

Likelihood-Ratio-Based Biometric Verification

Asker M. Bazen and Raymond N. J. Veldhuis

Abstract—This paper presents results on optimal similarity measures for biometric verification based on fixed-length feature vectors. First, we show that the verification of a single user is equivalent to the detection problem, which implies that for single-user verification the likelihood ratio is optimal. Second, we show that, under some general conditions, decisions based on posterior probabilities and likelihood ratios are equivalent and result in the same receiver operating curve. However, in a multi-user situation, these two methods lead to different average error rates. As a third result, we prove theoretically that, for multi-user verification, the use of the likelihood ratio is optimal in terms of average error rates. The superiority of this method is illustrated by experiments in fingerprint verification. It is shown that error rates below 10^{-3} can be achieved when using multiple fingerprints for template construction.

Index Terms—Biometric verification, fingerprint recognition, fixed-length feature vector, likelihood ratio, posterior probability density.

I. INTRODUCTION

BIOMETRIC verification systems are used to verify the claimed identity of a user by measuring specific characteristics of the body, such as a fingerprint, face, hand geometry, iris, or pressure patterns. Given a *test* feature vector \mathbf{v} that is obtained from a user requesting access to a biometric system, and a class w that represents the users claimed identity (represented by a *template* feature vector that has been recorded during the enrollment phase), the task of a biometric verification system is to decide whether the offered feature vector can be accepted as a member of the given class or not. For this purpose, the system determines a measure that represents the similarity between the test and the template measurements. If the similarity is larger than an acceptance threshold, the user is accepted; otherwise, the user is rejected.

The subject of this paper is the comparison of different similarity measures that can be used. This will be demonstrated in the context of fingerprint recognition. Most fingerprint matching systems use minutiae-based algorithms [1], [2], which are in general considered to be the most reliable. However, comparing two sets of minutiae is not a straightforward task. First, the number of minutiae that are extracted depends on the actual fingerprint. Second, it is not known beforehand which minutia in the template set corresponds to which one in the test set. Third, even if the first two problems are solved, the minutiae locations and associated characteristics cannot be compared directly due to translation, rotation, and scaling

of the fingerprints. Instead, comparing two sets of minutiae requires special point-pattern matching algorithms.

In contrast, this paper presents a fingerprint matching algorithm that uses fixed-length feature vectors, consisting of a number of measurements that are performed at some specific, fixed, locations in the fingerprint. The advantage of this approach is that, once the features are extracted (including the critical step of determining the accurate location of a reference point), the matching is very fast. This enables the search for a matching fingerprint in a large database.

This paper presents results on optimal similarity measures for general biometric verification based on fixed-length feature vectors. In Section II, three similarity measures are reviewed: Euclidean distance, posterior probability, and likelihood ratio. In that section, the general expressions for biometric system errors are derived. Next, the optimality of likelihood ratio-based decisions is proved in Section III. Then, Section IV presents experimental results on fingerprint verification, confirming the theory that using likelihood ratios instead of posterior-probability densities decreases the error rates, even though the feature vectors are the same for both decision methods.

II. SIMILARITY MEASURES

Various similarity measures for fixed-length feature vectors have been proposed in the literature. First, we give an overview of the three most widely used measures, being Euclidean distance, posterior probabilities, and likelihood ratios. Next, expressions for the biometric system errors are given.

A. Euclidean Distance

In [3], FingerCode is used as a feature vector for fingerprint verification. This feature vector contains the standard deviation of the responses of Gabor filters in specified locations in the fingerprint. For comparison of these feature vectors, the Euclidean distance is used. It is calculated by

$$d_{\text{Eucl}} = \|\mathbf{v} - \mathbf{c}\|^2 \quad (1)$$

where \mathbf{v} is the test vector and \mathbf{c} is the template vector corresponding to class w . The Euclidean distance treats all elements of the feature vector as equally important and uncorrelated. Although this is not a realistic assumption in practice, the authors present experiments with relatively good recognition performance.

B. Posterior Probability

In [4], biometric verification systems that are based on hand geometry and face recognition are presented. In that paper, it is claimed that decisions that are based on the posterior probability densities are optimal, where optimality means minimal

Manuscript received October 29, 2002; revised May 7, 2003.

The authors are with University of Twente, 7500 AE Enschede, The Netherlands (e-mail: a.m.bazen@utwente.nl, r.n.j.veldhuis@utwente.nl).

Digital Object Identifier 10.1109/TCSVT.2003.818356

error rates as defined in Section III. The posterior probability density of class w given observed feature vector \mathbf{v} is given by

$$p(w|\mathbf{v}) = \frac{p(\mathbf{v}|w) \cdot p(w)}{p(\mathbf{v})} \quad (2)$$

where $p(\mathbf{v}|w)$ is the probability density of the feature vectors given class w , $p(w)$ is the probability density of the class w , and $p(\mathbf{v})$ is the prior probability density of the feature vectors. The feature vector \mathbf{v} is accepted as member of the template class if its posterior probability density exceeds a threshold $t \in [0, t_{\max}]$.

C. Likelihood Ratio

On the other hand, in detection theory it has long since been known that the use of likelihood ratios is asymptotically optimal [5]. In detection, a given feature vector has to be classified as originating from a predefined situation (the presence of some object to be detected) or not. Since the detection problem is in some sense equivalent to the verification problem that is considered here, it is to be expected that using likelihood ratios for biometric verification is optimal as well. The likelihood ratio $L(\mathbf{v})$ is given by

$$L(\mathbf{v}) = \frac{p(\mathbf{v}|w)}{p(\mathbf{v}|\bar{w})} \quad (3)$$

where $p(\mathbf{v}|\bar{w})$ is the probability of \mathbf{v} , given that \mathbf{v} is not a member of class w . Since we assume infinitely many classes, exclusion of a single class w does not change the distribution of the feature vector \mathbf{v} . Therefore, the distribution of \mathbf{v} , given that \mathbf{v} is not a member of w , equals the prior distribution of \mathbf{v} as

$$p(\mathbf{v}|\bar{w}) = p(\mathbf{v}) \quad (4)$$

and the likelihood ratio is given by

$$L(\mathbf{v}) = \frac{p(\mathbf{v}|w)}{p(\mathbf{v})}. \quad (5)$$

In this framework, a test feature vector \mathbf{v} is accepted as member of the template class if its likelihood ratio exceeds a threshold $t \in [0, \infty)$. The acceptance region $A_{t,w}$ and rejection region $R_{t,w}$ can be defined in the feature space \mathbf{V} as

$$A_{t,w} = \{\mathbf{v} \in \mathbf{V} \mid L(\mathbf{v}) \geq t\} \quad (6)$$

$$R_{t,w} = \{\mathbf{v} \in \mathbf{V} \mid L(\mathbf{v}) < t\}. \quad (7)$$

The probability density functions in (2) and (5) are in practice usually modeled by multidimensional Gaussian distributions. More details of that case are given in Appendix A.

D. Biometric System Errors

In this section, we derive expressions for the system errors, using likelihood ratio-based decisions, as a function of the probability density functions of the feature vectors.

1) *False Rejection Rate*: The *false rejection rate* (FRR) measures the probability that a feature vector is rejected as a member of some class, although it does originate from that class. For a specific class w and a given threshold t , $\text{FRR}(t, w)$ is given by

$$\text{FRR}(t, w) = P(\mathbf{v} \in R_{t,w} \mid \mathbf{v} \in w) = \int_{R_{t,w}} p(\mathbf{v}|w) d\mathbf{v}. \quad (8)$$

Since $A_{t,w} + R_{t,w} = \mathbf{V}$, this can also be written as

$$\text{FRR}(t, w) = 1 - \int_{A_{t,w}} p(\mathbf{v}|w) d\mathbf{v}. \quad (9)$$

The (average) overall false rejection rate $\text{FRR}(t)$ is found by integrating over all classes as

$$\text{FRR}(t) = \int_W \text{FRR}(t, w) \cdot p(w) dw \quad (10)$$

where W is the space of all classes. The summation over all (discrete) classes is represented by a (continuous) integral to indicate the infinite number of classes.

2) *False Acceptance Rate*: The *false acceptance rate* (FAR) measures the probability that a feature vector is accepted as a member of some class, although it does not originate from that class. For a specific class w and a given a threshold t , $\text{FAR}(t, w)$ is given by

$$\text{FAR}(t, w) = P(\mathbf{v} \in A_{t,w} \mid \mathbf{v} \notin w) = \int_{A_{t,w}} p(\mathbf{v}) d\mathbf{v} \quad (11)$$

where again $p(\mathbf{v}|\bar{w}) = p(\mathbf{v})$ is used. Then, the (average) global false acceptance rate $\text{FAR}(t)$ is found by integrating over all classes to yield

$$\text{FAR}(t) = \int_W \text{FAR}(t, w) \cdot p(w) dw. \quad (12)$$

3) *Receiver Operating Characteristic*: The dependence of both error rates on the threshold can be visualized in a plot of FRR against FAR for varying threshold values, which is called the *receiver operating characteristic* (ROC). Using a procedure similar to the work in [4] on posterior probabilities, an expression that describes the tradeoff between FAR and FRR for a likelihood-ratio-based verification system can be derived as

$$\frac{d\text{FRR}(t, w)}{d\text{FAR}(t, w)} = -t. \quad (13)$$

The same expression has been derived in [5] using other methods.

III. OPTIMALITY OF LIKELIHOOD RATIOS

In this section, we prove that, for verification based on fixed-length feature vectors, the use of the likelihood ratio is optimal in terms of average overall error rates. In this context, optimality is defined as the lowest FAR for a given FRR or alternatively the lowest FRR for a given FAR. First, we consider the less complex case of single-user verification, where the system has to decide whether or not an input feature vector originates from the only user that is known to the system.

A. Single-User Verification

For single-user verification, there is one fixed distribution $p(\mathbf{v}|w)$ of feature vectors from the genuine user and one fixed distribution $p(\mathbf{v})$ of feature vectors from impostors. This situation is equivalent to the detection problem, which implies that the likelihood ratio is optimal for single-user verification [5].

The relation between the likelihood ratio and the posterior probability, which is derived from (2) and (5), is given by

$$p(w|\mathbf{v}) = p(w) \cdot L(\mathbf{v}). \quad (14)$$

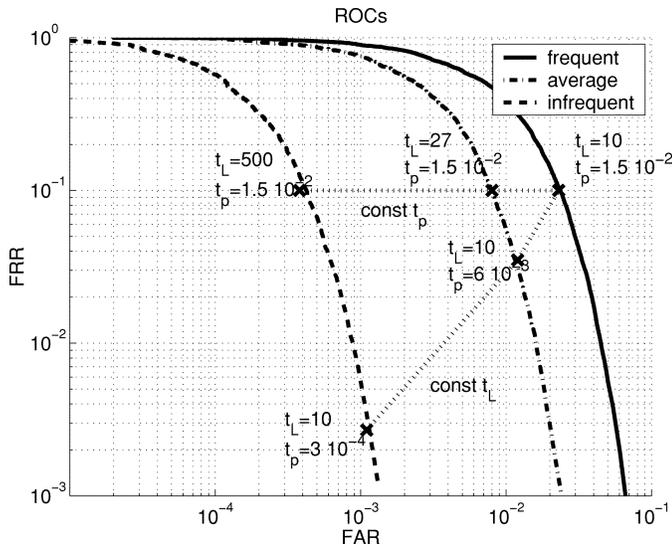


Fig. 1. Averaging ROCs of classes with different occurrence probabilities.

Since there is only one user in the system, $p(w)$ is a constant, and both methods provide the same error rates if the thresholds t_p for posterior probability density and t_L for likelihood ratio are set to $t_p = p(w) \cdot t_L$. This means that, for single-user verification, both methods provide the same ROC, with a different threshold parameterization along the curve. This is shown in Fig. 1, for instance by observing the leftmost curve, with the associated threshold values for both methods.

Please note that the ROCs are different for each of the users. Although the within-class variations are equal, the users with more common biometric characteristics are more difficult to identify than users with uncommon or rare characteristics, leading to higher biometric error rates. In other words, the error rate specifications of a biometric verification system only give average numbers. The error rates are higher for some of the users, while they are lower for others.

B. Multiuser Verification

In a multiuser situation, the two methods lead to different average error rates. The difference between averaging with the two methods is illustrated in Fig. 1, which shows three ROCs for frequently, average, and infrequently occurring classes in a synthetic data set. At some specific positions at the ROCs, the threshold values for both methods are indicated. It can be seen that combining individual ROCs with a constant posterior probability threshold t_p will take the average of different points on the individual ROCs than combining individual ROCs with a constant likelihood ratio threshold t_L . The question that is answered in the rest of this section is which of all possible averaging paths will lead to the minimum overall error rates.

It can be explained intuitively that the use of likelihood ratios will lead to better overall matching performance than the use of posterior probabilities. A fixed posterior-probability threshold requires equal similarity values for feature vectors from each individual, under the condition that the distribution of the feature vectors from the entire population is much wider than the distribution within one class. This means that a feature vector is relatively close to its associated class center, such that $p(\mathbf{v}) \approx p(w)$

[see (2)]. This condition is easily satisfied in practice. The effect is an equal FRR and a lower FAR for less frequently occurring classes.

On the other hand, a fixed likelihood-ratio threshold requires lower similarity values for less frequently occurring feature vectors [see (14)]. As a consequence, the acceptance region for less frequently occurring classes is larger, which has two effects. The reduction in FAR is smaller, but at the same time, FRR is reduced. The overall recognition performance can be optimized by choosing the right tradeoff between these two effects.

Next, we prove that using likelihood ratios in multiuser verification is optimal. First, define $\varphi(L|w)$ as the probability density function of the likelihood ratio of an observation vector \mathbf{v} that is taken from the true class w . Also, define $\varphi(L|\bar{w})$ as the probability density function of the likelihood ratio of an observation vector \mathbf{v} that is not taken from the true class w . For these probability density functions, the following well-known relation holds [5]:

$$\varphi(L|w) = L \cdot \varphi(L|\bar{w}). \quad (15)$$

The error rates FRR and FAR for class w , as a function of the threshold t , are given by

$$\text{FRR}(t, w) = \int_0^t \varphi(L|w) dL \quad (16)$$

and

$$\text{FAR}(t, w) = \int_t^\infty \varphi(L|\bar{w}) dL. \quad (17)$$

Next, we find expressions for the average $\text{FRR}(t(\cdot))$ and $\text{FAR}(t(\cdot))$ with a class-dependent threshold $t(w)$ by integrating over all classes as

$$\text{FRR}(t(\cdot)) = \int_W p(w) \int_0^{t(w)} \varphi(L|w) dL dw \quad (18)$$

$$\text{FAR}(t(\cdot)) = \int_W p(w) \int_{t(w)}^\infty \varphi(L|\bar{w}) dL dw. \quad (19)$$

For optimal verification performance, the question is how to choose the threshold t as a function of w , such that the resulting ROC is minimal. This is solved by Lagrange optimization (see, for instance, [6]). The objective is to minimize FRR, subject to the condition of a constant FAR. The threshold is chosen as $t = t_{\text{opt}}(w) + \varepsilon f(w)$, where $t_{\text{opt}}(w)$ is the optimal threshold, $f(w)$ is some function of w , ε is a small constant, and some specific value for FAR is chosen as additional condition. Then

$$J = \int_W p(w) \int_0^{t_{\text{opt}}(w) + \varepsilon f(w)} \varphi(L|w) dL dw + \lambda \left[\int_W p(w) \int_{t_{\text{opt}}(w) + \varepsilon f(w)}^\infty \varphi(L|\bar{w}) dL dw - \text{FAR}(t) \right] \quad (20)$$

has to be minimized by setting the derivative with respect to ε to zero as

$$\int_W p(w) \varphi(t_{\text{opt}}(w) + \varepsilon f(w)|w) f(w) dw - \lambda \int_W p(w) \varphi(t_{\text{opt}}(w) + \varepsilon f(w)|\bar{w}) f(w) dw = 0. \quad (21)$$

By realizing that this expression must hold for any $f(w)$, the integrals over all w can be omitted. Furthermore, since t_{opt} is optimal, ε is equal to zero, which further simplifies the expression to

$$\varphi(t_{\text{opt}}(w)|w) - \lambda\varphi(t_{\text{opt}}(w)|\bar{w}) = 0. \quad (22)$$

Applying (15) results in

$$t_{\text{opt}}(w)\varphi(t_{\text{opt}}(w)|\bar{w}) - \lambda\varphi(t_{\text{opt}}(w)|\bar{w}) = 0 \quad (23)$$

which, by dividing both sides by $\varphi(t_{\text{opt}}(w)|\bar{w})$ and rearranging the expression, gives

$$t_{\text{opt}}(w) = \lambda. \quad (24)$$

Since λ is a constant, the optimal threshold $t_{\text{opt}}(w)$ is constant too, independent of w . Therefore, using a constant likelihood ratio threshold when averaging over the classes gives the optimal verification results.

IV. MATCHING EXPERIMENTS

In this section, results of fingerprint-matching experiments are presented. The proposed similarity measures, which are Euclidean distance, posterior probabilities, and likelihood ratios, have been evaluated by applying them to Database 2 of FVC2000 [7]. This FVC2000 database consists of 880 8-b grayscale fingerprints, 8 prints of each of 110 different fingers. The images are captured with a capacitive sensor at 500 dpi, resulting in image sizes of 364 by 256 pixels.

A. Feature Vectors

We use two types of feature vectors that have been extracted from the grayscale fingerprint images. The first feature vector is the squared directional field that is defined in [8], which is calculated at a regular grid of 11×11 points with spacings of 8 pixels and is centered at the core point (see [8]). At each position in the grid, the squared directional field is coded in a vector of two elements. The resulting feature vector of length 242 is reduced to dimension 100 by principal component analysis (PCA) over the entire population. This method linearly projects each feature vector to a lower dimensional subspace that is optimized to maintain the largest variations that are present in the population. The directional field of an example fingerprint is shown in Fig. 2. For approximately 10% of the fingerprints, the automatic core-point extraction failed and for those fingerprints, the location of the core point was adjusted manually. The automatic core-point extraction errors could be resolved by two related methods. First, feature vectors could be extracted at many regularly spaced locations from the fingerprint. That one feature vector that results in the highest matching score is used. This solution is inspired by the feature space correlation method that is described in [9]. Second, feature vectors could be extracted at each location where a (possibly false) core is detected. Again the best matching feature vector is used. This would save a lot of processing time compared to the first method.

The second feature vector is the Gabor response of the fingerprint, which is discussed in [10]. After subtraction of the local

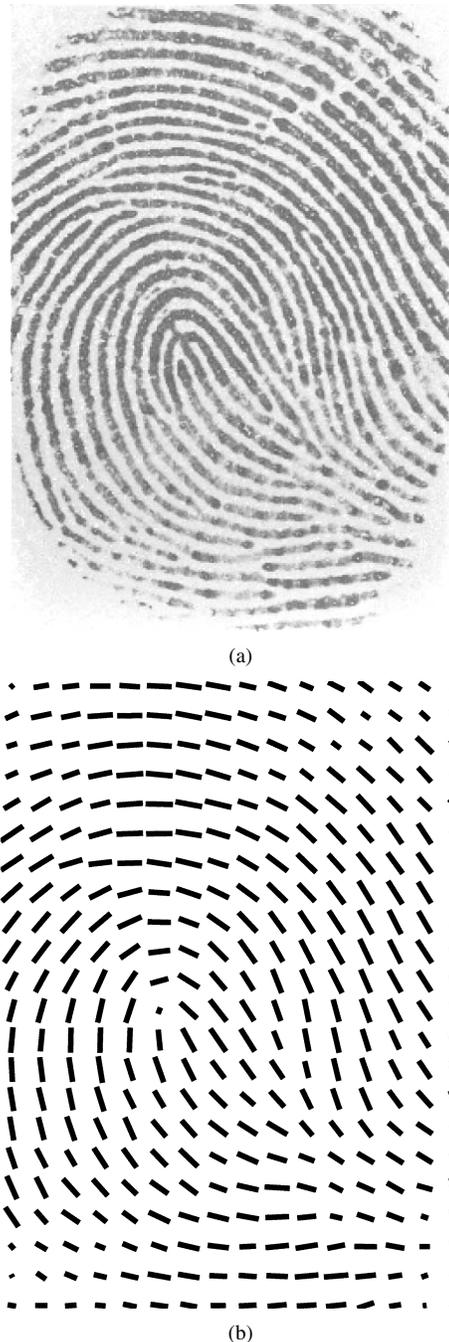


Fig. 2. (a) Fingerprint image and (b) its directional field.

mean, the fingerprint image is filtered by a set of four complex Gabor filters, which are given by

$$h_{\text{Gabor}}(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp(j2\pi f \cdot (x \sin \theta + y \cos \theta)). \quad (25)$$

The orientations θ are set to $0, \pi/4, \pi/2,$ and $3\pi/4$, the spatial frequency is set to $f = 0.125$, which corresponds to a ridge-valley period of 8 pixels, and the width of the filter is set to $\sigma = 3$. The absolute values of the output images are taken, which are subsequently filtered by a Gaussian window with $\sigma = 6$. The resulting images are shown in Fig. 3.

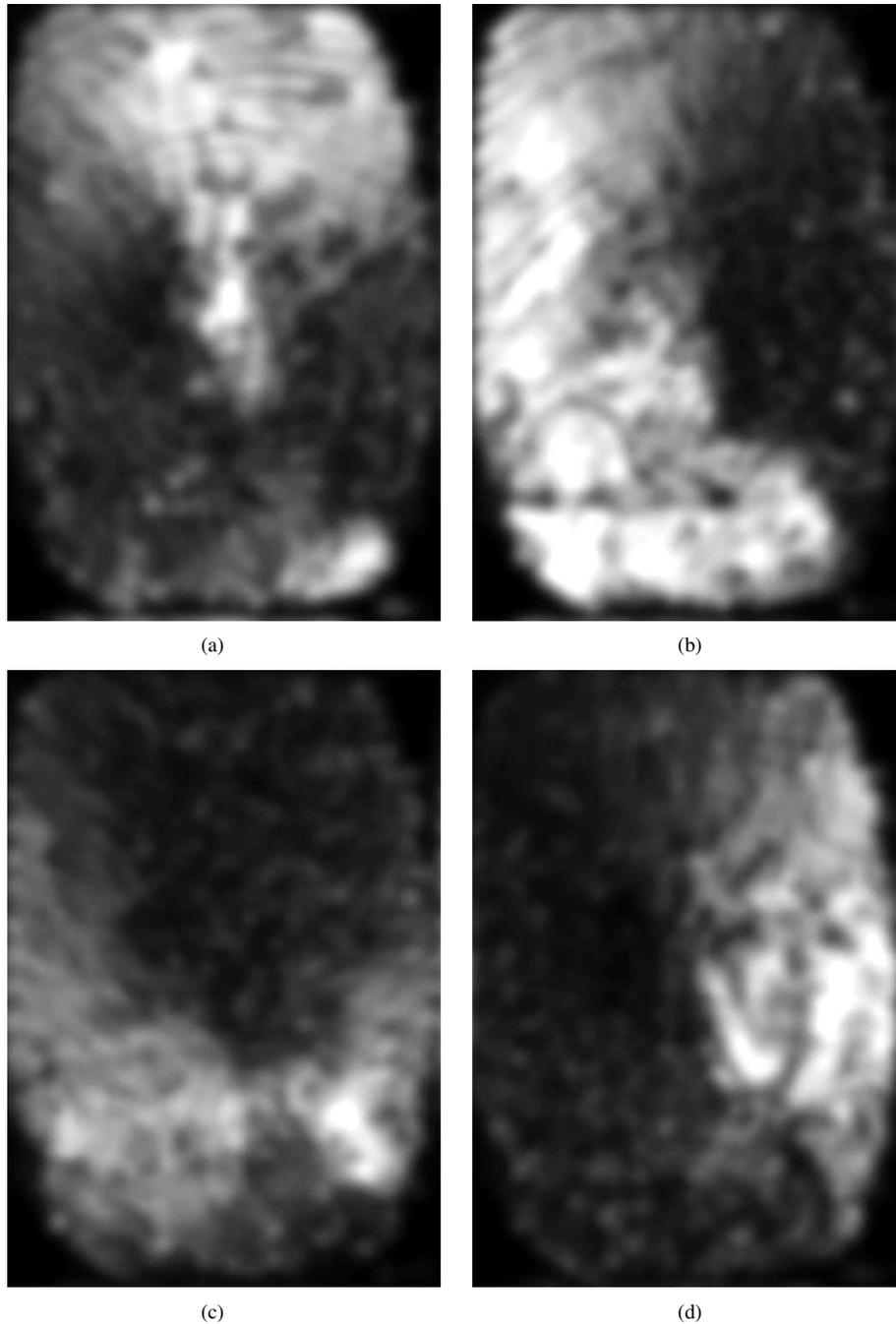


Fig. 3. Smoothed absolute value of the fingerprint image of Fig. 2(a) that has been filtered by Gabor filters with different orientations. (a) 0π . (b) $\pi/4$. (c) $\pi/2$. (d) $3\pi/4$.

Next, samples are taken at a regular grid of 11×11 points with spacings of 8 pixels and centered at the core point. The resulting feature vector of length 484 is reduced to dimension 200 by PCA of the entire population. This feature vector is inspired by FingerCode [3], but it can be calculated more efficiently since a rectangular grid is used rather than a circular one (see also [11]), and it performs slightly better.

B. Matching Algorithm

To enable calculation of the posterior-probability density and likelihood ratio, we assume a Gaussian probability density function (pdf) for the overall feature-vector distribution and

Gaussian pdfs with unequal means but equal covariance matrices for the feature vectors from all individual classes. The use of equal within-class covariance matrices for all users is motivated by the fact that in a biometric system in practice, only one or a few feature vectors are available as template. Therefore, no user-dependent covariance matrix can be determined reliably, and the best approximation possible is to use the average covariance matrix for all users.

There are a few motivations for using Gaussian distributions for the feature vectors. In general, measurements of natural features tend to a Gaussian distribution. Furthermore, as the dimension of the feature vectors is reduced by PCA, the feature vector

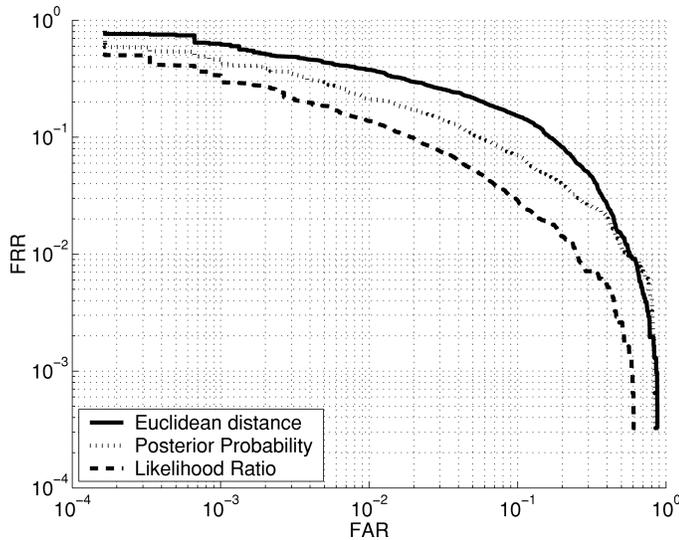


Fig. 4. Results of the directional field-based matching algorithm.

elements are weighted sums of measured features, which approximate a Gaussian distribution even better, as dictated by the central limit theorem. Appendix A provides expressions for the verification errors for Gaussian feature vectors.

The within-class covariance matrix Σ_W represents the differences between multiple prints of the same finger, like noise, partial impressions, and elastic deformations. The between-class covariance matrix Σ_B represents the differences between the templates of the different fingers, and the total covariance matrix Σ_T represents the variations of the feature vectors over the entire population, representing the differences between individual fingerprints. For both types of feature vectors, the three covariance matrices have been determined from the fingerprints in our database.

Next, the matching scores of 3080 genuine attempts and 5995 impostor attempts have been recorded using three similarity measures: the Euclidean distance d_{Eucl} , the posterior probability density $p(w|\mathbf{v})$, and the log-likelihood ratio $\Lambda(\mathbf{v})$. By incorporating the values of the constants into the threshold, and by using $p(\mathbf{v}) \approx p(w)$ if \mathbf{v} is close to w , the expressions to be tested are simplified to

$$d_{\text{Eucl}} = (\mathbf{v} - \mathbf{c})^T (\mathbf{v} - \mathbf{c}) \quad (26)$$

$$\log p(w|\mathbf{v}) \propto -(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c}) \quad (27)$$

$$\Lambda(\mathbf{v}) \propto -(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c}) + \mathbf{v}^T \Sigma_T^{-1} \mathbf{v}. \quad (28)$$

C. Matching Results

The resulting average ROC's of the matching experiment, using Euclidean distance, posterior probability, and likelihood ratios, are shown in Figs. 4 and 5. Both the directional field and the Gabor response perform at an equal-error rate (EER) of approximately 5% when using likelihood ratios. The figures show that the use of posterior probabilities results in a higher EER of 8%, while the Euclidean distance performs at an even worse EER of 15%. Fig. 5 shows that posterior probabilities perform better than likelihood ratios for FAR > 20%. For that setting, the acceptance region is very large, and the criterion $p(\mathbf{v}) \approx p(w)$

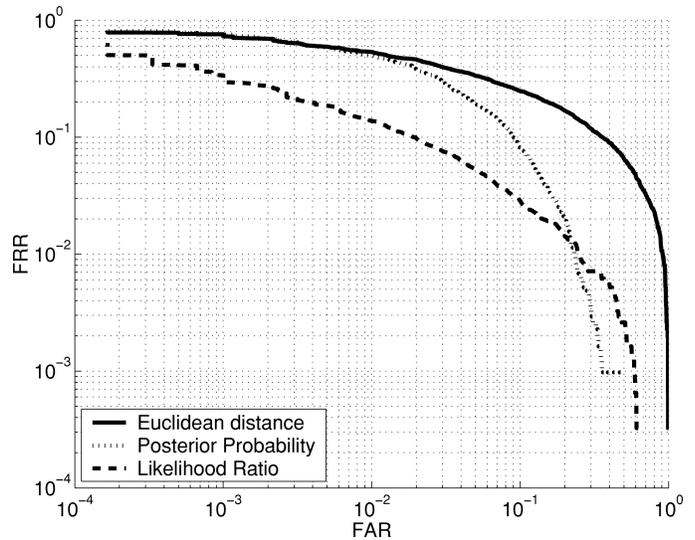


Fig. 5. Results of the Gabor response-based matching algorithm.

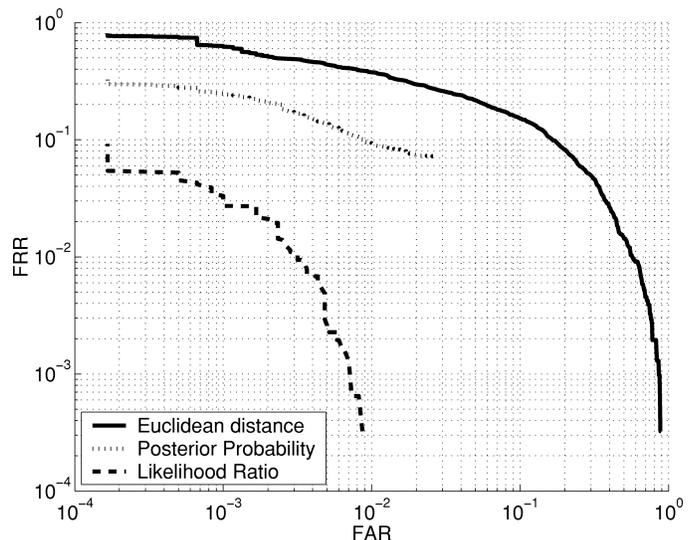


Fig. 6. Results of the matching algorithm that uses both the directional field and the Gabor response as features.

is no longer satisfied. Combined with possibly unequal within-class covariance matrices, this may explain the results.

Next, a new feature vector has been constructed by concatenating the directional field and Gabor response into one large feature vector of dimension 300, and new interclass and intraclass covariance matrices have been determined. For this combined feature vector, the performance differences are even more significant. Likelihood ratios perform at EER = 0.5%, posterior probability at EER = 7%, and Euclidean distance at EER = 12%, as shown in Fig. 6.

The matching algorithm takes less than 1 ms in Matlab on a 1-GHz P-III (optimized feature extraction takes 350 ms), compared to approximately 100 ms for a fast optimized minutiae matching algorithm (optimized feature extraction takes 450 ms); see, for instance, [1]. Using appropriate preprocessing such as simultaneous diagonalization which makes Σ_W identity and Σ_T diagonal [12], consecutive matches of a single test fingerprint to

a database of templates take less than $25 \mu\text{s}$ per template, which enables the search through a database that contains 40 000 entries in 1 s.

D. Unknown Versus Known Class Centers

For a biometric system in practice, the feature vectors that correspond to the true class centers are not known. More generally speaking, only one biometric measurement is available as template, and this is not necessarily a good representative of the class. As a solution, it has been shown in a number of papers that the use of multiple measurements from the same user for the construction of a template may improve the recognition rates. The combination approaches range from simply storing and comparing multiple templates per user [13] to the more complex combination of fingerprint images or minutiae sets [14]. The combination of fixed-length feature vectors that is presented here (averaging the feature vectors) is very simple, but at the same time very effective since it provides the optimum estimate of the class center.

In the context of likelihood ratios, the availability of only one template means that all elements of the intraclass covariance matrix are twice as large, compared to the situation with known class centers. As a result, the acceptance region has to be $2^{d/2}$ times as large (with d the dimension of the feature vector) for a specific FRR. If the interclass feature vector distribution is much wider than the intraclass distribution, the impostor distribution can be assumed constant for thresholds that correspond to a relatively high likelihood ratio. Consequently, FAR is also $2^{d/2}$ times as large as it would be with known class centers. This is confirmed in experiments with low-dimensional synthetic data sets.

The combined feature vector has dimension $d = 300$. This means that, at least theoretically, FAR can be reduced by a factor of $2^{300/2} \approx 10^{45}$ when using class centers as templates. To validate this effect, experiments have been set up where multiple fingerprints have been used for construction of the template. The template is taken as the average of multiple individual feature vectors, while only one of the remaining fingerprint is used as test vector. For this experiment, 6160 genuine attempts and 5995 impostor attempts have been carried out. The ROCs for the three similarity measures, using two fingerprints to construct the template, are shown in Fig. 7, and the equal error rates for likelihood ratios with one to five fingerprints for template construction are shown in Table I. The table shows that the performance gain in practice is not as large as it is in theory. But still, the matching performance can be increased enormously by using multiple fingerprints for template construction.

Two more remarks have to be made on this subject. First, the performance is evaluated in a database of 880 fingerprints, using approximately 6000 genuine and 6000 impostor attempts. In this evaluation set, error rates smaller than 0.1% cannot be estimated reliably. Therefore, the 0% in Table I does not mean that we have implemented the perfect biometric system, but only that it made no errors on our database. Second, the practical performance gain of using multiple feature vectors for template construction is smaller than the theoretic gain since the interclass covariance matrix is not much wider than the intraclass covariance matrix

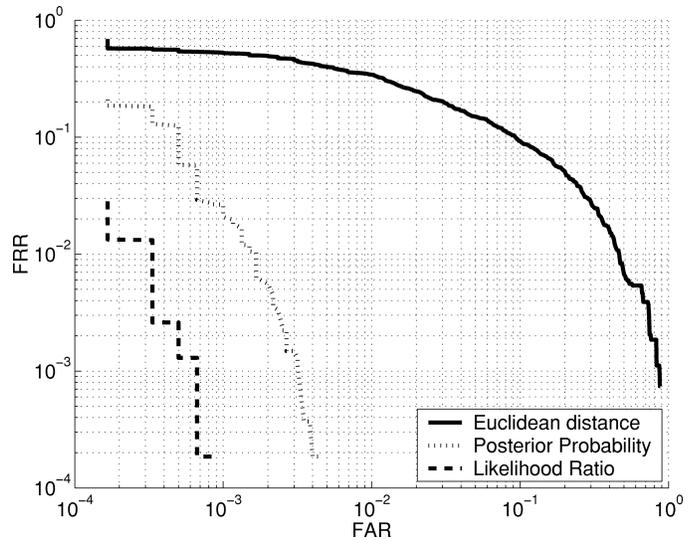


Fig. 7. Results of the matching algorithm that is based on both features and uses two fingerprints for template construction.

TABLE I
MATCHING PERFORMANCE USING MULTIPLE FINGERPRINTS
FOR TEMPLATE CONSTRUCTION

# templates	1	2	3	4	5
EER	0.5%	0.07%	0.03%	0.01%	0%

for most of the elements of the feature vector. Therefore, the assumption of a constant $p(\mathbf{v})$ is not true in practice, and the performance gain is smaller than predicted.

V. CONCLUSION

In this paper, we have shown that the verification of a single user is equivalent to the detection problem, which implies that, for single-user verification, the likelihood ratio is optimal. We have also shown that, in single-user verification, decisions based on posterior probability and likelihood ratio are equivalent, and result in the same ROC. However, in a multiuser situation, the two methods lead to different average error rates. As a third result, we have proven theoretically that, for multi-user verification, the use of the likelihood ratio is optimal in terms of average error rates.

The superiority of the likelihood-based similarity measure is illustrated by experiments in fingerprint verification. It is shown that error rates below 10^{-3} can be achieved when using multiple fingerprints for template construction. Since the algorithm is extremely fast, it can be used to search through large fingerprint databases. For automatic application of the algorithm, improvements have to be made in the automatic extraction of the core point. This could be circumvented by trying all detected cores, but that would slow down a database search.

Future research will address more reliable reference point detection and the reliability of high-dimensional covariance matrix estimates, with the purpose of minimizing the differences between estimates for the training set and the test set.

APPENDIX
GAUSSIAN APPROXIMATION OF ERROR RATES

Assume that the different observed feature vectors \mathbf{v} that are generated by class w , have a multi-dimensional Gaussian distribution with dimension d , mean \mathbf{c} and within-class covariance matrix Σ_W , independent of the class. Then, the conditional probability density $p(\mathbf{v}|w)$ is given by

$$\begin{aligned} p(\mathbf{v}|w) &= n(\mathbf{c}, \Sigma_W) \\ &= \frac{1}{(2\pi)^{d/2} \cdot |\Sigma_W|^{1/2}} \\ &\quad \cdot \exp\left(-\frac{1}{2}(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c})\right). \end{aligned} \quad (29)$$

Furthermore, assume that the centers \mathbf{c} of classes w have a Gaussian distribution with zero mean (which can be guaranteed by subtraction of the mean) and between-class covariance matrix Σ_B . Then, $p(\mathbf{c})$ is given by

$$\begin{aligned} p(\mathbf{c}) &= n(0, \Sigma_B) \\ &= \frac{1}{(2\pi)^{d/2} \cdot |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{c}^T \Sigma_B^{-1} \mathbf{c}\right). \end{aligned} \quad (31)$$

The prior probability density function $p(\mathbf{v})$ of all patterns is given by

$$\begin{aligned} p(\mathbf{v}) &= \int_W p(\mathbf{v}|w)p(w) dw \\ &= \int_{\mathbf{v}} n(\mathbf{c}, \Sigma_W) \cdot n(0, \Sigma_B) d\mathbf{c} \end{aligned} \quad (33)$$

which has a total covariance matrix $\Sigma_T = \Sigma_W + \Sigma_B$. Under the condition that the distribution over the entire population is much wider than the distribution within one class

$$|\Sigma_W| \ll |\Sigma_B| \quad (34)$$

then $p(\mathbf{v})$ is equal to $p(\mathbf{c})$, given by

$$\begin{aligned} p(\mathbf{v}) &= n(0, \Sigma_B) \\ &= \frac{1}{(2\pi)^{d/2} \cdot |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{v}^T \Sigma_B^{-1} \mathbf{v}\right). \end{aligned} \quad (35)$$

Using the acceptance condition $L(\mathbf{v}) \geq t$, and (5) for $L(\mathbf{v})$, an expression for the acceptance region $A_{t,w}$ can be derived as

$$p(\mathbf{v}|w) \geq t \cdot p(\mathbf{v}). \quad (37)$$

By substituting the Gaussian probability density functions (30) and (36), we obtain

$$\begin{aligned} &\frac{1}{(2\pi)^{d/2} \cdot |\Sigma_W|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c})\right) \\ &\geq t \cdot \frac{1}{(2\pi)^{d/2} \cdot |\Sigma_B|^{1/2}} \cdot \exp\left(-\frac{1}{2}\mathbf{v}^T \Sigma_B^{-1} \mathbf{v}\right) \end{aligned} \quad (38)$$

or

$$\begin{aligned} &\exp\left(-\frac{1}{2}(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c}) + \frac{1}{2}\mathbf{v}^T \Sigma_B^{-1} \mathbf{v}\right) \\ &\geq t \cdot \left(\frac{|\Sigma_W|}{|\Sigma_B|}\right)^{1/2}. \end{aligned} \quad (39)$$

By taking the natural logarithm and multiplying both sides by -2, we obtain

$$(\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c}) - \mathbf{v}^T \Sigma_B^{-1} \mathbf{v} \leq -2 \log\left(t^2 \cdot \frac{|\Sigma_W|}{|\Sigma_B|}\right) \quad (40)$$

and the acceptance region is given by

$$A_{t,w} = \{\mathbf{v} \in \mathbf{v} | (\mathbf{v} - \mathbf{c})^T \Sigma_W^{-1} (\mathbf{v} - \mathbf{c}) \leq \beta^2(t)\} \quad (41)$$

with

$$\beta^2(t) = -\log\left(t^2 \cdot \frac{|\Sigma_W|}{|\Sigma_B|}\right) + \mathbf{v}^T \Sigma_B^{-1} \mathbf{v}. \quad (42)$$

For threshold values that correspond to a relatively high likelihood ratio, \mathbf{v} is relatively close to \mathbf{c} , and \mathbf{v} can be replaced with \mathbf{c} in (42). Then, β is independent of \mathbf{v} , and $A_{t,w}$ is an ellipsoid region with a Mahalanobis distance less than β from the class center \mathbf{c} .

Now, the error rates can be calculated by substituting (41) into (9) and (11). To transform the multidimensional integrals into one-dimensional integrals, we use the volume $\text{Vol}(r, \Sigma)$ of a ellipsoid that is defined by covariance matrix Σ and Mahalanobis distance r , which is given by [4]

$$\text{Vol}(r, \Sigma) = V_{\text{unit}} \cdot |\Sigma|^{1/2} \cdot r^d \quad (43)$$

where V_{unit} is a constant that depends on the dimension d of the space

$$V_{\text{unit}} = \begin{cases} \pi^{1/2}, & d \text{ even} \\ \left(\frac{d}{2}\right)!, & \\ \frac{2^d \pi^{(d-1)/2} \frac{(d-1)}{2}}{d!}, & d \text{ odd.} \end{cases} \quad (44)$$

Using

$$d \text{ Vol}(r, \Sigma) = V_{\text{unit}} \cdot |\Sigma|^{1/2} \cdot d \cdot s^{d-1} ds \quad (45)$$

and (9), the false rejection rate is given by

$$\begin{aligned} \text{FRR}(t, w) &= 1 - \int_0^{\beta(t)} p(\mathbf{v}|w) d \text{Vol}(r, \Sigma) \\ &= 1 - \int_0^{\beta(t)} \frac{1}{(2\pi)^{d/2} |\Sigma_W|^{1/2}} \\ &\quad \cdot \exp\left(\frac{-s^2}{2}\right) V_{\text{unit}} |\Sigma_W|^{1/2} ds^{d-1} ds \\ &= 1 - \frac{\text{unit} \cdot d}{(2\pi)^{d/2}} \cdot \int_0^{\beta(t)} \exp\left(\frac{-s^2}{2}\right) s^{d-1} ds \end{aligned} \quad (46)$$

which corresponds to a χ^2 distribution with d degrees of freedom.

Using (11) and assuming that $p(\mathbf{v})$ is constant within class w , the false acceptance rate is given by

$$\text{FAR}(t, w) = p(\mathbf{v}) \cdot V_{\text{unit}} \cdot |\Sigma_W|^{1/2} \cdot \beta^d(t). \quad (47)$$

These expressions for the theoretic error rates can be calculated relatively easily. Experiments show that they provide a very accurate estimate of the error rates for low dimensions ($d < 10$), while the experimental results deviate from the predicted values for higher dimensions. This can be explained by the fact that $p(\mathbf{v})$ cannot be assumed to be constant within class w anymore.

REFERENCES

- [1] A. M. Bazen and S. H. Gerez, "Fingerprint matching by thin-plate spline modeling of elastic deformations," *Patt. Recognit.*, vol. 36, no. 8, pp. 1859–1867, Aug. 2003.
- [2] A. K. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 302–314, Apr. 1997.
- [3] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Trans. Image Processing*, vol. 9, pp. 846–859, May 2000.
- [4] M. Golfarelli, D. Maio, and D. Maltoni, "On the error-reject trade-off in biometric verification systems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 786–796, July 1997.
- [5] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.
- [6] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2000.
- [7] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2000: Fingerprint verification competition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 402–412, Mar. 2002.
- [8] A. M. Bazen and S. H. Gerez, "Systematic methods for the computation of the directional field and singular points of fingerprints," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 905–919, July 2002.
- [9] A. Ross, J. Reisman, and A. K. Jain, "Fingerprint matching using feature space correlation," in *Proc. Biometric Authentication, Int. ECCV 2002 Workshop*, vol. 2359 of LNCS, M. Tistarelli, J. Bigun, and A. K. Jain, Eds., Copenhagen, Denmark, June 2002, pp. 48–57.
- [10] A. M. Bazen, "Fingerprint identification—Feature extraction, matching, and database search," Ph.D. dissertation, Univ. of Twente, Enschede, The Netherlands, 2002.
- [11] A. Ross, A. K. Jain, and J. Reisman, "A hybrid fingerprint matcher," *Patt. Recognit.*, vol. 36, no. 7, pp. 1661–1673, July 2003.
- [12] S. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA: Academic, 1990.
- [13] A. K. Jain, S. Prabhakar, and A. Ross, "Fingerprint matching: Data Acquisition and Performance Evaluation," Dept. of Computer Science, Michigan State Univ., East Lansing, Tech. Rep. MSU-CPS-99-14, 1999.
- [14] A. K. Jain and A. Ross, "Fingerprint mosaicking," in *Proc. IEEE Int. Conf. on Acoustic Speech and Signal Processing (ICASSP)*, Orlando, FL, May 2002.



Asker M. Bazen was born in The Netherlands in 1973. He received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1998 and 2002, respectively. His master's research focused on high-resolution parametric radar processing, which he continued for one more year at Thomson-CSF Signaal. His doctoral work focused on various topics in fingerprint recognition, including robust minutiae extraction from low-quality fingerprints, matching elastically deformed fingerprints, and indexing large

fingerprint databases.

He currently has a research position at the University of Twente in the field of biometrics. His research interests include biometrics, signal and image processing, and statistical pattern recognition.



Raymond N. J. Veldhuis received the engineer 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. His dissertation was titled "Adaptive restoration of lost samples in discrete-time signals and digital images."

From 1982 until 1992, he worked as a Researcher at 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 until 2001 he worked at the IPO (Institute of Perception Research), Eindhoven, in speech signal processing and speech synthesis. From 1998 until 2001, he was program manager of the Spoken Language Interfaces research program. He is now an Associate Professor at the University of Twente, working in the fields of biometrics and signal processing. His expertise involves digital signal processing for audio, images and speech, statistical pattern recognition, and biometrics. He has been active in the development of MPEG standards for audio source coding.