

Eigenvalue correction results in face recognition

Anne Hendrikse Raymond Veldhuis Luuk Spreeuwers
University of Twente
Fac. EEMCS, Signals and Systems Group
Hogekamp Building, 7522 NB
Enschede, The Netherlands
a.j.hendrikse@utwente.nl

Asker Bazen
Uniqkey Biometrics
The Netherlands
a.m.bazen@uniqkey.com

Abstract

Eigenvalues of sample covariance matrices are often used in biometrics. It has been known for several decades that even though the sample covariance matrix is an unbiased estimate of the real covariance matrix [3], the eigenvalues of the sample covariance matrix are biased estimates of the real eigenvalues [6]. This bias is particularly dominant when the number of samples used for estimation is in the same order as the number of dimensions, as is often the case in biometrics. We investigate the effects of this bias on error rates in verification experiments and show that eigenvalue correction can improve recognition performance.

1 Introduction

In biometrics the objective is to automate decisions about peoples identities based on measurable physiological characteristics, e.g. characteristics of the face. Covariance matrices are used to describe relations between these characteristics and relations between the identity of people and their characteristics. Using a Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which make extensive use of these covariance matrices, we construct a system which can verify identity claims using photo's of the face.

However, the required covariance matrices are not available on forehand and need to be estimate from examples. One estimator is the sample covariance matrix given by:

$$\hat{\Sigma} = \frac{1}{N-1} X \cdot X^T \quad (1)$$

where X is a matrix with zero mean data. Each of the N columns of X contains a sample. Any covariance matrix can be decomposed into eigenvector and eigenvalues according to

$$\Sigma = \Phi \cdot \Lambda \cdot \Phi^T \quad (2)$$

where the columns of orthonormal matrix Φ contain the eigenvectors and their corresponding eigenvalues are on the diagonal of the diagonal matrix Λ .

When N is in the same order as the dimensionality of the samples, p , the eigenvalues of $\hat{\Sigma}$ have a significant bias, which means they significantly deviate from the real eigenvalues. In this article we will apply methods to correct the bias in eigenvalues estimated from face data. We will show that a correction algorithm we denote as Two Subset is capable of reducing the error rates of the verification system.

The paper has the following setup. In section 2 we will describe the verification system we used to test the influence of eigenvalue correction. We will indicate at which point eigenvalue correction makes sense. An important aspect of the system description is the data model we assumed. To describe the problems with the eigenvalue estimates, we will introduce the bias and the variance of estimators in section 3. In section 4 a few correction algorithms will be presented which we used to remove the bias in the eigenvalue estimates.

To demonstrate the use of eigenvalue corrections we present two experiments. The first experiment, in which we used synthetic face like data, is reported in section 5. The second experiment, in which we used real face data, is reported in section 6. Finally some conclusions are presented in section 7.

2 Verification system setup

We will briefly describe the PCA-LDA verification system we have used. For a more extended discussion on such a system, see [1] or [8].

The input of the system are facial images. With some preprocessing, feature vectors are generated from these images. We assume a feature vector x can be modeled as a composition of a within part x_w and a between part x_b via $x = x_b + x_w$. x_b models the differences between different people and x_w models the variations between different feature vectors of the same person. The distribution of x_b is modeled by $N(\mu_t, \Sigma_b)$, a Gaussian distribution with mean μ_t and a covariance matrix Σ_b . The distribution of x_w is modeled by $N(0, \Sigma_w)$. The distribution of x is given by $N(\mu_t, \Sigma_t)$.

After the preprocessing, a transformation matrix T is constructed which transforms a feature vector from the feature space to a classification space. To find T a set of labeled example images is needed. This set is denoted as training set.

The first step to determine T is to perform PCA on the training set. With the results of PCA, a transformation matrix is constructed which compresses the data and also whitens the data. Whitening means that the data components stay uncorrelated under any rotation.

By applying LDA in the whitened space [3], the most discriminating basis can be determined. Combining this projection with the projection from PCA results in the following transformation matrix:

$$T = \hat{\theta}_b^T \cdot \hat{\Lambda}_t^{-\frac{1}{2}} \cdot \hat{\Phi}_t^T \quad (3)$$

where $\hat{\Lambda}_t$ is a diagonal matrix with the M largest eigenvalues of $\hat{\Sigma}_t$ on the diagonal, $\hat{\Phi}_t$ is a matrix of which the columns are the eigenvectors belonging to the eigenvalues in $\hat{\Lambda}_t$, and $\hat{\theta}_b$ are the eigenvectors belonging to the Q largest eigenvalues of the between sample covariance matrix in the whitened space.

To verify if a feature vector v belongs to class c a matching score based on log likelihood is calculated:

$$L(v, c) = -(T \cdot (v - \mu_c))^T \cdot \hat{\Gamma}_w^{-1} \cdot (T \cdot (v - \mu_c)) + (T \cdot (v - \mu_t))^T \cdot (T \cdot (v - \mu_t)) \quad (4)$$

where $\hat{\Gamma}_w$ is a diagonal matrix with the eigenvalues of the within sample covariance matrix in the classification space. If the matching score exceeds a threshold, the claim that v belongs to class c is accepted.

2.1 Modifications for eigenvalue correction

In this verification system, there are two points where eigenvalue correction may improve results: in the whitening step where the data is scaled based on eigenvalue estimates and in the matching score calculation where the eigenvalues of the within covariance matrix in the classification space are needed. We performed eigenvalue correction after the compression, but before the whitening step.

At first sight, it seems that the eigenvalues of $\hat{\Sigma}_t$ need to be corrected. However, under the assumed model, Σ_t can be written as $\Sigma_b + \Sigma_w$. These matrices are estimated with $N_b = C - 1$ and $N_w = N - N_b - 1$ samples respectively (C is the number of classes). This results in a different bias of the eigenvalues of sample estimates of Σ_b and Σ_w . We therefore implemented the correction in the following manner:

1. Estimate Σ_w and Σ_b .
2. Decompose both covariance matrices in eigenvectors and eigenvalues.
3. Construct new estimates of the covariance matrices using the original eigenvector estimates and the corrected eigenvalues.
4. Sum the two estimates to get a new estimate of Σ_t .

The corrected estimate of the covariance matrices is given by:

$$\tilde{\Sigma}_r = \hat{\Phi}_r \cdot f_{N_r}(\hat{\Lambda}_r) \cdot \hat{\Phi}_r \quad (5)$$

where r is either w or b and $f_{N_r}(\hat{\Lambda}_r)$ is an eigenvalue correction algorithm.

3 Estimator statistics

In the previous section we required estimates of eigenvectors and eigenvalues. One estimator of the eigenvalues is the set of eigenvalues of the sample covariance matrix. A good estimator should provide estimates close to the real value. Two measures indicating the accuracy of an estimator are the mean and the variance of estimates generated by the estimator.

One method of retrieving these statistics of an estimator is General Statistical Analysis (GSA) [4]. In GSA the analysis the following limit situation is considered: $N \rightarrow \infty$ while $\frac{p}{N} \rightarrow \gamma$, where N is the number of samples, p is the dimensionality of the samples and γ is some positive constant. In face recognition, both N and p are large, while their ratio is close to 1. Therefore results from GSA may apply in face recognition. Under GSA, the mean estimate of the sample eigenvalues is not equal to real set of eigenvalues, the sample eigenvalue estimator is therefore biased.

Some estimators can trade bias for variance in the estimate. The variance in the estimate of eigenvalue i is given by $E[(\Lambda_i - \hat{\Lambda}_i)^2]$. In general decreasing the number of samples used for estimation, increases the variance of the estimate. The Two Subset correction we describe in section 4.4 reduces the number of samples used for eigenvalue estimation, therefore increasing the variance. But, judging from the results of the experiments, the reduction in bias justifies this cost.

4 Correction algorithms

To reduce the bias in the estimated eigenvalues, we have used a number of correction algorithms. We will describe them briefly in this section. We will use the following conventions: the genuine eigenvalue i is denoted by λ_i and is referred to as the population eigenvalue i . The sample estimate of this population eigenvalue i is referred to as the sample eigenvalue i and is denoted by l_i . The corrected eigenvalues are denoted by $\hat{\lambda}_i$.

4.1 Muirhead correction

The first correction algorithm has been proposed by Muirhead [5]. Muirhead derived expressions of the distribution of the eigenvalues under the condition that the samples are Gaussian distributed. Based on this distribution, he derived an expression for the most likely set of population eigenvalues given a set of sample eigenvalues:

$$\hat{\lambda}_i = l_i - \frac{1}{n} l_i \sum_{\substack{j=1 \\ j \neq i}}^m \frac{l_j}{l_i - l_j} + O(n^{-2}) \quad (i = 1, \dots, m) \quad (6)$$

If neighboring eigenvalues are very close in value, the correction may become unstable. We therefore choose to ignore some of the closest eigenvalues in the correction, so $j \neq \{(i - K_{low}) \dots (i + K_{high})\}$.

4.2 Karoui correction

The second algorithm is proposed by Karoui. The algorithm is based on the Marcenko Pastur equation. This equation describes the relation between the sample eigenvalues and population eigenvalues in the GSA limit. The eigenvalue sets then converge to distributions. The equation is valid under some mild conditions on the population eigenvalue distribution [7].

There is no closed form expression which gives the density of the population eigenvalues when the sample eigenvalues are known or visa versa. Therefore Karoui derived an optimization problem from the equation and designed an algorithm which performs this optimization. For the algorithm itself we refer to [2], because a more elaborate discussion would fall outside the scope of this article. The result of this algorithm is a description of the population density in the form of a weighed sum of delta pulses, bars and triangles. We did not use triangles in our implementation. We translate this density back into a set of eigenvalues.

4.3 Iterative feedback correction

The iterative feedback correction is a new algorithm to our knowledge. It attempts to correct the eigenvalues empirically. The algorithm assumes some initial guess of the population eigenvalues. Using a random number generator, synthetic data is generated with these eigenvalues. The sample eigenvalues of this synthetic data are compared with the originally measured eigenvalues. Based on differences between the two sets of sample eigenvalues the guessed population eigenvalues are adjusted. The process is repeated until the sample eigenvalues of the synthetic data closely match the measured eigenvalues. See figure 1 for a schematic representation.

As shown in figure 1 we implemented the comparison between the synthetic sample eigenvalues and the measured sample eigenvalues by dividing the two sets element wise. The resulting vector is then multiplied element wise with the guessed population eigenvalues to get a new guess at the population eigenvalues.

4.4 Two Subset correction

The Two Subset correction is a classical technique in statistics to remove bias in estimates. In the Two Subset correction, the input data X is split in Two Subsets X_1 and X_2 . From the first subset eigenvectors are estimated, denoted $\hat{\Phi}_1$. In the second

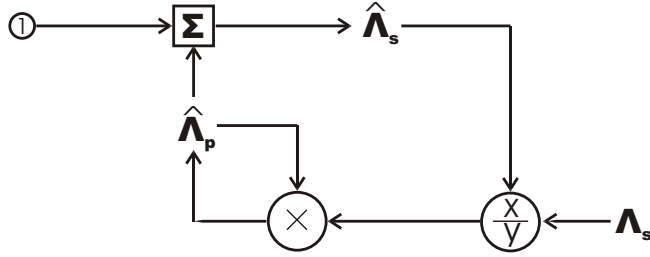


Figure 1: Schematic representation of iterative feedback correction algorithm.

set the variances along these estimated eigenvectors are estimated, denoted $\hat{\Lambda}_2$, via:

$$\hat{\Lambda}_2 = \text{diag} \left\{ \hat{\Phi}_1^T \cdot \frac{1}{N_2 - 1} X_2 X_2^T \cdot \hat{\Phi}_1 \right\} \quad (7)$$

$\hat{\Lambda}_2$ will be used as estimate for the population eigenvalues. The eigenvalues in $\hat{\Lambda}_2$ do not contain the bias of the original estimates. However, since the estimation is performed on half of the original set, the variance in the estimate will have increased, as described in section 3.

5 Synthetic facial data test

To show the effects of the correction algorithms in verification problems, we present two experiments. In the first experiment synthetic data is used. By using synthetic data a comparison can be made between the true population eigenvalues and the reconstructed population eigenvalues (see for instance [2]). In biometrics, we are more interested in lower error rates. With synthetic data, the real variances on the estimated eigenvectors can be calculated, giving a theoretical correction. This correction gives an indication of the best improvement.

In section 2 we presented a model for facial data. The synthetic data is generated using this model. The generation process can be split into two steps. In the first step random samples are drawn from a between class distribution. Each sample will be the class average of an individual. The input of this stage is a Gaussian random number generator which draws samples from $N(0, I)$. The samples are then scaled and rotated such that their covariance matrix equals Σ_b . Finally a total mean μ_t may be added, but in our experiment it is set to zero.

In the second step a second random number generator draws a number of samples per class to model the within class variation. The samples are scaled and rotated such that their covariance matrix matches Σ_w . Next a class mean vector, generated in the first step, is added. The result is a set of synthetic samples.

To generate synthetic data similar to facial data, we extracted model parameters from a set of face images in the FRGC database. 8941 photos were selected in the database. The photos originated from 515 individuals. Most photo's were taken under controlled conditions with limited variations in pose and illumination. Also the faces had a neutral expression and nobody had glasses.

To compare the different correction algorithms, a verification experiment was performed with a training set of limited size. The generated training set contained 70 synthetic classes. Each class had 4 samples, resulting in 280 samples. The initial compression reduced the data dimensionality to 150 components. In the LDA step,

the dimensionality was further reduced to 60. 6 correction options were tested: no correction, theoretical correction, Muirhead correction, Iterative Feedback correction, Two Subset correction and Karoui correction.

To test the trained transformation matrices, a test set was generated containing 1000 classes. For each class 10 enrollment samples and 10 probe samples were generated. The Detection Error Trade-off (DET) curves are shown in figure 2.

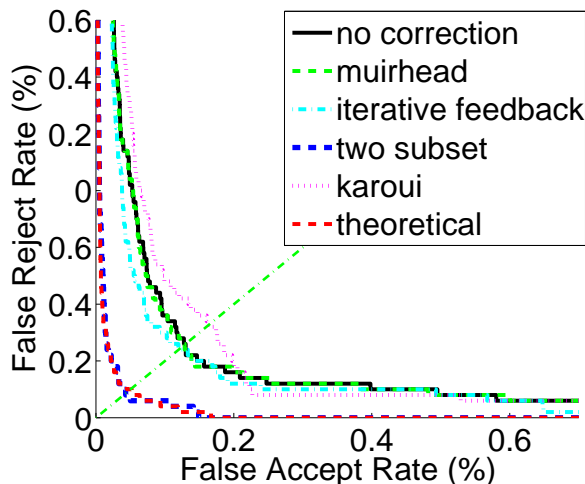


Figure 2: DET curves for test with synthetic faces.

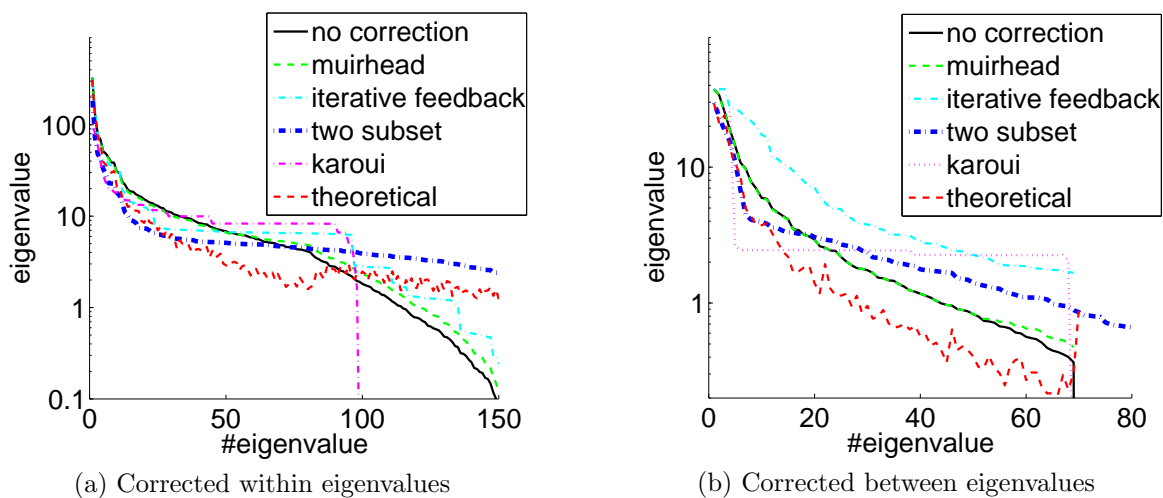


Figure 3: Eigenvalues of the synthetic data test.

The error rates achieved in the synthetic experiment are more than an order lower than the error rates in real facial data. This seems to indicate that the bias in eigenvalue is not the only factor causing errors.

The theoretical line indicates that if the true variances along the eigenvectors can be found, the EER is decreased from 0.13% to 0.05%. This is a considerable improvement. The Two Subset correction achieves comparable error rates as the theoretical correction.

Apparently the increase in variance of the estimates of the eigenvalues is compensated by the removal of the bias. The other correction algorithms only fluctuate around the no correction line, and provide no structural improvement. The Karoui correction even deteriorates the performance.

The corrected eigenvalues are shown in figure 3. There is a considerable difference between the theoretical correction and the Two Subset correction, but their DET curves are nearly the same. The difference is the largest for the smaller between eigenvalues, suggesting the error rates are more influenced by changes in the largest between eigenvalues.

Karoui changed all within eigenvalues above number 100 to almost zero. If there is any between variance in those directions, then the LDA step will identify them as highly discriminating directions. This may explain the poor performance of the Karoui correction.

Combining the results from the Two Subset correction and the Karoui correction, corrections in the largest between eigenvalues and the smallest within eigenvalues seem to have the most effect on the error rates. This may be explained by the fact that LDA finds directions in which the ratio of the between variance over the within variance is the largest.

6 FRGC facial data test

The synthetic data test shows that reduction of the error rates is possible when the assumed model is accurate. In the next test we used real face data. To find the parameters of the face model in the synthetic data experiment, a set of images from the FRGC database was selected. In this test the same set is used. The set is split in a training set and a test set. For the training set 70 identities were selected with a maximum of 5 samples per identity. The remaining 445 identities are used as test set.

There is a small deviation from the setup as described in section 2. The dimension reduction is only used to remove zero eigenvalues. After applying the correction, a further reduction to 150 dimensions is performed. The same correction methods as used in the synthetic data test are used, except for the theoretical correction, which is not possible when the population eigenvalues are unknown.

The experiment is repeated 5 times with different data divisions. Resulting EERs are shown in tabel 1. The last column shows the average of the EERs over the 5 runs. Karoui deteriorated the EER in most runs. The iterative feedback fluctuates around the no correction performance. The Muirhead correction gives a very small improvement of the EER. The Two Subset correction never worsens the performance and gives on average a significant improvement in EER of 0.4%.

Table 1: EER of 5 runs in percentages

iteration	1	2	3	4	5	AV
no correction	3,57	3,49	2,90	3,68	3,82	3,49
muirhead	3,50	3,38	2,50	3,54	3,64	3,31
iterative feedback	3,79	3,84	2,17	3,75	3,64	3,44
two subsets	3,24	3,49	2,17	3,42	3,03	3,07
karoui	4,43	4,71	2,17	3,58	3,68	3,71

Figure 4 shows the DET curves of the correction algorithms for one run. The Two Subset is significantly below the no correction curve. The error rates of the Karoui correction are considerably worse than the no correction.

Figure 5 shows the corrected eigenvalues. The Two Subset correction considerably lowered the first few eigenvalues, especially the within eigenvalues. The smaller eigenvalues have become more similar. The Muirhead and Iterative Feedback correction do not differ considerably from the no correction curve. The Muirhead correction only altered the last eigenvalues. The Karoui correction reduced within eigenvalues number 110 and larger to almost 0. This is the same behaviour as in the synthetic data experiment. Karoui also changed most of the between eigenvalues to a value of around 20. Even though these are large changes, the DET curve stays in the same order as the no correction curve.

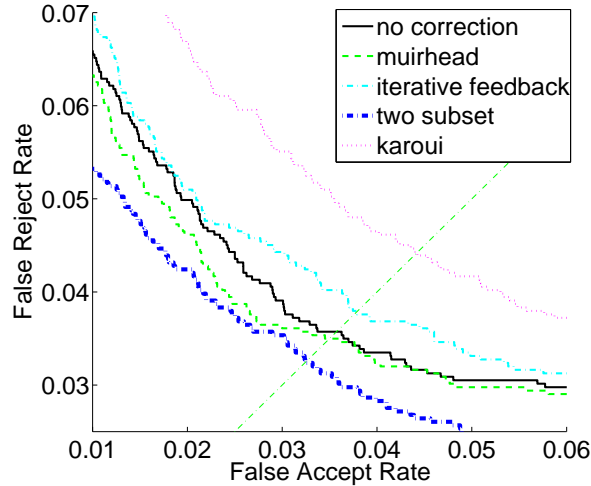


Figure 4: DET curves of FRGC facial data test.

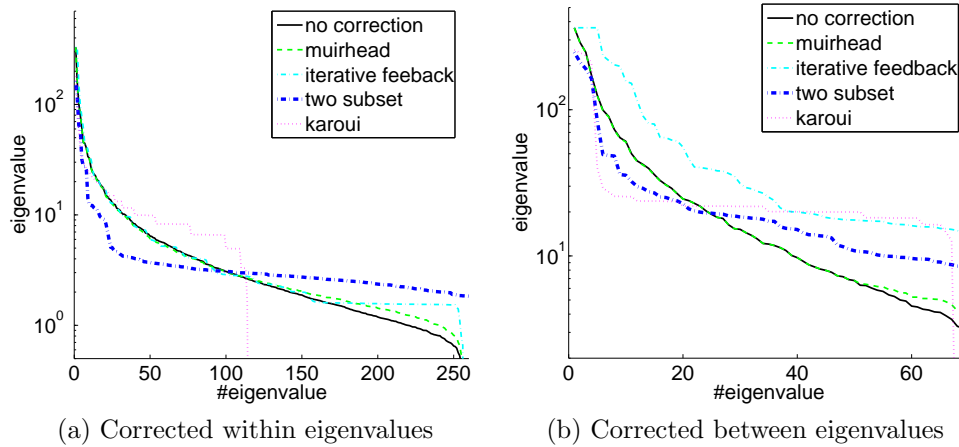


Figure 5: Corrected eigenvalues of FRGC facial data test.

7 Conclusion

In the experiments the more advanced correction algorithms provided little improvement at best, while the empirical based Two Subset correction provided significant

improvements in both the synthetic data test and the real face test. The corrected eigenvalues of the Two Subset correction show that the largest eigenvalues are over estimated, while the smallest eigenvalues are under estimated. Except for the iterative feedback, the other correction algorithms showed similar behaviour. We intend to publish a more detailed analysis in future publications. From the success of the Two Subset we also conclude that the error rates are much more influenced by the bias in the eigenvalue estimates than their variance.

In the synthetic data experiment, the Two Subset correction provided the same improvement as the theoretical correction. If this relation holds as well for real facial data, the relative small error reduction of the Two Subset correction indicates that only a small error reduction can be achieved. A reason for this might be that the error rates in real facial data are largely determined by other factors than the bias in the eigenvectors. This conclusion is also supported by the more than an order difference between the error rates of real facial data experiment and synthetic data experiment.

The influence of the eigenvalue corrections on the error rates is somewhat difficult to determine. The large changes the Karoui correction made, had limited effect on the error rates. Still it appears that corrections in the smallest within eigenvalues and the largest between eigenvalues are the most important. This is in line with the objective of LDA to find the basis with small within variances and large between variances.

References

- [1] G. M. Beumer, A. M. Bazen, and R. N. J. Veldhuis. On the accuracy of eers in face recognition and the importance of reliable registration. In *5th IEEE Benelux Signal Processing Symposium (SPS-2005), Antwerp, Belgium*, pages 85–88, secretariaat in Delft, April 2005. IEEE Benelux Signal Processing Chapter.
- [2] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *ArXiv Mathematics e-prints*, september 2006.
- [3] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [4] V.L. Girko. *Theory of Random Determinants*. Kluwer, 1990.
- [5] Robb John Muirhead. *Aspects of multivariate statistical theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, inc, 1982.
- [6] Jack W. Silverstein. Eigenvalues and eigenvectors of large dimensional sample covariance matrices. In *Random Matrices and their Applications*, volume 50 of *Contemporary Mathematics*, pages 153–159. American Mathematical Society, 1986.
- [7] Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivar. Anal.*, 55(2):331–339, 1995.
- [8] R. N. J. Veldhuis, A. M. Bazen, W. Booij, and A. J. Hendrikse. Hand-geometry recognition based on contour parameters. In A. K. Jain and N. K. Ratha, editors, *SPIE Biometric Technology for Human Identification II, Orlando, FL, USA*, pages 344–353, Washington, March 2005. SPIE – The Int. Society for Optical Engineering.