A Comparison of Hand-Geometry Recognition Methods Based on Low- and High-Level Features

Raymond Veldhuis, Asker Bazen, Wim Booij, Anne Hendrikse
University of Twente, Department of Electrical Engineering
P.O. box 217, 7500 AE Enschede, The Netherlands
Phone: +31 (0)53 489 2838 Fax +31 (0)53 489 1060
E-mail: R.N.J.Veldhuis@utwente.nl

Abstract—**This paper compares the performance of hand-geometry recognition based on high-level features and on low-level features. The difference between high- and low-level features is that the former are based on interpreting the biometric data, e.g. by locating a finger and measuring its dimensions, whereas the latter are not. The low-level features used here are landmarks on the contour of the hand. The high-level features are a standard set of geometrical features such as widths and lengths of fingers and angles, measured at preselected locations.**

Keywords— **Biometric verification, hand geometry, hand contour, landmarks.**

I. Introduction

Reported systems for hand-geometry recognition, e.g. [1], [2], [3], and [4] for an overview, use high-level geometrical features as inputs. Examples of such features are the widths and the lengths of fingers and of parts of the palm, and the angles between line segments connecting certain points. These features are measured from a black-and-white or gray-level image of the hand. The image processing needed to obtain these features is simple. For instance, the widths of the hand and the fingers, are measured on fixed line segments. Other features may require some simple type of processing, such as determining the fingertips and the interfinger points. Registration can be omitted, because the position of the hand is fixated by means of alignment pegs. A black-and-white image of the hand, including a side view, is shown in Figure 1. The lengths of the line segments and the angles are the features. The alignment pegs appear as black disks. The three larger black disks are for calibration.

The performance of hand-geometry recognition is, in spite of its simplicity, quite acceptable. Equal-error rates of about 0.5% have been reported[1], which is almost as good as the equal-error rate that is obtained by state-of-the art, but much more advanced, fingerprint recognition.

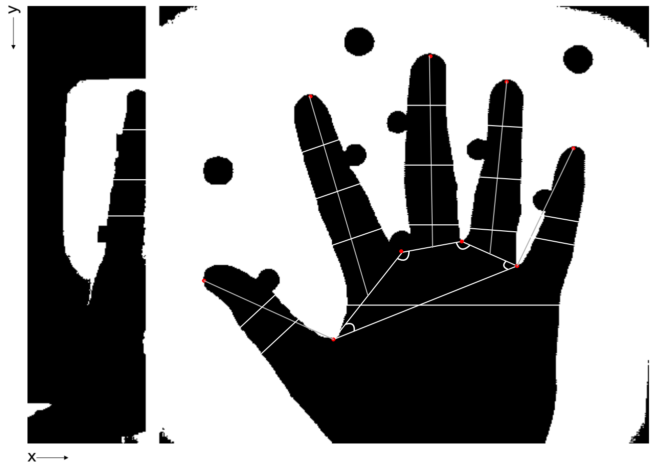

Fig. 1. Binary image of the hand and geometrical features. The lengths of the line segments and the angles in the image are used as features.

This paper demonstrates the feasibility of a method of contour-based hand-geometry recognition. The contour is completely determined by the black-and-white image of the hand and can be derived from it by means of simple image-processing techniques. It can be modelled by parameters, or features, that capture more details of the shape of the hand than the standard geometrical features do. An example of such a set of parameters are Fourier descriptors [5]. The features considered in this paper are the spatial coordinates of certain landmarks on the contour, similar to the landmarks used in the statistical shape model of the hand in [6]. These features are low-level, because they are not based on an interpretation of the data. Section II discusses the features and the recognition method.

Another contour-based method of hand-geometry recognition was published in [7]. This method does not use landmarks, but the fingers are extracted from the contour and aligned pairwise. The mean alignment error is used to compare contours. An equal-error rate of about

[1] In [1] an equal-error rate of 0.12% is claimed, but it is said that 'at the cross-over point we observed 1 FR and 118 FA'. In [1] the number of tests was 800 for the false-reject rate and 9900 for the false-accept rate. Therefore, the equal-error rate may be anywhere between 0.12% and 1.2%.

2.5% was reported.

The usefulness of the new method has been evaluated experimentally in a verification context. The verification performance obtained with contour-based features has been compared with the verification performance obtained with standard high-level features. The experiment and the results are presented in Section III.

II. Contour-based recognition

Images of the right hand are used for recognition. The part of the contour that is used runs counterclockwise from a point at a fixed distance below the basis of the little finger to a point at a fixed distance below the basis of the thumb. The parts of the contour below those points are not used, because they are unreliable due to sleeves or cuffs that may appear in the image. The alignment pegs are removed from the extracted contour. Possible dents at their locations are smoothed by linear interpolation. The number of landmarks on a contour can be chosen freely, but the minimum set consists of 11 reference landmarks. These are: the start and end point of the contour, the fingertips and the interfinger points. A number of $n_l \geq 0$ additional landmarks can be placed on the contour at equidistant positions between adjacent reference landmarks. This means that there are $l = 10n_l + 11$ landmarks in total. Their spatial coordinates (x, y) constitute the feature vector. The dimensionality m of the feature vector is, therefore, twice the number of landmarks. The landmarks can be aligned to a reference set by means of a rotation and a translation, but this hardly improves the recognition performance.

The verification is based on a log-likelihood-ratio classifier. It is assumed that the feature vectors have multivariate Gaussian probability densities. The total probability density, i.e. the probability density of a feature vector x without prior knowledge of the specific class of x, is

$$p(x) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_{\mathrm{T}}|^{\frac{1}{2}}} \mathrm{e}^{-\frac{(x-\mu_{\mathrm{T}})^{\mathrm{T}} \Sigma_{\mathrm{T}}^{-1} (x-\mu_{\mathrm{T}})}{2}}, \qquad (1)$$

with m the dimensionality of the feature space, μ_{T} the total mean and Σ_{T} the total covariance matrix. The superscript T denotes vector or matrix transposition. It is assumed that a class c is characterized by its class mean μ_c and that all classes have the same within-class covariance matrix Σ_{W}. The within-class probability density, i.e. the probability density of a feature vector $x \in c$, is

$$p(x|c) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_{\mathrm{W}}|^{\frac{1}{2}}} \mathrm{e}^{-\frac{(x-\mu_c)^{\mathrm{T}} \Sigma_{\mathrm{W}}^{-1} (x-\mu_c)}{2}}. \qquad (2)$$

Prior to classification the feature vector is mapped onto a lower-dimensional subspace by means of a linear transform. The $d \times m$ transform matrix M simultaneously diagonalizes the within-class and the total covariance matrix, such that the latter is an identity matrix. This results in a log-likelihood-ratio classifier that has a computational complexity that is linear, rather than quadratic, with the dimensionality d. The log-likelihood-ratio is then given by

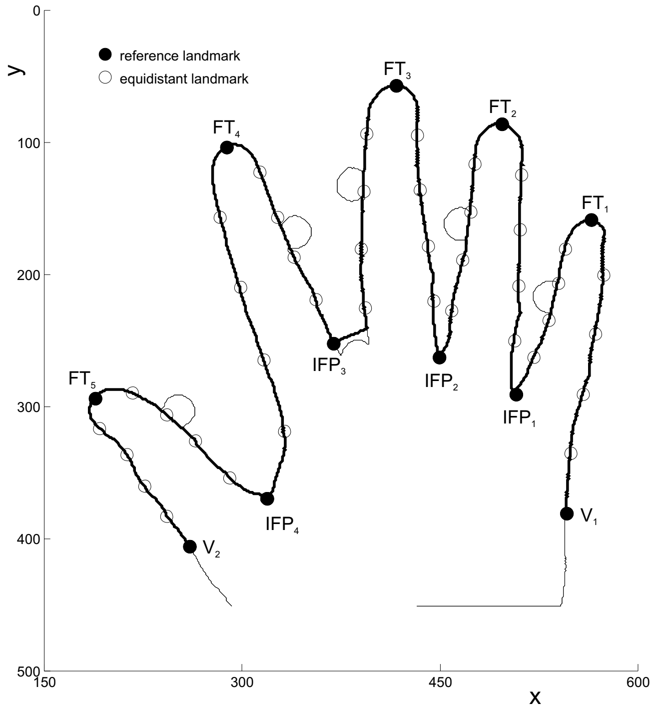

Fig. 2. Original contour (*thin*) and final contour (*thick*) with 51 landmarks $n_l = 4$.

$$\begin{aligned} l(y) = & \qquad (3) \\ & -\frac{1}{2}(y-\nu_c)^{\mathrm{T}} \Lambda^{-1} (y-\nu_c) + \frac{1}{2}(y-\nu_{\mathrm{T}})^{\mathrm{T}} (y-\nu_{\mathrm{T}}) \\ & -\frac{1}{2}\log(|\Lambda|), \end{aligned}$$

with $y = Mx$, $\nu_c = M\mu_c$, $\nu_{\mathrm{T}} = M\mu_{\mathrm{T}}$, and $\Lambda = M^{\mathrm{T}}\Sigma_{\mathrm{W}}M$ a diagonal matrix. If $l(y)$ is above a threshold T, the user is accepted, otherwise he is rejected.

The coefficients of the transformation matrix M and the parameters $(\nu_c, \nu_{\mathrm{T}}, \Lambda)$ of the classifier must be estimated from training data consisting of the landmarks of a number of s subjects. This training procedure is as follows. The columns of the $m \times n_{\mathrm{ex}}$ matrix X contain the feature vectors of the s subjects (=classes). I.e. $X = (X_1 \ldots X_s)$, with X_c the matrix containing the training data from subject c. Let $(1 \ldots 1)$ denote a row vector with all elements equal to 1. As a first step, an estimate

$$\hat{\mu}_{\mathrm{T}} = \frac{1}{n_{\mathrm{ex}}} X (1 \ldots 1)^{\mathrm{T}} \qquad (4)$$

of the total mean μ_{T} is computed and subtracted from the vectors in the training set. The resulting data matrix is denoted as $X^0 = (X_1^0 \dots X_s^0)$. The dimensionality of the feature space is reduced by projecting the zero-mean training set onto its p *most* significant principal components, thus retaining only the strongest modes of variation. The principal components are computed by means of a singular-value decomposition. I.e.

$$X^0 = U_X S_X V_X^{\mathrm{T}}, \tag{5}$$

with U_X an $m \times n_{\mathrm{ex}}$ orthonormal matrix spanning the column space of X^0, S_X an $n_{\mathrm{ex}} \times n_{\mathrm{ex}}$ diagonal matrix of which the diagonal elements are the singular values of X^0 in descending order, and V_X an $n_{\mathrm{ex}} \times n_{\mathrm{ex}}$ orthonormal matrix spanning the row space of X^0. The dimensionality reduction and the whitening are achieved as follows. Let the $n_{\mathrm{ex}} \times p$ matrix U_{PCA} be the submatrix of U_X consisting of the first $p < n_{\mathrm{ex}}$ columns. Furthermore, let the $p \times p$ matrix S_{PCA} be the first principal $p \times p$ submatrix of S_X. Finally, let the $n_{\mathrm{ex}} \times p$ matrix V_{PCA} be the submatrix of V_X consisting of the first p columns. The whitened data matrix with reduced dimensionality is denoted as

$$Y = \sqrt{n_{\mathrm{ex}} - 1} V_{PCA}^{\mathrm{T}}. \tag{6}$$

The columns are p-dimensional feature vectors with zero mean and independent, identically distributed (i.i.d.) elements. The next step is a linear-discriminant analysis on this set. Note that $Y = (Y_1 \dots Y_s)$, with Y_c the matrix containing the transformed training data from subject c. Let $n_{\mathrm{ex,c}}$ denote the number of columns of Y_c. Estimates

$$\hat{\mu}_c = \frac{1}{n_{\mathrm{ex,c}}} Y_c \left(1 \dots 1\right)^{\mathrm{T}}, \; c = 1, \dots, s, \tag{7}$$

of the within-class variations of each user are computed and subtracted from the corresponding Y_c. This results in a matrix Y^0 of within-class variations. The principal components of the within-class variations are also computed by means of a singular-value decomposition. I.e.

$$Y^0 = U_Y S_Y V_Y^{\mathrm{T}}, \tag{8}$$

with U_Y an $p \times p$ orthonormal matrix spanning the column space of Y^0, S_Y an $p \times p$ diagonal matrix of which the diagonal elements are the singular values of Y^0 in descending order, and V_Y an $n_{\mathrm{ex}} \times p$ orthonormal matrix spanning the row space of Y^0.

A projection onto the d *least* significant principal components of the within-class variations is the last factor of the transform. The resulting dimensionality d of the feature vector satisfies $d \leq \min(p, s-1)$. It is shown in the appendix that a higher dimensionality does not contribute to the likelihood ratio. Therefore, let U_{LDA} denote the submatrix of U_Y consisting of the last d columns and let the $d \times d$ matrix S_{LDA} denote the last principal $d \times d$ submatrix of S_Y. The sequence of transformations described above can be replaced by one multiplication by the $d \times m$ matrix

$$M = \sqrt{n_{\mathrm{ex}} - 1} U_{\mathrm{LDA}}^{\mathrm{T}} S_{\mathrm{PCA}}^{-1} U_{\mathrm{PCA}}^{\mathrm{T}}. \tag{9}$$

The parameters $(\nu_c, \nu_{\mathrm{T}}, \Lambda)$ of the log-likelihood ratio (3) are given by

$$\nu_c = U_{\mathrm{LDA}}^{\mathrm{T}} \hat{\mu}_c + \nu_{\mathrm{T}}, \; c \in \{1, \dots, s\}, \tag{10}$$

with $\hat{\mu}_c$ defined in (7),

$$\nu_{\mathrm{T}} = M \hat{\mu}_{\mathrm{T}}, \tag{11}$$

$\hat{\mu}_{\mathrm{T}}$ defined in (4), and

$$\Lambda = \frac{1}{n_{\mathrm{ex}} - 1} \, S_{\mathrm{LDA}}^2. \tag{12}$$

The effect of the total transform is that the observation space is i.i.d. and that the within-class variations are uncorrelated. The transformed total mean and the class means serve as templates in the verification process.

Instead of the class-independent reduction of dimensionality described above, a class-dependent dimensionality reduction is possible. In [8] a method is described that consists of two steps. First, by means of the above procedure, the dimensionality is reduced to $s-1$. Then a class-dependent transform is applied, which further reduces dimensionality under the constraint that the discrimination between $p(y|c)$ and $p(y)$ is maximum.

III. Experimental evaluation and results

A lab system, similar to the one described in [3], has been realized. A standard web cam captures a color image of the hand, which is converted to a black-and-white image. The position of the hand is fixated by means of 6 reference pins. A side view of the hand at the thumb-side is captured via a mirror that is positioned under 45 degrees. The side view was only used for measuring the heights of the hand at a few positions, which belong to the set of standard high-level features. A black-and-white image with the references pins is shown in Figure 1. The geometrical features are indicated in this figure.

The lab system was used for an experimental comparison of two methods: a standard method based on 30 high-level features, similar to those described in [3], and the contour-based method described above. The standard method also uses a log-likelihood-ratio classifier based on Gaussian probability densities and the dimensionality is reduced by the same procedure as is used for the contour parameters. A database containing 10 to 20 black-and-white images of the right hand of each of 51 subjects was collected. It contains a total of about 850 images.

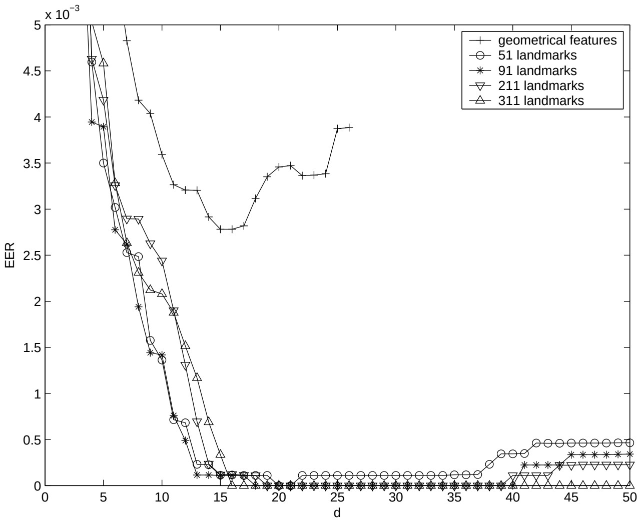

Fig. 3. Equal-error rates as functions of the final dimensionality d, obtained the standard method with 30 high-level features ($+$), and with the new method with 51 ($\circ$), 91 ($*$), 211 ($\bigtriangledown$) and 311 ($\bigtriangleup$) landmarks.

The equal-error rate was taken as the measure of performance. It was estimated as the average outcome of 20 experimental trials. In each trial the feature vectors of each subject were randomly divided into 2 groups: a fraction of 75% was added to a training set; the remaining 25% were added to a test set. The transform coefficients and the classifier parameters were estimated from the training set. The class means served as templates in the verification process. Therefore, the enrollment was part of the training. The equal-error rates were estimated from the test set. Log-likelihood ratios were computed for each feature vector in the test set. This means that there were about 213 (25% of 850) genuine attempts and 10650 imposter attempts (50×213) in each trial. With 20 trials this amounts to a total of 4260 genuine and 213000 imposter attempts. These attempts are not statistically independent.

The parameters of the trails were the number of most significant principal components p, the final dimensionality d of the feature vector, and the number of landmarks $l = 10n_{\mathrm{l}} + 11$. Figure 3 presents the equal-error rates as functions of the final dimensionality d, obtained with the standard method with 30 high-level features ($+$), and with the new method with 51 ($\circ$), 91 ($*$), 211 ($\bigtriangledown$) and 311 ($\bigtriangleup$) landmarks. The number of most significant principal components p, was 26 for the standard and 65 for the new method. The precise value of p is not critical in the new method. Similar results are obtained with any choice of p between 55 and 170.

The best results obtained with the standard method were an equal-error rate of 0.27% $(p = 26, d = 15)$. This is close to the equal-error rates reported for other standard systems, such the equal-error rate of 0.5% in [3]. All tested versions of the new method achieve an equal-error rate of 0% for some range of d. The range of d for which an equal-error rate of 0% is achieved seems to increase with the number of landmarks. For the version with 51 landmarks, this range is smallest $d = 20, 21$. For the version with 311 landmarks, it is largest $15 \leq d \leq 50$. Note that an equal-error rate of zero does not mean that the new method is errorless. Because of the limited size of the data set and the type of experiment, error-rates below about 10^{-4} cannot be measured. Besides, there is also some (unknown) variance on the measured errors. However, it can safely be concluded that the new method yields much better verification results that the standard method for hand-geometry verification. It is also better than the contour-based method in [7], which had an equal-error rate of about 2.5%[2]

IV. Conclusion

A new method for hand-geometry verification, based on a model of the contour of the hand, has been presented. The feature vectors consist of the spatial coordinates of landmarks placed on the contour. The verification is based on a log-likelihood-ratio classifier. Prior to classification, the dimensionality of the feature vector is reduced by a combination of principal component and linear discriminant analysis. In an evaluation experiment based on a data set containing a total of 850 hand contours of 51 subjects, an equal-error rate of 0% was measured. This is substantially better the equal-error rate of the standard method for hand-geometry verification, measured on the same data.

Appendix

Define

$$r = \min(s-1, p). \tag{13}$$

In order to simplify the derivation we redefine the $p \times n_{\mathrm{ex}}$ matrix Y in (6) as

$$Y = V_{\mathrm{PCA}}^{\mathrm{T}}. \tag{14}$$

Without the second dimensionality reduction from p to d, the diagonal matrix Λ in the log-likelihood ratio (3) would be given by

$$\Lambda = S_Y^2 \tag{15}$$

instead of (12). The factor $\frac{1}{n_{\mathrm{ex}}-1}$ in (12) has disappeared because of the redefinition (14). The local means would be given by

$$\nu_c = U_Y^{\mathrm{T}} \hat{\mu}_c + \nu_{\mathrm{T}},\ c \in \{1, \ldots, s\}, \tag{16}$$

[2]The equal-error rates in [7] and [2] were computed from 1-to-1 comparisons, whereas the equal-error rates reported in this paper were computed from 1-to-template comparisons. The latter method produces somewhat lower equal-error rates on the same data. The differences are not large enough to affect the conclusions on the performance.

with $\hat{\mu}_c$ defined in (7). We will prove that

$$\Lambda_{ii} = 1,\ i = r+1, \dots, p, \tag{17}$$

and that

$$(\nu_c)_i = (\nu_{\mathrm{T}})_i,\ c = 1, \dots, s,\ i = r+1, \dots, p. \tag{18}$$

This means that the last $p-r$ dimensions of y do not contribute to $l(y)$.

The matrix Y has the property that $YY^{\mathrm{T}} = I$. We assume that it has full column rank, i.e. $p \leq n_{\mathrm{ex}}$. Recall that $Y = (Y_1 \dots Y_s)$ and that $n_{\mathrm{ex,c}}$ denotes the number of columns of Y_c. Of course

$$YY^{\mathrm{T}} = \sum_{c=1}^{s} Y_c Y_c^{\mathrm{T}}. \tag{19}$$

The matrix Y^0 of within-class variation can be written as

$$Y^0 = (Y_1 - \hat{\mu}_1(1 \dots 1) \dots Y_s - \hat{\mu}_s(1 \dots 1)), \tag{20}$$

with $\hat{\mu}_c$ defined in (7). We now have that

$$\begin{aligned} Y^0(Y^0)^{\mathrm{T}} &= \sum_{c=1}^{s}(Y_c - \hat{\mu}_c(1\dots1))(Y_c - \hat{\mu}_c(1\dots1))^{\mathrm{T}} \\ &= \sum_{c=1}^{s} Y_c Y_c^{\mathrm{T}} - \\ &\quad \hat{\mu}_c(1\dots1)Y_c^{\mathrm{T}} - Y_c(1\dots1)^{\mathrm{T}}\hat{\mu}_c^{\mathrm{T}} + \\ &\quad \hat{\mu}_c(1\dots1)(1\dots1)^{\mathrm{T}}\hat{\mu}_c^{\mathrm{T}} \\ &= \sum_{c=1}^{s} Y_c Y_c^{\mathrm{T}} - n_{\mathrm{ex,c}}\hat{\mu}_c\hat{\mu}_c^{\mathrm{T}} \\ &= YY^{\mathrm{T}} - W \begin{pmatrix} n_{\mathrm{ex,1}} & & \\ & \ddots & \\ & & n_{\mathrm{ex,s}} \end{pmatrix} W^{\mathrm{T}} \\ &= I - W \begin{pmatrix} n_{\mathrm{ex,1}} & & \\ & \ddots & \\ & & n_{\mathrm{ex,s}} \end{pmatrix} W^{\mathrm{T}}, \end{aligned} \tag{21}$$

with $W = (\hat{\mu}_1 \dots \hat{\mu}_s)$. Because Y has zero column mean, we have that $W(n_{\mathrm{ex,1}} \dots n_{\mathrm{ex,s}})^{\mathrm{T}} = 0$. Therefore, the columns of W are linearly dependent and the maximum rank of W is $\min(s-1, p) = r$. Consider the singular-value decomposition

$$W \begin{pmatrix} n_{\mathrm{ex,1}} & & \\ & \ddots & \\ & & n_{\mathrm{ex,s}} \end{pmatrix}^{\frac{1}{2}} = U_W S_W V_W^{\mathrm{T}}, \tag{22}$$

with U_W an $n_{\mathrm{ex}} \times n_{\mathrm{ex}}$ orthonormal matrix of which the first r columns span the column space of W, S_W an $n_{\mathrm{ex}} \times n_{\mathrm{ex}}$ diagonal matrix of which at most the first r diagonal elements are non-negative and all the others are zeros, and V_W an $s \times n_{\mathrm{ex}}$ orthonormal matrix spanning the row space of W. Pre- and post-multiplying (21) with U_W^{T} and U_W, respectively, and using (22) yields

$$U_W^{\mathrm{T}} Y^0 (Y^0)^{\mathrm{T}} U_W = I - S_W^2. \tag{23}$$

Since all matrices on the right-hand side of (23) are diagonal, U_W must diagonalize $Y^0(Y^0)^{\mathrm{T}}$, such that

$$U_Y = U_W \tag{24}$$

and

$$S_Y^2 = I - S_W^2, \tag{25}$$

with U_Y and S_Y defined in (8). The result (17) follows from (15) and (25) and because the last $p-r$ diagonal elements of S_W^2 are zero. The result (18) follows because the last $p-r$ columns of U_W, and thus of U_Y, are orthogonal to the column space of $W = (\hat{\mu}_1 \dots \hat{\mu}_s)$, which means that the last $p-r$ elements of $U_Y^{\mathrm{T}}\hat{\mu}_c$ are zero.

REFERENCES

[1] M. Golfarelli, D. Maio, and D. Maltoni. On the error-reject trade-off in biometric verification systems. *IEEE Trans. PAMI*, 19(7):786–796, July 1997.

[2] A.K Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. In *Proceedings of the second International Conference on Audio- and Video-Based Personal Authentication (AVBPA)*, pages 166–171, Washington, March 1999.

[3] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos. Biometric identification through hand geometry measurements. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1168–1171, October 2000.

[4] N. Pavešić, S. Ribarić, and D. Ribarić. Personal authentication using hand-geometry and palmprint features – the state of the art. In *Workshop Proceedings – Biometrics: Challenges arising from Theory to Practice*, pages 17–26, Cambridge, UK.

[5] G.H. Granlund. Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers*, 21(2):195–201, 1972.

[6] T.F Cootes and C.J. Taylor. Statistical models of appearance for computer vision. Internal report, Image Science and Biomedical Engineering, University of Manchester, 2001.

[7] A.K. Jain and N. Duta. Deformable matching of hand shapes for verification. In *Proc. of IEEE Int. Conf. on Image Prcoessing*, Kobe, Japan, October 1999.

[8] R.N.J. Veldhuis and A.M. Bazen. Maximum discrimination analysis (MDA) as a means for dimension reduction in biometric verification. In *Workshop Proceedings – Biometrics: Challenges arising from Theory to Practice*, pages 31–34, Cambridge, UK.