A Likelihood Ratio Classifier for Histogram Features

Raymond Veldhuis
University of Twente
Eindhoven, The Netherlands
r.n.j.veldhuis@utwente.nl

Kiran Raja
University of South-Eastern Norway
Norwegian Biometrics Laboratory, NTNU - Norway
kiran.raja@usn.no

Raghavendra Ramachandra
Norwegian Biometrics Laboratory, NTNU - Norway
raghavendra.ramachandra@ntnu.no

Abstract

In a number of classification problems, the features are represented by histograms. Traditionally, histograms are compared by relatively simple distance measures such as the chi-square, the Kullback-Leibler, or the Euclidean distance. This paper proposes a likelihood ratio classifier for histogram features that is optimal in Neyman-Pearson sense. It is based on the assumptions that histograms can be modelled by a multinomial distribution and the bin probabilities of the histograms by a Dirichlet probability density. A simple method to estimate the Dirichlet parameters is included. Feature selection prior to classification improves the classification performance. Classification results are presented on periocular and face data from various datasets. It is shown that the proposed classifier outperforms the chi-square distance measure.

1. Introduction

In a number of biometric comparison methods, features are represented by histograms that indicate the frequencies of occurrence of local descriptors. Well-known examples of local descriptors that are aggregated in histograms are Local Binary Patterns (LBP) [1], Binarized Statistical Image Features (BSIF) [8], and Histograms of Gradients (HOGs) [4]. The comparison of histograms that count occurrences of words or of n-grams is an elementary method for document classification since the introduction of the bag-of-words concept by [17]. Traditionally, histograms are compared by relatively simple measures such as the chi-square (e.g. in [1]), the Kullback-Leibler [9], or the Euclidean distance, although in [4] a more complex Support Vector Machine (SVM) is proposed as a classifier. An overview of local descriptors and similarity measures that are used to compare histograms is presented in [10]. Apart from the SVM, the distance measures for histograms reviewed in [10] are simple and robust, because they do not require any training, but they may not provide the best recognition performance.

The objective of this paper is to introduce a Likelihood Ratio (LR) classifier for histogram features and to demonstrate its potential. Likelihood ratio classifiers take the likelihood ratio as a similarity score and compare it to a threshold. They are optimal in Neyman-Pearson sense [19]. In a biometric context this means that at any given False-Match Rate (FMR), they will achieve the maximum True-Match Rate (TMR) or Genuine Match Rate (GMR).

Likelihood ratio classifiers can be derived for features with known probability densities (PDFs). In practice this is a limitation, because these are usually unknown. In some cases, parametric densities can be estimated from training data, see, for instance, [2, 7, 13] for normal PDFs. The method proposed here is based on the assumption that the bin probabilities of the histograms can be modelled by a Dirichlet probability density [3]. The parameters of this probability density serve as a background model that models – in analogy to face space – histogram space. Further, they must be estimated from a training set.

Two variants of likelihood ratio classifiers for histograms are presented in this work. The first, 1:1 comparison, compares 2 histograms to determine whether they have the same or different bin probabilities. In biometric recognition, one may assume that 2 histograms measured from the same individual will have the same bin probabilities and that histograms measured from different individuals will have different bin probabilities. This method lends itself well for multiple-enrolment or multiple-probe applications, since, the respective enrolment or probe histograms can simply be added bin-by-bin. The second variant is user-specific comparison, as it compares a single histogram with the bin probabilities of a specific user. User-specific classifiers are
useful for application in personal devices such as smartphones, where the user can enrol a number of biometric samples. Feature selection prior to classification will improve the classification performance of both variants.

Although the presented work and the results can be adopted to areas of computer vision and document classification employing histogram features, we restrict and discuss the results in context of biometric recognition alone in this work. In what follows, we will derive the likelihood ratio classifier for 1:1 comparison in Section 2, and the one for user-specific comparison in Section 3. Section 4 proposes a method to estimate the Dirichlet parameters as a background model from a training set. This method is based on the Nelder-Mead simplex downhill method [12]. Feature selection is discussed in Section 5. Section 6 presents the results of experiments on face data from FRGCv2 dataset [14] and periocular data from [16]. It is shown that the proposed classifiers outperform the chi-square distance measure. Finally, Section 7 presents conclusions.

2. A LR classifier for comparing histograms

Let vectors $x \in \mathbb{N}^n$ and $y \in \mathbb{N}^n$ denote 2 histograms, e.g. counting the occurrences of LBP or BSIF descriptors in facial images, with $\sum_{i=1}^n x_i = X$ and $\sum_{i=1}^n y_i = Y$, then the problem at hand is to decide whether $x$ and $y$ are obtained from the same source. I.e., in the face recognition example, from the same individual. We assume that a probability $p_i \geq 0$, such that $\sum_{i=1}^n p_i = 1$ is associated to every bin $i$, $i = 1, \ldots, n$ of a histogram, conveniently arranged into a vector $p$. Furthermore, we assume that a histogram $x$ is a realisation of a random vector $\mathbf{x}$. In this paper, random vectors and variables will be underlined, in contrast to their realisations, which will not be.

A histogram is filled by distributing the outcomes of observations – values of local descriptors in face recognition – over the bins of the histogram. We assume that the allocation to bins of these outcomes are statistically independent. Under that assumption, the probability of a realisation $x$ of $\mathbf{x}$ is the multinomial distribution with parameters $p$, given by

$$P\{\mathbf{x} = x|p\} = \text{Mult}(x|p) = \frac{X!}{\prod_{i=1}^n x_i!} \prod_{i=1}^n p_i^{x_i},$$

(1)

For histograms that are biometric features, it is plausible to assume that each individual is characterised by its own vector of probabilities $p$, which can then be taken as a realisation of a random vector $p$. The classification problem at hand then translates to the question whether two histograms $x$ and $y$ share the same probability vector $p$ or are characterised by two probability vectors $p$ and $q$, respectively.

In general biometric comparison, given two feature vectors $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$, we can define a likelihood ratio classifier that compares the likelihood ratio $lr(u, v)$ to a predefined threshold in order to determine whether $u$ and $v$ originate from the same or from different individuals. This likelihood ratio is defined as

$$lr(u, v) \overset{\text{def}}{=} \frac{f_{u,v}(u, v|S)}{f_{u,v}(u, v|D)},$$

(2)

where $f_z(z)$ denotes the probability density function of a random variable $z$ evaluated in $z$, and $S$ and $D$ denote the conditions that $u$ and $v$ originate from the same or from different individuals, respectively [7, 13]. For the comparison of histograms this translates to

$$lr(x, y) = \frac{P\{x = x, y = y|S\}}{P\{x = x\}P\{y = y\}},$$

(3)

$S$ and $D$ denote the conditions that $x$ and $y$ share the same probability vector $p$ or are characterised by two different probability vectors, respectively. When $x$ and $y$ result from different individuals they are statistically independent, hence (3) can be written as

$$lr(x, y) = \frac{P\{x = x, y = y|S\}}{P\{x = x\}P\{y = y\}},$$

(4)

Under both conditions $S$ and $D$ are parameters for the multinomial distributions that are unknown, but we may have an idea – possibly based on measurements – about their range of values that we could model probabilistically by means of a prior probability density $f_x(p)$. With that, and using the conditional independence of $x$ and $y$ given $p$, we can rewrite (4) as

$$lr(x, y) = \frac{\int_p \text{Mult}(x|p)\text{Mult}(y|p)f_x(p)dp}{\int_p \text{Mult}(x|p)f_x(p)dp\int_q \text{Mult}(y|q)f_y(q)dq}.$$

(5)

The choice of the prior probability densities is important, as it expresses what we know about the underlying probabilities. In this context the Dirichlet density is relevant for two reasons. The first is that it is sufficiently flexible to model a wide range of underlying bin probabilities. The second is that it is the conjugate prior density of the multinomial distribution [3]. This means that when we observe a probability vector $p$ that has a prior Dirichlet probability density through a counting experiment that follows a multinomial distribution with parameter vector $p$, we can update the prior distribution to a posterior distribution of $p$ that is again a Dirichlet posterior density. As a side-effect the integrals in (5) become tractable. For $p_i \geq 0, i = 1, \ldots, n$, $\sum_{i=1}^n p_i = 1$, and parameter vector $\alpha \in \mathbb{R}^n$, with $\alpha_i > 0, i = 1, \ldots, n$, the Dirichlet density is defined as

$$\text{Dir}(p|\alpha) \overset{\text{def}}{=} \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n p_i^{\alpha_i - 1},$$

(6)

$$= \frac{1}{\text{B}(\alpha)} \prod_{i=1}^n p_i^{\alpha_i - 1}$$

(7)
Here $\Gamma(\alpha)$ is the gamma function [5]. In (7) the quotient of gamma functions is replaced by the multivariate beta function $B(\alpha)$ [5]. This is done to keep notations and derivations as simple as possible. Note that for $\alpha_i = 1$, $i = 1, \ldots, n$, we have that $\text{Dir}(\mathbf{p}|\alpha)$ is uniform on the simplex $p_i \geq 0$, $i = 1, \ldots, n$, $\sum_{i=1}^n p_i = 1$, and can serve as an uninform prior. By choosing appropriate $\alpha_i$, a prior probability density on $\mathbf{p}$ can be approximated.

With the definition of the multinomial distribution (1) and the Dirichlet density (7), the equation (5) can be worked out. We will start with the factors in the denominator, assuming that $\mathbf{p} \sim \text{Dir}(\mathbf{p}|\alpha)$ and $\mathbf{q} \sim \text{Dir}(\mathbf{q}|\alpha)$ for some $\alpha$ that can be estimated from a training set. We write

\[
\int_{\mathbf{p}} \text{Mult}(\mathbf{x} | \mathbf{p}) f_{\mathbf{p}}(\mathbf{p}) d\mathbf{p} = \int_{\mathbf{p}} \text{Mult}(\mathbf{x} | \mathbf{p}) \text{Dir}(\mathbf{p}|\alpha) d\mathbf{p} = \frac{X!}{\prod_{i=1}^n x_i!} \prod_{i=1}^n \frac{p_i^{x_i}}{B(\alpha)} \prod_{i=1}^n \frac{1}{p_i^{\alpha_i - 1}} d\mathbf{p} = \frac{X!}{\prod_{i=1}^n x_i!} \frac{1}{B(\alpha)} \int_{\mathbf{p}} \prod_{i=1}^n p_i^{x_i + \alpha_i - 1} d\mathbf{p} = \frac{X!}{\prod_{i=1}^n x_i!} \frac{B(\mathbf{x} + \alpha)}{B(\alpha)} \prod_{i=1}^n \frac{1}{p_i^{\alpha_i}} d\mathbf{p} = \frac{X!}{\prod_{i=1}^n x_i!} \frac{B(\mathbf{x} + \alpha)}{B(\alpha)} .
\] (8)

In the last step, we used that the integral over a probability density function amounts to 1. For the second factor in denominator of (5), we obtain

\[
\int_{\mathbf{p}} \text{Mult}(\mathbf{y} | \mathbf{p}) f_{\mathbf{p}}(\mathbf{p}) d\mathbf{p} = \frac{Y!}{\prod_{i=1}^n y_i!} \frac{B(\mathbf{y} + \alpha)}{B(\alpha)} .
\] (9)

For the numerator of (5), we can derive in a similar way that

\[
\int_{\mathbf{p}} \text{Mult}(\mathbf{x} | \mathbf{p}) \text{Mult}(\mathbf{y} | \mathbf{p}) f_{\mathbf{p}}(\mathbf{p}) d\mathbf{p} = \frac{X!}{\prod_{i=1}^n x_i!} \frac{Y!}{\prod_{i=1}^n y_i!} \frac{B(\mathbf{x} + \mathbf{y} + \alpha)}{B(\alpha)} .
\] (10)

On combination of (8), (9), and (10) we obtain for (5)

\[
lr(\mathbf{x}, \mathbf{y}|\alpha) = B(\alpha) \frac{B(\mathbf{x} + \mathbf{y} + \alpha)}{B(\mathbf{x} + \alpha) B(\mathbf{y} + \alpha)} .
\] (11)

This is an elegant expression that is symmetric in $\mathbf{x}$ and $\mathbf{y}$. We included the $\alpha$ in $\text{lr}(\mathbf{x}, \mathbf{y}|\alpha)$ to emphasise that it depends on background model parameters. How these parameters can be estimated from training data will be explained in Section 4. An interesting property of $\text{lr}(\mathbf{x}, \mathbf{y}|\alpha)$ is that $\text{lr}(\mathbf{x}, \mathbf{y}|\alpha) = 1$ if either $\mathbf{x} = 0$ or $\mathbf{y} = 0$. A likelihood ratio of 1 indicates that the comparison does not provide evidence in either direction. This is exactly the case if no measurements for $\mathbf{x}$ or $\mathbf{y}$ are available.

The likelihood ratio in (11) may be hard to compute directly because of the high values that may occur when evaluating the gamma functions, which are generalisation of factorials, that constitute the beta functions. However, this problem can be evaded by computing the Log-Likelihood Ratio (LLR)

\[
\text{llr}(\mathbf{x}, \mathbf{y}|\alpha) = \begin{cases} \log (B(\alpha) + \log (B(\mathbf{x} + \mathbf{y} + \alpha)) \quad & \text{if } \mathbf{x} + \mathbf{y} + \alpha > 0 \\ 0 & \text{otherwise} \end{cases} ,
\] (12)

which is fully equivalent to the likelihood ratio. This simplifies the computations because in (12) the log-beta functions can be expanded as sums and differences of log-gamma functions, which can be computed efficiently without high values using the factorial property of the gamma function [5]. In the log domain this is given by $\log(\Gamma(m + r)) = \log(\Gamma(r)) + \sum_{k=1}^m \log(\Gamma(k + r))$, with $m \geq 1$, $m \in \mathbb{N}$ and $r \in [0, 1)$.

The likelihood ratio (11) or the log-likelihood ratio (12) can be generalised to multiple-enrolment and multiple-probe cases. Assume that $m_e \geq 1$ enrolment histograms $\{y_i\}_{i=1}^{m_e} \geq 1$ and $m_p$ probe histograms $\{x_i\}_{i=1}^{m_p}$ are available, then the corresponding likelihood ratio is given by

\[
lr(\{x_i\}_{i=1}^{m_p}, \{y_i\}_{i=1}^{m_e}|\alpha) = \frac{B(\alpha) B\left(\sum_{i=1}^{m_p} x_i + \sum_{i=1}^{m_e} y_i + \alpha\right)}{B\left(\sum_{i=1}^{m_p} x_i + \alpha\right) B\left(\sum_{i=1}^{m_e} y_i + \alpha\right)} ,
\] (13)

and a similar result can be given for the log-likelihood ratio in (12).

3. The user-specific case

If the likelihood ratio classifier is to recognise one specific individual, the results (11) and (12) simplify. It is then designed to test whether the histogram $\mathbf{x}$ originates from a specific individual with specific bin probabilities $\mathbf{q}$ or from an arbitrary individual with bin probabilities drawn from the background model. For this case the likelihood ratio is given by

\[
lr(\mathbf{x}|\alpha, \mathbf{q}) = \frac{\text{Mult}(\mathbf{x}|\mathbf{q})}{\int_{\mathbf{p}} \text{Mult}(\mathbf{x} | \mathbf{p}) \text{Dir}(\mathbf{p}|\alpha) d\mathbf{p}} .
\] (14)

By using (1) and (8) we obtain

\[
lr(\mathbf{x}|\alpha, \mathbf{q}) = \frac{B(\alpha)}{B(\mathbf{x} + \alpha)} \prod_{i=1}^n q_i^{x_i} .
\] (15)

In order to avoid large numbers in the evaluation of (15), computing the log likelihood ratio is preferred, which is
given by
\[
\text{llr}(x|\alpha, q) = \log (B(\alpha)) - \log (B(x + \alpha)) + \sum_{i=1}^{n} x_i \log (q_i).
\]

The last equation provides use with an initial estimate \(\hat{\alpha}_{\text{init}}\) for \(\alpha\). All terms in (22) may serve as estimator for \(\alpha_0\). They are averaged to increase accuracy.

The bin probabilities that \(\hat{\mu}_i\) and \(\hat{\sigma}^2_i\) are estimated from must be estimated from sets of histograms of subjects in a training set. For that it is required that multiple histograms are available per subject. Let \(x_{j,k}, k = 1, \ldots, K_j\) denote the histograms that are available of subject \(j, j = 1, \ldots, J\), and \(X_{j,k}\) the sum of the elements of \(x_{j,k}\). Then \(p_j\) is estimated as
\[
\hat{p}_j = \frac{\sum_{k=1}^{K_j} x_{j,k} + 1}{\sum_{k=1}^{K_j} X_{j,k} + n},
\]
with 1 the all-ones vector and \(n\) the number of bins in the histograms. This estimator is the expectation of a posteriori Dirichlet probability density of \(p_j\), given the histograms \(x_{j,k}, k = 1, \ldots, K_j\), assuming an uninformative Dirichlet prior PDF with parameters \(\alpha = 1\). The ones in the numerator keep the initial estimates of the \(\alpha_i\) strictly positive, as is required for Dirichlet parameters. The \(\hat{p}_j\) are used to compute \(\hat{\mu}_i\) and \(\hat{\sigma}^2_i\), which are used in (22) and (23). They can also serve as the user-specific parameters \(q\) in (15) and (16).

Figure 1 illustrates how the proposed method for background model estimation converges. The figure shows the results of a synthetic data experiment. For 400 users, bin probabilities \(p_j, j = 1, \ldots, 400\), for histograms with \(n = 50\) were drawn from a Dirichlet probability density with \(\alpha = 0.7843 (1, \ldots, n)^\top\). For each user \(j, j = 1, \ldots, 400\), 10 histograms were then drawn from a multinomial distribution with parameters \(p_j\). From these histograms initial estimates \(\hat{\alpha}_{\text{init}}\) were computed as given by (22) and (23), which were used to compute estimates \(\hat{\alpha}\) for the Dirichlet parameters given by (19) using Nelder-Mead minimisation.

In Figure 1 the circles represent the true values of \(\alpha_i\), the stars the initial estimated \(\hat{\alpha}_{\text{init},i}\), and the triangles the final estimates \(\hat{\alpha}_i\). The differences between initial estimate and final estimate are minimal.

5. Feature selection

In order to prevent over-fitting, i.e. tuning model parameters too much to the peculiarities of a training set, such that the performance on testing data decreases, parameter or feature reduction is applied in many trained classifiers, e.g. [6]. We found that in the present case feature reduction, removing bins from the histograms can also improve the recognition results. Empirically we found that removing bins with the highest global bin probabilities measured over the entire training set improves the recognition performance the most. Figure 2 illustrates this for BSIF histograms with \(n = 256\) [8] obtained from the left periocular region. The feature extraction is described in [16]. The experimental settings are given in Section 6. The graph shows the TMR at FMR=0.001 for the 1:1 likelihood ratio classifier (solid...
that variabilities in the data are not well modeled. Another possible explanation would be that bins with high global probabilities may be less discriminative as they may occur in all subjects, but in that case the TMR of the chi-square classifier should also attain a maximum at a number of bins below 256.

6. Experimental results

The goal of the experiments presented here is to illustrate the potential of the proposed likelihood ratio classifiers for histogram features. For that reason we will compare them to variants of the chi-square classifier. A more extensive comparison such as the one in [10] will be part of future work. The choice for comparing with chi-square is motivated by the fact that in the extensive comparisons presented in [10] the chi-square distance often comes out as the best classifier. A comparison with SVM is not included here, because in biometrics SVM is commonly used in a one-versus-all scenario. This requires a specific SVM for each subject in the enrolment set, which is unsuitable for the 1:1 comparison scheme that is tested here, but could have been included in the user-specific case.

Two recognition scenarios are tested: 1:1 comparison and user-specific comparison. In 1:1 comparison, \( \text{llr}(x, y; \hat{\alpha}) \) from (12), with \( \hat{\alpha} \) obtained from training, is used to compare two histograms \( x \) and \( y \) to decide whether or not they originate from the same individual. The chi-square distance that we use to compare the proposed method with in the 1:1 scenario is given by

\[
d_{\chi^2}(x, y) = 2 \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{x_i + y_i}.
\]

In user-specific comparison it is decided whether a histogram results from a specific or from an arbitrary individual. In the user-specific scenario the multiple enrolment variant \( \text{llr}(x, \sum_{m_e=1}^{m_e} y_i; \hat{\alpha}) \) in (13) of (12) is tested with \( m_p = 1 \) and a certain \( m_e \), as well as the user-specific classifier \( \text{llr}(x; \hat{\alpha}, \hat{q}) \) from (16), with \( \hat{q} \) computed from the \( m_e \) enrolment histograms \( y_i \) of a specific subject as in (24). As in the 1:1 comparison, \( \hat{\alpha} \) is obtained from training. The chi-square distance that we use to compare the proposed method with in the user-specific scenario is given by

\[
d_{\chi^2}(x; \hat{q}) = \sum_{i=1}^{n} \frac{(x_i - n\hat{q}_i)^2}{n\hat{q}_i}.
\]

Results are presented as Equal-Error Rates (EERs), TMRs at an FMR of 0.0001, and receiver operating characteristics (ROCs), plotting the TMR as a function of the FMR. We present the results of experiments on two datasets: the Smartphone Biometric Dataset used in [16] and on the FRGCv2 dataset [14].
Smartphone Biometric Dataset  The dataset collected in [16] contains images of the face and of the left and right periocular regions from 73 users taken by smartphone camera. The size of face region downsampled to 256 × 256 pixels. The size of the periocular region of one eye is 120 × 88 (w × h) pixels. 15 samples of each user from different sessions are available. BSIF features were extracted from patches of 9 × 9 pixels in 8 layers as described in [16], resulting in histograms of length \( n = 256 \).

For the face data, we used the images of 50 users for training the background model \( \alpha \) and of the remaining 23 for testing. In order to extend the number of features for training and testing for periocular data, we mirrored the images and used the mirrored right periocular regions as left periocular images from an additional 73 users. In this way we created datasets with periocular images from 146 users. The first 100 disjoint subjects in the dataset were used for training the background model \( \alpha \). The remaining 46 for testing. Prior to training, we reduced the feature size to 180 by discarding the 76 bins with the highest global bin probabilities in the training set.

For the 1:1 comparison, we produced for each classifier 2415 mated (genuine) and 56925 non-mated (impostor) comparison scores for the facial data and 7455 mated comparison scores and 559125 non-mated comparison scores for the periocular data. In the user-specific comparison \( m_e = 10 \) enrolment histograms were randomly chosen from the 15 samples that were available from all users. This was repeated 15 times. Thus we produced for each classifier 1725 mated and 113850 non-mated comparison scores for the facial data and 5325 mated comparison scores and 1118250 non-mated comparison scores for the periocular data.

The results are presented in Table 1. ROC curves, plotting TMR as a function of the FMR are shown in Figure 3 for the left periocular data. The solid blue line in this figure shows the ROC obtained with \( \text{llr}(x, y | \hat{\alpha}) \). The dash-dotted blue line is obtained with the chi-square distance in (25). The solid red line in this figure shows the ROC obtained with \( \text{llr}(x, \sum_{i=1}^{m_e} y_i | \hat{\alpha}) \). The dashed red line is obtained with \( \text{llr}(x | \hat{\alpha}, \hat{q}) \). The dash-dotted red line is obtained with the chi-square distance in (26).

We observe that the proposed methods outperform the chi-square distances in terms of EERs and that at FMR=0.0001 the TMRs of the proposed methods are similar to those obtained with the chi-square distances, except for 1:1 comparison on the right periocular region, where chi-square is much better. The ROCs of the left periocular region show that for a wide range of FMR the proposed method outperforms chi-square. Both proposed user-specific methods show almost identical ROCs and subject-specific comparison is in all cases better than 1:1 comparison, as can be expected.

FRGCv2 dataset  We employ the FRGCv2 dataset [14] corresponding to the protocol known as Experiment-1. The dataset consists of 222 users in a training set and 466 users in a testing set. In order to test the robustness of the method, we have only considered the disjoint set of users who are only present in the testing set of 466 users. Furthermore, we employ 192 users from the testing set to derive the background model and the rest of the 274 disjoint users are employed for biometric performance evaluation. Thus, the background training data consists of 8016 images from 192 users. The remaining 8012 images are used for evaluation of the proposed approach. Prior to testing and training, each facial image in the dataset is cropped to 140 × 120 (w × h) pixels and rotation correction is applied based on the eye coordinates provided along with the dataset. BSIF features were extracted from patches of 17 × 17 pixels in 8 layers, resulting in histograms of length \( n = 256 \).

For the 1:1 comparison, we produced 174114 mated comparisons and 31917952 non-mated comparisons. In the user-specific comparison \( m_e = 5 \) enrolment histograms were randomly chosen from the samples that were available from all users. This was repeated 15 times. Thus we produced 68400 mated comparison scores and 21315600 non-mated comparison scores for each classifier.

The results from the experiments on FRGC dataset are provided in Table 2 and ROCs are provided in Figure 4. The solid blue line in this figure shows the ROC obtained with \( \text{llr}(x, y | \hat{\alpha}) \). The dash-dotted blue line is obtained with the chi-square distance in (25). The solid red line in this figure shows the ROC obtained with \( \text{llr}(x, \sum_{i=1}^{m_e} y_i | \hat{\alpha}) \). The dashed red is obtained with \( \text{llr}(x | \hat{\alpha}, \hat{q}) \). The dash-dotted red line is obtained with the chi-square distance in (26).
Table 1. Results obtained on Smartphone Biometric Dataset.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>EER × 10⁻² at FMR=0.0001</th>
<th>TMR × 10⁻² at FMR=0.0001</th>
<th>EER × 10⁻² at FMR=0.0001</th>
<th>TMR × 10⁻² at FMR=0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1 LLR</td>
<td>0.25</td>
<td>97.51</td>
<td>1.81</td>
<td>91.96</td>
</tr>
<tr>
<td>1:1 $\chi^2$</td>
<td>0.49</td>
<td>98.55</td>
<td>2.29</td>
<td>91.77</td>
</tr>
<tr>
<td>Multiple Enrolment LLR</td>
<td>0.02</td>
<td>99.88</td>
<td>0.72</td>
<td>98.51</td>
</tr>
<tr>
<td>Subject Specific LLR</td>
<td>0.01</td>
<td>99.88</td>
<td>0.77</td>
<td>98.51</td>
</tr>
<tr>
<td>Subject Specific $\chi^2$</td>
<td>0.11</td>
<td>99.76</td>
<td>0.99</td>
<td>97.53</td>
</tr>
</tbody>
</table>

Table 2. Results obtained on the FRGCv2 dataset.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>EER × 10⁻² at FMR=0.0001</th>
<th>TMR × 10⁻²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1 LLR</td>
<td>12.69</td>
<td>34.82</td>
</tr>
<tr>
<td>1:1 $\chi^2$</td>
<td>15.26</td>
<td>32.81</td>
</tr>
<tr>
<td>Multiple Enrolment LLR</td>
<td>3.59</td>
<td>73.39</td>
</tr>
<tr>
<td>Subject Specific LLR</td>
<td>3.55</td>
<td>73.61</td>
</tr>
<tr>
<td>Subject Specific $\chi^2$</td>
<td>4.93</td>
<td>65.04</td>
</tr>
</tbody>
</table>

We observe that the proposed methods outperform the chi-square distances in terms of EERs and in terms of TMRs at FMR=0.0001. The ROCs show that for a wide range of FMR, the proposed methods outperform comparison based on the chi-square distance. Both proposed user-specific methods show almost identical ROCs and subject-specific comparison is in all cases better than 1:1 comparison, as can be expected.

The recognition performance on the FRGC data is state-of-the-art [15]. This is not surprising, because the proposed method based on BSIF features is actually simple and not designed for this task. However, the purpose of the experiment was to demonstrate that the proposed classifiers are promising and that they seem to be because they outperform other classifiers on the same features. A subdivision into facial regions as is done in [1] for LBP features, each producing a histogram, may bring some improvement, but this is a topic for further investigations.

7. Conclusion

We derived likelihood ratio based classifiers for histogram features. The derivations were based on the assumptions that histograms can be modeled as draws from a multinomial distribution and that the bin probabilities of the histograms, i.e. the parameters of the multinomial distribution, can be modeled as draws from a Dirichlet probability density. A method for estimating the parameters of the Dirichlet probability density from training data was presented. It was found that feature selection by discarding bins that have a high global bin probability improves the recognition performance. Although possible explanations for this phenomenon were given, it still requires further study.

The classifiers were derived for 1:1 comparison, as can, for example, be found in automated border control applications that do not allow for the design of a dedicated classifier for every user, and for user-specific comparison, as we can find in authentication in personal devices that do allow for user-specific training. They were tested on histograms of BSIF features, extracted from facial and periocular images from smartphones and from the FRGCv2 dataset and compared to classifiers based on the chi-square distance, which are commonly used for histogram comparison. The results are promising and the proposed methods outperform the classifiers based on the chi-square distance for a wide range of FMRs.

The current work has not yet extensively studied the impact of the classifier on other protocols to gain more insights on the scalability and adoption of LR classifier. Furthermore, the current work has also not conducted a detailed experimental evaluation to determine the behavior of LR classifier with respect to various textural descriptors. Finally, comparing the proposed method with other (trained) classifiers, including SVM, for instance, is needed. Together with the elaboration on feature selection, this shall be considered in the future work.
Acknowledgement

This work is carried out under the partial funding of the Research Council of Norway under Grant No. IKTPLUSS 248030/O70.

References


