselection is to build up statistical models, like the ASM, of the facial landmark distributions and to select the most likely combination of detected facial features according to them [6], [12], [13]. The underlying assumption of these models is that the detected features are distributed in a probabilistic way around the true locations. This assumption, however, is not true for the multiple detections that resulted from the Viola–Jones method. Furthermore, for a user-specific system, such a model is preferably user specific, but in practice, this is not possible (no ground-truth data of the user available), and variances of other subjects will be introduced by building a general shape model from a number of other training subjects. Fig. 5(a) shows a typical example of the left-eye detection. The face sample is from the FERET database [47], while the left-eye detectors are trained from the manually labeled BioID database [46].

We build up a new error model for the false detected facial features. Fundamentally, the Viola–Jones detector uses a combination of local structures as the template, so all the patterns that have more or less similar structures are likely to be detected. With respect to this mechanism, two types of false acceptances can be identified. The *type I false acceptance* is the acceptance of the background patches that coincidentally have comparable local structures with the facial feature. The chin shadow that is falsely detected as an eye, as shown in Fig. 5(a), is a good example of the type I error, as the shadow has roughly the bright–dark–bright pattern that resembles the eye texture. The *type II false acceptance* is the acceptance of

¹Shape content refers to the relatively consistent layout of the object pattern. For example, the face has consistent and sufficient texture patterns of the facial features (from up down: eyebrows, eyes, nose, mouth) that provide abundant information to discriminate them from nonface patterns and facilitate the construction of the detector. For facial features, e.g., eyes and eyebrows, however, the shape content is far less and more unstable, leading to considerable overlaps in the distributions of the facial-feature and non-facial-feature classes and making it difficult to learn the detector.

Fig. 6. Examples of facial feature detection from (a) BioID, (b) FERET, and (c) YaleB databases and (d) random unconstrained Internet images, with different illumination, size, pose, and expression.

the patches centered at approximately the same position as the true facial features, but larger in sizes. A type II error is caused by the fact that the search is through different scales, and at a slightly coarser (larger) scale around the true position, the detected image patches often have similar structures as the facial feature patch. Both error types can be observed in Fig. 5(a), and Fig. 5(c) illustrates the proposed model in detail, in comparison to the conventional probabilistic model depicted in Fig. 5(b). Obviously, the proposed model better describes the distribution of false detections.

To remove most of the type I false detections, we predefine a corresponding region of interest (ROI) before detecting certain facial features. The ROI acts as a geometrical constraint as in AAM [9] and EBGM [51], thus precluding a large percentage of false detections. For those false detections that remain within the ROI, we observed that the scale information of the detection, which was normally neglected, actually provides a very interesting insight. A concise principle to remove the false acceptances is proposed: a minimal-scale detection within the maximal-scale detection is mostly likely to be the true facial landmark location. The reasoning of this principle is directly related to the mechanism of the Viola–Jones method: First, the detections within the maximal-scale detection have less chance of being the type I false acceptances (i.e., random errors), as they have been confirmed multiple times by the overlapped detections. Second, the minimal-scale detection within the maximal-scale detection is most likely to be the accurate one, excluding the type II false acceptances, as it is detected on the finest scale. The type II false acceptances, therefore, are employed as extra information to confirm the localization, but are finally eliminated. In contrast, an additional statistical model can potentially eliminate the type I false detections but, in principle, cannot deal with the type II false detections, as they are virtually very close.

Fig. 6 shows some examples from databases and unconstrained Internet and real-time images. To make the registration sufficiently reliable in the automatic system, we have trained

13 facial landmarks from the BioID databases, as shown in the first image of Fig. 6. For registration, in theory, two landmarks are enough to compute the rigid transformation, but more landmarks are more robust in a minimum square error sense. Occasional misses of the detection of certain landmarks occur, but in most cases, the number of detected landmarks is large enough for a reliable registration.

The robustness and speed of the facial feature detectors are inherited from the Viola–Jones method, and the postselection strategy that we proposed further strengthened their applicability on an MPD. In summary, the proposed facial feature detectors are extremely fast and self-standing, not requiring any additional face shape or texture models, nor any interactive optimization. The trouble of learning statistical constraint models, together with the modeling error, is avoided.

IV. ILLUMINATION NORMALIZATION

The variability on the face images brought by illumination changes is one of the biggest obstacles for face verification. It has been suggested that the variability caused by illumination changes easily exceeds the variability caused by identity changes [34]. Illumination normalization, therefore, is a very important preprocessing step before verification. There has been an intensive study on this topic in the literature, which can be categorized into two different methodologies. The first category tries to study the illumination problem in a fundamental way by building up the physics imaging model and restoring the 3-D face surface model, either in an explicit or an implicit way. We call this category the 3-D methods, which include the linear subspace method [40], illumination cone [2], spherical harmonics [1], quotient image [41], etc. The second category, however, does not rely on recovering the full 3-D information, instead, they work directly on the 2-D image pixel values. We call this category the 2-D methods, which include histogram equalization [27], linear and homographic filters [22], the Retinex method [29], the diffusion method [7], etc.

The 3-D methods aim at utilizing the 3-D information that is robust to illumination. However, as converting the 3-D objects to the 2-D images is a process with loss of information, the reverse process will unavoidably introduce regulations or restrictions to make up for such loss, like fixed surface normals [41], absence of shadow or specular reflections [40], etc. In reality, such assumptions are very often violated and introduce artifacts in the processed images [42]. Furthermore, due to the complexity of the algorithm, 3-D methods normally load a heavy computational burden on the face authentication system. The 2-D methods, in contrast, are more direct and much simpler. Because of this directness and simplicity, on the other hand, it is not possible for 2-D methods to achieve illumination invariance, as has been theoretically proved in [8]. Therefore, instead of pursuing illumination invariance, we aim for illumination insensitivity and the compromise between computational cost and system performance. We proposed a 2-D illumination *insensitive* filter for the face authentication system on the MPD, called the simplified local binary pattern (LBP) filter.

LBP was initially proposed to solve the texture recognition problem [35]. The fundamental idea is stated as follows: a 3×3 neighborhood block in the image is thresholded by the value of the center pixel and results in eight binary values. This eight-bit binary sequence is then converted into a decimal value ranging from 0 to 255, representing the type of the texture pattern in the neighborhood of this central point. The distribution of the LBP patterns throughout the image is then used as the feature of the image. LBP histogram is recognized as a robust measure of the local textures, insensitive to illumination and camera parameters.

The LBP histogram is a good representation for images with more or less uniform textures, but for face images, it is insufficient. A distribution disassociates the connection between the patterns and their relative positions on a face and potentially decreases differences among subjects and mingles them up in space. To include the positional information, LBP can instead be used as a preprocessing method on the image values [23]. Essentially, this acts as a nonlinear high-pass filter. As a result, it emphasizes the edges that contain significant changes of pixel values, but at the same time, it also emphasizes the noise that involves only small changes of pixel values. The original weights of LBP, i.e., exponentials of 2, differ greatly within the neighborhood. For two neighboring pixels in the neighborhood, they are at least two times different, and in worst cases, their difference can be as large as 128 times. As noise occurs in a random manner with respect to the eight directions, the way of converting binary value to decimal value, i.e., the exponential weights assigned on the neighbors, renders the LBP filtering noise sensitive. To make the filtering more robust, we propose to simplify the weighting process, assigning equal weights on each of the eight neighbors. The noise is suppressed by not emphasizing circular differences and potentially averaging them out within the neighborhood. We show some examples from the YaleB database [20] in Fig. 7, in which the proposed simplified LBP filtering and the original LBP filtering are compared. The histogram equalization method is also illustrated as a reference. The performance of the three illumination normalization methods will be compared in Section VII.

It is observed that the simplified LBP filtering produces stable patterns under diverse illuminations, even under extreme illuminations. There are several advantages brought by the simplified LBP filtering. First, the LBP is a local measure, so the LBPs in a small region are not affected by the illumination conditions in other regions. Second, the measure is relative, and therefore, is strictly invariant to *any* monotonic transformation, such as shifting, scaling, or logarithm, of the pixel values. Third, we largely reduce the sensitivity of the LBP value to noise by assigning uniform weights in all eight directions. Finally, even for the MPD with limited computation resources, the proposed filtering operation is extremely fast.

The immediate concern is that simplified LBP filtering may filter out too much information from the face image, as both illumination-sensitive components and some face-related components are discarded. This problem can be solved in a systematic manner by introducing a classification method that has high discrimination capability such that the part of information loss caused by simplified LBP filtering is negligible. In Section V,

Fig. 7. Comparison of the illumination normalization methods. (a) Original face images. (b) Histogram-equalized images. (c) Images filtered by the original LBP. (d) Images filtered by the simplified LBP.

Fig. 8. Distributions of the two classes for the face detection case and the face verification case, respectively. Decision boundaries are also illustrated.

with special regards to this problem, we will discuss the generalization and discrimination capabilities of the classifier in a high-dimensional space and justify the simplified LBP filter for the illumination normalization purpose.

It is worth mentioning that the illumination problem can also be tackled from a hardware point of view. The near-infrared illumination, for example, provides more consistent images in diverse lighting situations and also enables face authentication at night [28].

V. FACE VERIFICATION

In this section, we address the problem of verifying the detected, registered, and normalized image pattern. Under the context of our application, this is a two-class classification problem, with the two classes defined as the *user* and the *nonuser* (or impostor) classes.

Although similar to the face detection problem in the sense that both are two-class classification problems, the face verification problem is different in the distribution of the two classes. An illustration is given in Fig. 8. In the verification case, the user class and the impostor class are more closely distributed in space than the two classes in the detection case. In other words, the chance that an impostor face resembles a user face is much higher than the chance that a random background patch resembles a face patch. In the detection case, a number support vectors, as shown in Fig. 8, are sufficient to “support” the decision boundary. In the recognition case, however, the distribution of the two classes are intermingled in a more complex way. This implies that the boundary-based classification methods that work well on the face detection problem, like the support vector machine method [11], which relies explicitly on the support vectors, or the Viola–Jones Adaboost method,

which relies implicitly on the highly weighted samples, are no longer suitable for the verification problem. The better solution, instead, is to classify in the overlapped regions with a minimal possible error, from a statistical point of view. For this reason, we propose to verify the feature vectors in a statistically optimal way using the likelihood ratio.

The likelihood ratio is an optimal statistic in the Newman–Pearson sense [49]: At a given FAR, the likelihood ratio achieves a minimal FRR; or at a given FRR, the decision-fused classifier reaches a minimal FAR. The likelihood ratio classification rule is defined as

$$L(x) = \frac{p(x|\omega)}{p(x|\bar{\omega})} > T \quad (1)$$

where x is the preprocessed face image, which is stacked into a vector, ω is the user class, $\bar{\omega}$ is the nonuser class, T is the threshold. When $L(x) > T$, x is accepted as the genuine user; otherwise, it is rejected. Since we assume infinitely many subjects in the sets $\omega \cup \bar{\omega}$, exclusion of a single subject ω from it virtually does not change the distribution of x . Therefore, the following holds:

$$p(x|\bar{\omega}) \simeq p(x) \quad (2)$$

which facilitates an even simpler modeling of the two classes conceptually as two overlapping clouds in a high-dimensional space, as shown in Fig. 8. The two classes are now the *user-face* class and the *all-face* (or background) class.

To obtain the likelihood ratio of an input feature vector x with respect to two classes ω and $\bar{\omega}$, the probability density functions of the two classes $p(x|\omega)$ and $p(x)$ are first estimated. The Gaussian assumption is often applied on a large set of data samples, which is motivated by the central limit theorem [16]. Given N sample feature vectors of the face x_i , $i = 1, \dots, N$, both μ and Σ can be estimated as follows:

$$\mu = \frac{1}{N-1} \sum_{i=1}^N x_i \quad \Sigma = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T. \quad (3)$$

To avoid the influence of extreme samples, which are possibly caused by extraordinary illumination, pose, expression, or misregistration, μ can also take the median of the sample vectors at every element: $\mu = \text{median}(x_1, \dots, x_N)$.

The two classes involved in face verification are the user class ω_{user} and the background class ω_{bg} . Equivalently, the likelihood ratio in (1) can be rewritten as

$$\begin{aligned} \ln L(x) &= \ln p(x|\omega_{\text{user}}) - \ln p(x|\omega_{\text{bg}}) \\ &= \frac{1}{2} \left(\ln |\Sigma_{\text{bg}}| + (x - \mu_{\text{bg}})^T \Sigma_{\text{bg}}^{-1} (x - \mu_{\text{bg}}) \right) \\ &\quad - \frac{1}{2} \left(\ln |\Sigma_{\text{user}}| + (x - \mu_{\text{user}})^T \Sigma_{\text{user}}^{-1} (x - \mu_{\text{user}}) \right) \\ &= \frac{1}{2} (d_{\text{Maha}}(x|\omega_{\text{bg}}) - d_{\text{Maha}}(x|\omega_{\text{user}})) + c \end{aligned} \quad (4)$$

where μ_{user} , μ_{bg} , Σ_{user} , Σ_{bg} are the means and covariances of the user and background classes, respectively. The term $c = 1/2(\ln |\Sigma_{\text{bg}}| - \ln |\Sigma_{\text{user}}|)$ is a constant that can be absorbed into the threshold T in (1) without influencing the final receiver operating characteristic (ROC). As (4) shows, the logarithm essentially reduces the likelihood ratio to the difference between the two squared Mahalanobis distances in the user and background classes.

It is useful to further study the discrimination capability and generalization capability of the likelihood ratio classifier. Generalization capability and discrimination capability are two equally important aspects in verification. For our MPD application, they are closely related to the convenience requirement and the security requirement, respectively.

We notice that the face verification is normally in a very high dimensional space. A small-enough face image, for example, of size 32×32 , already has 1024 pixels, which implies a 1024-dimensional feature vector. High-dimensional space potentially has great power of discrimination but is relatively difficult to generalize [44]. We explain this with a simple example using the model in Fig. 8(b). Suppose each of the user and the background class take up a hypersphere with radii r_{user} and $r_{\text{bg}} = a \cdot r_{\text{user}}$, $a > 1$, in an N -dimensional space. For a single dimension, the ratio of volume between the two spaces is $V_{\text{bg}}/V_{\text{user}} = a$, which means that given an arbitrary point in the 1-D space, the chances that it belongs to the background class ω_{bg} is a times of the chance that it belongs to the user class ω_{user} . From all the N dimensions, however, the ratio becomes $V_{\text{bg}}/V_{\text{user}} = a^N$. When N is large, e.g., $N = 1000$, and a takes a moderate value, e.g., $a = 1.5$, $a^N = 1.5^{1000} \sim 10^{176}$ is almost infinite. This implies that for an arbitrary N -dimensional feature vector, the chance that it falls into the user class ω_{user} is almost none. In other words, the discrimination capability of such a likelihood-ratio classifier in the high-dimensional space is very high, whereas to generalize, the feature vector of the user image taken under different situations must be able to stay within an extremely small region.

This trait, on the other hand, justifies the proposed illumination normalization method in Section IV, in which more emphasis is put on maintaining the generalization capability, rather than the discrimination capability. The large reduction of the image information (restricted LBP values) by the simplified LBP filter makes both class much smaller in volume after illumination normalization. In comparison to the user class, the background class is more substantially reduced, as

the method also discards certain information that is useful for discriminating different subjects. Consequently, the relative volume between ω_{bg} and ω_{user} is reduced, or equivalently, a is reduced. When a^N is not so prohibitively high, the generalization becomes easier. Most importantly, the discarded information contains a large illumination-sensitive component, which greatly increases the generalization capability across different illuminations. Meanwhile, enough discrimination capability is preserved because of the high dimensionality of the space.

VI. INFORMATION FUSION

Fusion is a popular practice to increase the reliability of the biometric verification [37], [45] and has been applied in face recognition tasks [28], [31], [32]. In our face authentication system, as shown in Fig. 2, the fusion is done between different frames. This not only improves the system performance, but also realizes the ongoing authentication, as introduced in Section I.

From each frame, we obtain a value of its likelihood ratio and compare it to a threshold to make the decision.² We have compared the following three different types of fusion methods 1) sum of the scores [17] and 2) AND and 3) OR of the decisions, which are equivalent to the min and max rules of the scores [43]. Theoretically, summation of log likelihood ratios acts in a similar way as a naive Bayes classifier [16] and can achieve the nearly optimal performance despite certain dependencies between consecutive frames [15]. In practice, however, we observed that the OR rule decision fusion yields still better results in our authentication system. This is, again, explained by the fact that the classifier possesses strong discrimination capability that tends to reject the authentic user occasionally. An OR operation, obviously, makes the classifier more prone to accept than to reject. As a result, it decreases the FRR, but at a very low FAR, because the chance that the two successive frames is falsely accepted are even lower than what we discussed in Section V. To illustrate the fusion performances, ROCs will be shown in Section VII.

VII. EXPERIMENTS AND RESULTS

A. Data Collection

To learn the probability density functions of the user class $p(x|\omega_{\text{user}})$ and the background class $p(x|\omega_{\text{bg}})$, a large number of samples are required.

The background sample set is taken from public face databases. In the experiments, we adopt the following four databases: 1) the BioID database [46]; 2) the FERET database [47]; 3) the YaleB database [20]; 4) and the FRGC database [48]. The faces are detected and registered using the methods proposed in Sections II and III. Each face is registered to the size of 32×32 and elongated into a 1024-dimensional feature vector. The databases, in total, result in more than 10 000 samples for training in the background class.

²To compute the ROC, the threshold is a changing value [16], and its value in the final system should be calibrated on the ROC based on certain requirements of the FAR or FRR.

Fig. 9. Examples of four session images of the same subject.

The user sample set is obtained by taking face images from the MPD. We used the Eten M600 PDA as the mobile device. In practice, the user set is convenient to collect: at a frame rate of 15 fps, a 2-min video results in 1800 frames of face images. In total, the data of 20 users have been collected from volunteers, with four independent sessions for each subject, taken at different times and under different illuminations. Fig. 9 shows examples of the same subject taken in four sessions. Additionally, to increase the variance of the samples as well as the tolerance to registration, the training set is extended by flipping and creating slightly shifted, rotated, and scaled versions of the face image.

B. Test Protocol

For one specific user, we learn the density of the user class from the user data of one session and compute the testing genuine scores, i.e., the log likelihood ratios, from the user data of the other three independent sessions. The density of the background class is learned from the public database, and the impostor scores are computed from our collected data of the other 19 subjects. As a result, for the user, we have around 1800 images for training and 5400 for validation. On the impostor side, we have around 10 000 public database faces for training and 100 000 collected faces of other subjects for validation. Note that the training and testing data of the background class are independent in the setting. Using the public databases as the training data is convenient for the MPD implementation, as the background parameters need to be calculated only once and stored for all the users. On the other hand, obtaining the impostor scores from our own database is of more interest than obtaining them from the public database because the impostor face images in our own database are collected under more or less the same situations as in the enrollment and, thus, more meaningful and critical for testing the verification performance.

Given the likelihood ratios obtained from the two classes, the ROC can be obtained to evaluate the performance of the system. Information fusion is further done, as described in Section VI.

C. Results on Mobile Data

We show the system performance of fusing two frames with certain intervals t . The longer the interval is, the lower the

dependency. We test different time intervals t of 0.2, 1, and 30 s. Fig. 10 shows the scatter plot of log likelihood ratios of two frames, as well as the ROCs of fusion at different time intervals t . The ROCs are drawn in the logarithm scale. It can be observed that the AND rule decision fusion does not bring any improvement; instead, it degrades the performance due to the fact that it increases the discrimination capability that is already very high. In comparison, the OR rule decision fusion works particularly well and outperforms the sum rule. As can be further observed, with the increase of t , the improvement of performance by fusion becomes more pronounced. When $t = 1$ s, the equal-error rate (EER) of the ROC by the OR rule decision fusion is already reduced to half of the original. For the MPD application, if we do ongoing authentication at a time interval of 30 s, an EER of 2% can be achieved. Fusing more frames will further increase the performance. Experiments have also shown that a much lower EER can be achieved if the enrollment and testing are done across sessions of closer illuminations.

D. Results on YaleB Data

The algorithm has also been tested on Yale database B [20], which contains the images of ten subject, each seen under 576 viewing conditions (9 poses $\times$ 64 illuminations). For the YaleB database, which emphasizes on illumination, we compare three different illumination normalization methods, namely, simplified LBP preprocessing, original LBP preprocessing, and histogram equalization. Examples of Yale database B and the effects of illumination normalization are shown in Fig. 7. The result of unprocessed images is also presented for reference.

In our test, for each subject, the user data are randomly partitioned into 80% for training and 20% for testing. The data of the other nine subjects are used as the impostor data. The face verification is exactly the same as in the mobile device case. To illustrate the performances, we used the EER as the performance measure. The random partition process is carried out 20 rounds for each subject. We obtain an average EER for each of the ten subjects. As a result, the performances of different illumination methods are compared in Fig. 11.

It can be seen in Fig. 11 that for all the subjects in the Yale database B, the simplified LBP preprocessing consistently achieves the best performance. This indicates that the simplified LBP preprocessing has higher robustness to large illumination variability. As the proposed system balances between generalization and discrimination by putting much emphasis on generalization among different imaging situations in the *illumination normalization* part, while on discrimination between user and impostor in the *face verification* part, the experimental results on the YaleB database indicates that the distribution of emphasis is effective in practice, i.e., on a system level, the generalization capability is guaranteed at little loss of discrimination capability.

E. Implementation

The efficiency of the proposed system enables realistic implementation of this system on an MPD. We chose the Eten M500 Pocket PC for demonstration and transformed our

Fig. 10. (Left column) Scatter plot and (right column) ROC comparison at different time intervals t : 0.2, 1, 30 s. (a) $t = 0.2$ s. (b) $t = 1$ s. (c) $t = 30$ s.

algorithms that are written in the C language onto the Windows Mobile 5 platform of the device. We used the Intel OpenCV library [26] to facilitate the implementation.

In the preliminary experiments, the enrollment is on the PC: the MPD takes a sequence of the user images of about

2 min and transfers them to the PC to process them, with the user mean and covariance extracted for calculating the Mahalanobis distance in the user class. The background mean and covariance have been prestored in the mobile device for calculating the Mahalanobis distance in the background class.

Fig. 11. Comparison of the ROCs using different illumination normalization methods (top down): simplified LBP, Gaussian derivative filter, zero-mean and unit-variance normalization, original LBP, histogram equalization, high-pass filtering, and the unprocessed.

The user image sequences then pass through the diagram in Fig. 2 until the final decision of acceptance or rejection is made. The implementation of the system in the project framework has been reported in [14].

Even without optimization, our system has already achieved a frame rate of 10 frames/s on the laptop with an Intel(R) central processing unit, 1.66 GHz, and 2 GB of random access memory (RAM). On the mobile device, with the Samsung S3C2440 400 MHz processor and 64 MB of synchronous dynamic RAM, the time is longer—about 8 s a frame. Profiling of the system indicates that the face detection and registration are still the most time-consuming part, compared to the illumination normalization and verification components that are extremely fast. This system will become practical in use with further optimization on both hardware and software, particularly when detection and registration can be implemented by hardware, transferring the Haar-like features into fast circuit units.

VIII. CONCLUSION

Face verification on the MPD provides a secure link between the user and the PN. In this paper, we have presented a biometric authentication system, from face detection, face registration, illumination normalization, face verification, to information fusion. Both theoretical concerns and practical concerns are given.

The series of solutions to the five modules proves to be efficient and robust. In addition, the different modules collaborate with each other in a systematic manner. For example, downscaling the face in the detection module provides very fast localization of the face, at the cost of coarser scales, but the subsequent registration module immediately compensates for the accuracy. The same is true for the illumination normalization and verification modules, where the latter well accommodates the former. The final system achieves an equal error rate of

2%, under challenging testing protocols. The low hardware and software cost of the system makes it well adaptable to a large range of security applications.

REFERENCES

- [1] R. Basri and D. Jacobs, "Lambertian reflectances and linear subspaces," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2001, pp. 383–390.
- [2] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 245–260, Jul. 1998.
- [3] G. Beumer, A. Bazen, and R. Veldhuis, "On the accuracy of EERs in face recognition and the importance of reliable registration," in *Proc. SPS IEEE Benelux DSP Valley*, 2005, pp. 85–88.
- [4] G. Beumer, Q. Tao, A. Bazen, and R. Veldhuis, "A landmark paper in face recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2006, pp. 73–78.
- [5] R. Bolle, J. Connell, and N. Ratha, "Biometric perils and patches," *Pattern Recognit.*, vol. 35, no. 12, pp. 2727–2738, Dec. 2002.
- [6] M. Burl, T. Leung, and P. Perona, "Face localization via shape statistics," in *Proc. Int. Workshop Autom. Face Gesture Recog.*, 1995, pp. 154–159.
- [7] T. Chan, J. Shen, and L. Vese, "Variational PDE models in image processing," *Not. Amer. Math. Soc.*, vol. 50, no. 1, pp. 14–26, 2003.
- [8] H. Chen, P. Belhumeur, and D. Jacobs, "In search of illumination invariants," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2000, pp. 254–261.
- [9] T. Cootes, G. Edwards, and J. Taylor, "Active appearance models," in *Proc. Eur. Conf. Comput. Vis.*, 1998, pp. 484–498.
- [10] T. Cootes and J. Taylor, "Active shape models—Smart snakes," in *Proc. Brit. Mach. Vis. Conf.*, 1992, pp. 266–275.
- [11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [12] D. Cristinacce and T. Cootes, "Facial feature detection using Adaboost with shape constraints," in *Proc. 14th Brit. Mach. Vis. Conf.*, 2003, pp. 231–240.
- [13] D. Cristinacce, T. Cootes, and I. Scott, "A multi-stage approach to facial feature detection," in *Proc. 15th Brit. Mach. Vis. Conf.*, 2004, pp. 277–286.
- [14] F. den Hartog, M. Blom, C. Lageweg, M. Peeters, J. Schmidt, R. van der Veer, A. de Vries, M. R. van der Werff, Q. Tao, R. Veldhuis, N. Baken, and F. Selgert, "First experiences with personal networks as an enabling platform for service providers," in *Proc. 2nd Int. Workshop Personalized Netw.*, Philadelphia, PA, 2007, pp. 1–8.
- [15] P. Domingos and M. Pazzani, "Beyond independence: Conditions for the optimality of the simple Bayesian classifier," in *Proc. 13th Int. Conf. Mach. Learn.*, 1996, pp. 105–112.
- [16] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [17] M. Faundez-Zanuy, "Data fusion in biometrics," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 1, pp. 34–38, Jan. 2005.
- [18] R. Feris, J. Gemmell, K. Toyama, and V. Krueger, "Hierarchical wavelet networks for facial feature localization," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2001, pp. 118–123.
- [19] Freeband, PNP2008: Development of a User Centric Ambient Communication Environment. [Online]. Available: <http://www.freeband.nl/project.cfm?language=en&id=530>
- [20] A. Georgiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [21] V. Govindaraju, "Locating human faces in photographs," *Int. J. Comput. Vis.*, vol. 19, no. 2, pp. 129–146, Aug. 1996.
- [22] G. Heusch, F. Cardinaux, and S. Marcel, "Lighting normalization algorithms for face verification," IDIAP, Martigny, Switzerland, Tech. Rep. 03, 2005.
- [23] G. Heusch, Y. Rodriguez, and S. Marcel, "Local binary patterns as image preprocessing for face authentication," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2006, pp. 9–14.
- [24] E. Hielmas and B. Low, "Face detection: A survey," *Comput. Vis. Image Underst.*, vol. 83, no. 3, pp. 235–274, Sep. 2001.
- [25] M. Hunke and A. Waibel, "Face locating and tracking for human-computer interaction," in *Proc. 28th Asilomar Conf. Signals, Syst., Images*, Monterey, CA, 1994, pp. 1277–1281.

- [26] Intel, Open Computer Vision Library. [Online]. Available: <http://sourceforge.net/projects/opencvlibrary>
- [27] S. King, G. Y. Tian, D. Taylor, and S. Ward, "Cross-channel histogram equalisation for colour face recognition," in *Proc. AVBPA*, 2003, pp. 454–461.
- [28] A. Kumar and T. Srikanth, "Online personal identification in night using multiple face representation," in *Proc. Int. Conf. Pattern Recog.*, Tampa, FL, 2008.
- [29] E. Land and J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, Jan. 1971.
- [30] J. Linnartz and P. Tuyts, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Proc. 4th Conf. Audio Video-Based Biometric Person Verification*, Guildford, U.K., 2003, pp. 393–403.
- [31] A. Lumini and L. Nanni, "Combining classifiers to obtain a reliable method for face recognition," *Multimed. Cyberscape J.*, vol. 3, no. 3, pp. 47–53, 2005.
- [32] G. Marcialis and F. Roli, "Fusion of appearance-based face recognition algorithms," *Pattern Anal. Appl.*, vol. 7, no. 2, pp. 151–163, Jul. 2004.
- [33] S. McKenna, S. Gong, and J. Collins, "Face tracking and pose representation," in *Proc. Brit. Mach. Vis. Conf.*, Edinburgh, U.K., 1996, vol. 2, pp. 755–764.
- [34] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," in *Proc. Eur. Conf. Comput. Vis.*, 1994, pp. 286–296.
- [35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [36] T. Riopka and T. Boulton, "The eyes have it," in *Proc. ACM SIGMM Multimed. Biometrics Methods Appl. Workshop*, 2003, pp. 9–16.
- [37] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*. New York: Springer-Verlag, 2006, ser. International Series on Biometrics.
- [38] T. Sakai, M. Nagao, and S. Fujibayashi, "Line extraction and pattern detection in a photograph," *Pattern Recognit.*, vol. 1, no. 3, pp. 233–248, Mar. 1969.
- [39] R. Schapire, Y. Freund, P. Bartlett, and W. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, pp. 322–330.
- [40] A. Shashua, "Geometry and photometry in 3D visual recognition," Ph.D. dissertation, MIT, Cambridge, MA, 1997.
- [41] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129–139, Feb. 2001.
- [42] Q. Tao and R. Veldhuis, "A study on illumination normalization for 2D face verification," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, Madeira, Portugal, 2008, pp. 42–49.
- [43] Q. Tao and R. N. J. Veldhuis, "Threshold-optimized decision-level fusion and its application to biometrics," *Pattern Recognit.*, vol. 42, no. 5, pp. 823–836, May 2009.
- [44] D. Tax, "One class classification," Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, 2001.
- [45] B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan, "Studies of biometric fusion," NIST, Gaithersburg, MD, NIST Tech. Rep. IR 7346, 2006.
- [46] BioID, BioID Face Database. [Online]. Available: <http://www.humanscan.de/>
- [47] FERET, FERET Face Database. [Online]. Available: <http://www.itl.nist.gov/iad/humanid/feret/>
- [48] FRGC, FRGC Face Database. [Online]. Available: <http://face.nist.gov/frgc/>
- [49] H. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1969.
- [50] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
- [51] L. Wiskott, J. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.
- [52] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.
- [53] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

Qian Tao received the B.Sc. and M.Sc. degrees from Fudan University, Shanghai, China, in 2001 and 2004, respectively.

In 2004, she joined the Signals and Systems Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands, as a Ph.D. student. She participated in the personal network pilot (PNP) project of the Freeband, The Netherlands, and the main focus of her work is face authentication via mobile devices. Her research interests include biometrics, artificial intelligence, image processing, and statistical pattern recognition.

Raymond Veldhuis received the M.Eng. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1981 and the Ph.D. degree from Nijmegen University, Nijmegen, The Netherlands, in 1988, with a thesis titled, "Adaptive restoration of lost samples in discrete-time signals and digital images."

From 1982 to 1992, he was a Researcher with Philips Research Laboratories, Eindhoven, The Netherlands, in various areas of digital signal processing, such as audio and video signal restoration and audio source coding. From 1992 to 2001, he was with the Institute of Perception Research (IPO), Eindhoven, working on speech signal processing and speech synthesis. From 1998 to 2001, he was a Program Manager of the Spoken Language Interfaces research program. He is currently an Associate Professor with the Signals and Systems Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, working in the fields of biometrics and signal processing. He is the author of over 120 papers published in international conference proceedings and journals. He is also a coauthor of the book *An Introduction to Source Coding* (Prentice-Hall) and the author of the book *Restoration of Lost Samples in Digital Signals* (Prentice-Hall). He is the holder of 21 patents in the fields of signal processing. His expertise involves digital signal processing for audio, images and speech; statistical pattern recognition and biometrics.