

Analysis of Eigenvalue Correction Applied to Biometrics

Anne Hendrikse¹, Raymond Veldhuis¹, Luuk Spreeuwiers¹, and Asker Bazen²

¹ University of Twente, Fac. EEMCS, Signals and Systems Group, Hogeekamp Building, 7522 NB, Enschede, The Netherlands

a.j.hendrikse@utwente.nl

² Uniqkey Biometrics, The Netherlands

a.m.bazen@uniqkey.com

Abstract. Eigenvalue estimation plays an important role in biometrics. However, if the number of samples is limited, estimates are significantly biased. In this article we analyse the influence of this bias on the error rates of PCA/LDA based verification systems, using both synthetic data with realistic parameters and real biometric data. Results of bias correction in the verification systems differ considerable between synthetic data and real data: while the bias is responsible for a large part of classification errors in the synthetic facial data, compensation of the bias in real facial data leads only to marginal improvements.

1 Introduction

An important aspect of biometrics is data modeling. Modeling the statistics of data by covariance matrices is an example. Two techniques which rely on modeling by covariance matrices are Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Because the covariance matrix of the data generating process, Σ , is usually unknown, it needs to be estimated from a training set. An often used estimate is the sample covariance matrix:

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

where the columns of matrix X contain the training samples with the mean subtracted and N is the number of samples in the set.

In the modeling process we are often more interested in functions of the covariance matrix than in the covariance matrix itself. A commonly used function is the decomposition of the covariance matrix in eigenvectors and eigenvalues. The decomposition results we call population eigenvectors and population eigenvalues when derived from Σ and we call them by sample eigenvectors and sample eigenvalues when derived from $\hat{\Sigma}$. The i^{th} population eigenvalue is denoted by λ_i and the i^{th} sample eigenvalue is denoted by l_i . Though $\hat{\Sigma}$ is an unbiased estimate of Σ [1], the estimate of λ_i by l_i does have a bias.

In this article, we analyse the effect of this bias with two verification experiments. In the first experiment we use synthetic data so we can compare the verification performance of the system with and without the bias. In both the synthetic data and the real biometric data we compare performance improvement when applying several bias correction algorithms in several configurations.

An analysis of the bias is given in section 2.1. In section 2.2 we present a number of algorithms which reduce the bias. In section 3 we describe the verification system used in the experiments. We indicate where the bias will have its largest effect and where it should be compensated.

In section 4.1 we present an experiment with synthetic facial data, to determine the effect of the bias when the assumed model is correct. In section 4.2 we repeat the experiment with real facial data. In section 5 we present conclusions.

2 Eigenvalue Bias Analysis and Correction

2.1 Eigenvalue Bias Analysis

To find the statistics of estimators often Large Sample Analysis (LSA) is performed. The sample eigenvalues show no bias in this limit case where the number of samples is large enough that it solely determines the statistics of the estimator. However, in biometrics, the number of samples is often in the same order as the number of dimensions or even lower. Therefore, in the analysis of the statistics of the sample eigenvalues the following limit may be considered: $N, p \rightarrow \infty$ while $\frac{p}{N} \rightarrow \gamma$. Here N is the number of samples used, p is the number of dimensions and γ is some positive constant. Analysis in this limit are denoted General Statistical Analysis (GSA) [2]. In GSA the sample eigenvalues do have a bias.

To demonstrate GSA, we estimated sample eigenvalues of synthetic data with population eigenvalues chosen uniformly between 0 and 1. We kept $\gamma = \frac{1}{5}$ while we varied the dimensionality between 4, 20 and 100. In Figure 1 we show both the population eigenvalue probability function and the sample eigenvalue probability functions for 4 repetitions, given by

$$F_p(l) = p^{-1} \sum_{i=1}^p u(l - l_i) \quad (2)$$

where $u(l)$ is the step function. The empirical probability functions converge with increasing dimensionality, and they converge to a different probability function as the population probability function, due to bias. This example also shows that bias reduction is only possible for a minimum dimensionality, because only then the largest part of the error in l_i as estimate of λ_i is caused by the bias.

2.2 Eigenvalue Bias Correction Algorithms

The bias is a deterministic error and can therefore be compensated. In this section we present a number of correction algorithms we used in the verification experiments to reduce the bias. The correction algorithms provide new estimates of the population eigenvalues, which are denoted by $\hat{\lambda}_i$.

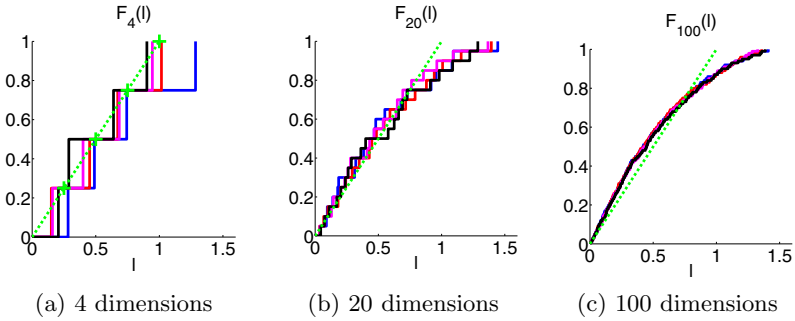


Fig. 1. Examples of eigenvalue estimation bias toward the GSA limit. All lines indicate empirical probability functions based on sets of eigenvalues (see equation 2). The dashed line indicates the population distribution, the four solid lines are the empirical sample distribution.

1. The *Muirhead correction* [3] is given by a maximum likelihood estimate of the population eigenvalues:

$$\hat{\lambda}_i = l_i - \frac{1}{n} l_i \sum_{j=1 \dots i-K, i+K \dots p} \frac{l_j}{l_i - l_j} \tag{3}$$

In the original formula K was set to one. However, to prevent strong fluctuations, we set $K = 50$, which is a simplified version of the Stein[4] algorithm.

2. The *Karoui correction* [5] is based on the Marčenko Pastur equation [6] which gives a relation between sample eigenvalues and the population eigenvalues in the limit considered in GSA. The algorithm finds an estimate of the empirical population eigenvalue probability function (Equation 2, with l replaced by λ) as a weighed sum of fixed probability functions, in our case a set of delta pulses and bar functions.
3. The *Iterative feedback* algorithm was developed by the authors and is new to our knowledge. To find the population eigenvalues the algorithm starts with an initial guess for the population eigenvalues, $\hat{\lambda}_{i,1}$. In the m^{th} iteration of the algorithm, synthetic data is generated with population eigenvalues equal to $\hat{\lambda}_{i,m}$. The sample eigenvalues $\hat{l}_{i,m}$ of this synthetic data are determined. $\hat{\lambda}_{i,m+1}$ is constructed via $\hat{\lambda}_{i,m+1} = \hat{\lambda}_{i,m} \cdot \frac{l_i}{\hat{l}_{i,m}}$. These steps are repeated until $\sum_{i=1}^p \left(\hat{l}_i - l_i \right)^2$ is below a preset threshold or $m > m_{\max}$.
4. The *Two Subset* correction is a classical technique in statistics to remove bias in estimates, where X is split in two subsets X_1 and X_2 . From $(N/2 - 1)^{-1} X_1 X_1^T$ eigenvectors are estimated, denoted $\hat{\Phi}_1$. The variances in the second set along these estimated eigenvectors are used as $\hat{\lambda}_i$'s, so $\hat{\lambda}_i = \hat{\Phi}_{1,i}^T \cdot \frac{1}{N_2-1} X_2 X_2^T \cdot \hat{\Phi}_{1,i}$. The $\hat{\lambda}_i$'s do not contain the bias of the original estimates. However, since the estimation is performed on half of the original set, the variance of the estimate increases. This might explain why this correction is not commonly used.

3 Verification System Description

3.1 System Setup

In our experiments we test the influence of the bias of eigenvalues in biometric systems, using a well known baseline PCA LDA system in our experiments. In this section we give a brief description of this system. For a more detailed discussion we refer to [7].

The input of the verification system are facial images. On these images some standard preprocessing is done, which results in a data sample \mathbf{x} for each image. To transform these input vectors to a space where classification is possible, a transformation matrix \mathbf{T} is determined in 3 steps based on a training set of example samples. In the first two steps we use PCA to reduce the dimensionality and whiten the data.

In the third step a projection to the most discriminating subspace is determined by modeling each data sample as $\mathbf{x} = \mathbf{x}_w + \mathbf{x}_b$. Variations between samples from the same class are modeled by \mathbf{x}_w , which is distributed as $N(0, \Sigma_w)$, a multi variate normal distribution with mean 0 and covariance matrix Σ_w . We model the variations between classes by \mathbf{x}_b , which is distributed as $N(\mu_t, \Sigma_b)$. Since the data is whitened, the most discriminating subspace is the subspace of the largest eigenvalues of Σ_b . Therefore the transformation matrix \mathbf{T} is given by:

$$\mathbf{T} = \hat{\Phi}_{b,C_2}^T \cdot \hat{\Lambda}_{t,C_1}^{\frac{1}{2}} \cdot \hat{\Phi}_{t,C_1}^T \quad (4)$$

where $\hat{\Phi}_{t,C_1}$ are the first C_1 eigenvectors of $\hat{\Sigma}_t$, the covariance matrix of the training set, and $\hat{\Lambda}_{t,C_1}$ is a diagonal matrix with as diagonal the first C_1 eigenvalues of $\hat{\Sigma}_t$. $\hat{\Phi}_{b,C_2}$ are the first C_2 eigenvectors of $\hat{\Sigma}_b$.

After projecting samples in the classification space, we compare sample \mathbf{x} with class c by calculating a matching score. We accept an identity claim if the score is above a certain threshold. The score is based on the log likelihood:

$$L(\mathbf{x}, c) = -(\mathbf{T} \cdot \mathbf{x} - \mu_c)^T \cdot \hat{\Sigma}_w^{-1} \cdot (\mathbf{T} \cdot \mathbf{x} - \mu_c) + (\mathbf{T} \cdot \mathbf{x} - \mu_t)^T \cdot (\mathbf{T} \cdot \mathbf{x} - \mu_t) \quad (5)$$

3.2 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 perform eigenvalue correction after the dimensionality reduction, 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, the total covariance matrix Σ_t can be written as $\Sigma_b + \Sigma_w$. These matrices are estimated by $(C - 1)^{-1} \sum_{c=1}^C \mu_c \mu_c^T$ and $(N - C)^{-1} \sum_{i=1}^N (\mathbf{x}_i - \mu_{\ell(\mathbf{x}_i)})(\mathbf{x}_i - \mu_{\ell(\mathbf{x}_i)})^T$ respectively, where C is the number of classes in the training set, μ_c is the mean of the training samples of class c , and

$\ell(\mathbf{x}_i)$ returns the class index of sample \mathbf{x}_i . Because both matrices are estimated with a different number of samples, their eigenvalues have a different bias. We therefore perform 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 matrix is given by $\tilde{\Sigma}_r = \hat{\Phi}_r \cdot f_{N_r}(\hat{\Lambda}_r) \cdot \hat{\Phi}_r^T$, where r is either w or b and $f_{N_r}(\hat{\Lambda}_r)$ is an eigenvalue correction algorithm.

4 Experiments

In this section we describe two verification experiments with the system presented in the previous section. In the first experiment we used synthetic facial data while in the second experiment we used real facial data.

4.1 Synthetic Data Experiment

To generate synthetic data close to real facial data, we determined the data structure of a large set of face images in the FRGC database. The data contained 8941 facial images. All facial images were taken under controlled conditions with limited variations in pose and illumination. Also the faces in the facial images had a neutral expression and nobody wore glasses.

We model the facial data with the model in section 3. For generating synthetic data adhering to this model with parameters close to real facial data, we estimated the within class covariance matrix Σ_w and the between class covariance matrix Σ_b from the FRGC data. Since the eigenvalues of these estimates also contain a bias, we corrected their eigenvalues with the Two Subset correction, knowing from previous experiments that this correction led to better estimates of eigenvalues [8]. We kept μ_t zero.

We generated a small training set of 70 identities, with 4 samples per identity, so the bias should be comparable to small real face data sets. This training set was used to train a verification system. In the dimensionality reduction stage of the training the dimensionality was reduced to 150. In the LDA step, the 60 most discriminating features were retained.

We tested the following corrections: no correction, Muirhead correction, Karoui correction, Iterative Feedback correction, Two Subset correction and a lower bound correction. With the lower bound correction, we use the true covariance matrices of the synthetic data to calculate the actual variances along the estimated eigenvectors and use these values as $\hat{\lambda}_i$'s. We assumed this correction would give an indication of the best possible error reduction.

We generated a test set with 1000 identities. For each identity 10 enrollment samples and 10 probe samples were generated. During the experiment 3 configurations were tested: correction of only the within class eigenvalues, correction

of only the between class eigenvalues and correction of both the within and the between class eigenvalues. The DET curves of the three configurations are shown in Figure 2. In Figure 4a we show the relative EER improvement averaged over 5 repetitions.

The within class eigenvalues correction configuration shows a large difference between the no correction DET curve and the lower bound correction. Therefore the bias in the within class eigenvalues seems to have a large effect on the error rates. The Two Subset correction achieves on average slightly better results as the lower bound correction, but this is probably due to measurement noise. The performance of the Karoui correction fluctuates when the experiment is repeated. In some repetitions the Karoui correction reduces the error rates by half, but on average it increases the error rates as shown in Figure 4a.

The between class eigenvalues correction configuration shows hardly any difference between the different correction algorithms. It seems that the bias in the between class eigenvalues have little influence on the verification scores. The curve of both eigenvalue sets corrected shows no significant difference with the within only correction.

In Figure 3a and Figure 3b we show the corrected within class eigenvalues and between class eigenvalues respectively. The lower bound correction shows considerable fluctuations in the curve. This indicates that the ordering of the sample eigenvectors is wrong.

The lower bound curve is much flatter for the small eigenvalues in the within class correction than the no correction curve. The Two Subset correction also makes the curve much flatter for the smaller eigenvalues, although the eigenvalues are considerably larger than the lower bound correction. Considering the error rates are almost the same, the similarity in flatness seems more important than the actual value of the eigenvalues.

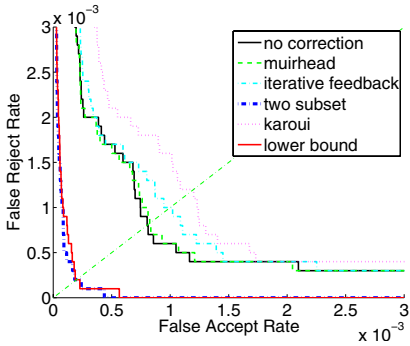
The Karoui correction shows a similar flatness until the 78th eigenvalue. After the 92th eigenvalue, all remaining eigenvalues are set to 0. This seems to have only a small effect on the error rates. This is remarkable since 0 within class variance would indicate very good features, while we know from the lower bound correction that the within class variance is non zero. However, if the between class variance is also zero, the direction will be neglected.

4.2 FRGC Facial Data Experiment

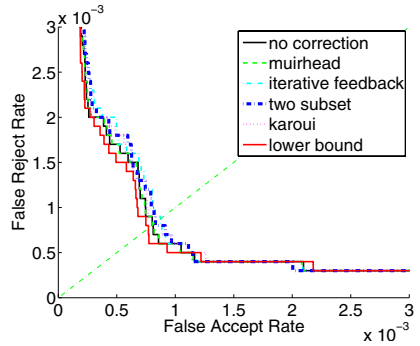
Eigenvalue correction with synthetic facial data caused a significant reduction of the error rates. In the next experiment we replaced the synthetic facial data with the face data set from the FRGC database. This data set is the set used in the previous experiment to determine the facial data structure.

The data set is split in a training set and a test set. The training set contained 70 randomly chosen identities, with a maximum of 5 samples per identity. The test set contained the remaining 445 identities. At most 5 samples per identity are used for enrolling, at least 1 sample is used as probe per identity.

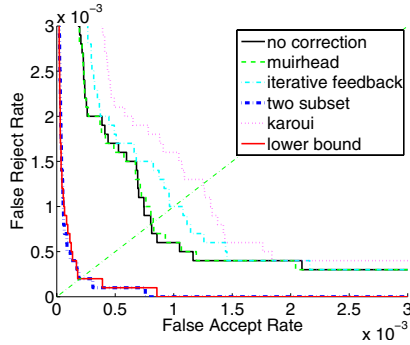
In the training stage instead of reducing the dimensionality to 150, as described in section 3, only the null space is removed. After correction of the



(a) Within eigenvalue correction only

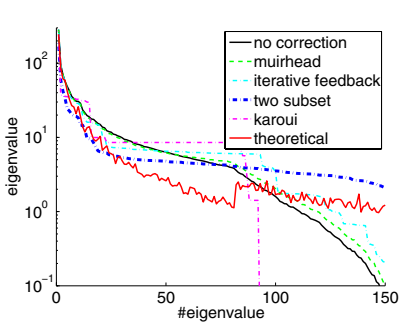


(b) Between eigenvalue correction only

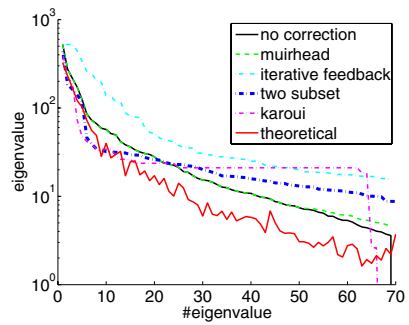


(c) Both within and between class eigenvalue correction

Fig. 2. DET curves for the synthetic data experiment



(a) Within class eigenvalues



(b) Between class eigenvalues

Fig. 3. Scree plots of the corrected eigenvalues of synthetic data

eigenvalues, the dimensionality is reduced to 150. The correction algorithms described in section 2.2 are compared.

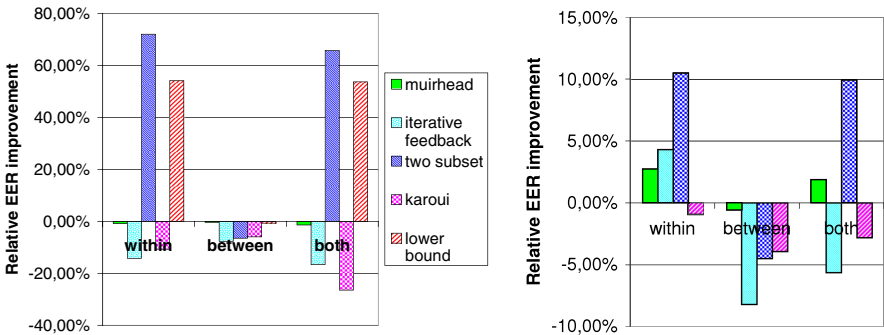
The experiment is repeated 5 times for the same 3 configurations as in the synthetic data experiment. For each correction algorithm in each configuration we determined the Equal Error Rate (EER). This EER is compared with the no correction EER. The average over 5 repetitions of the relative improvement of EER is shown in figure 4b.

The results show that correcting only the between class eigenvalues increases the EER for all correction algorithms. The within correction decreases the EER for most algorithms. Correcting both eigenvalue sets decreases the EER for the iterative feedback algorithm and the Two Subset algorithm. But this decrease in EER is less than the decrease in EER if only the within class eigenvalues are corrected.

Comparing the different correction methods shows that in the within correction and both eigenvalue sets correction the Two Subset correction performs considerably better than the other corrections. The Karoui correction always increases the EER.

In Figure 5 we show the results of the first repetition. The Karoui corrections sets a large set of small eigenvalues to zero. This had remarkably little effect on the error rates. The Two Subset correction on the other hand assigns non zero values to eigenvalues which were originally zero.

Most correction algorithms show a trend: the largest eigenvalues are reduced while the smaller eigenvalues are increased. This effect is the strongest with the Two Subset correction. Since this correction method achieved the lowest error rates, it seems that in face recognition indeed the largest eigenvalues are over estimated while the smallest are under estimated, at least in the within class estimation.



(a) synthetic facial data correction averaged over 5 repetitions (b) real facial data correction averaged over 5 repetitions

Fig. 4. Relative Equal Error Rate improvement for each correction method. There are three configurations: only within class eigenvalues correction, only between class eigenvalues correction and both eigenvalue sets correction.

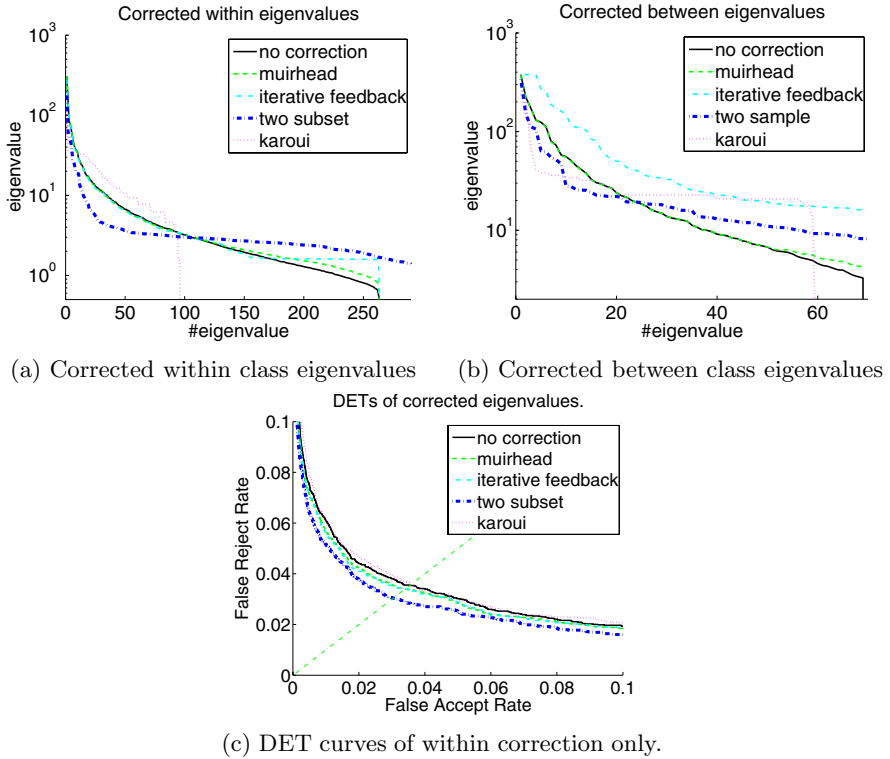


Fig. 5. Results of the first repetition of real facial data experiment

Comparing the results of the real facial data test with the results from the synthetic data shows that the EER's in real data are an order higher than the EER's in synthetic data. This suggests that the model we used is not sufficiently accurate for describing real facial data. However, in both experiments the Two Subset method showed the highest reduction in EER.

5 Conclusion

We showed that the GSA provides more accurate analysis of the sample eigenvalue estimator than LSA in biometrics: GSA on the estimator predicts that the estimates in biometrics will have a bias, which is observed in synthetic data, especially for the smaller eigenvalues.

Correcting only the within class eigenvalues has demonstrated the most effect. This is related to the previous conclusion: the best features are determined by the ratio of between class over within class variance. Therefore the best features probably lie in space spanned by the largest between class eigenvalues and the smallest within class eigenvalues. Since the smaller eigenvalues have more bias, within class correction has the most effect.

The Two Subset correction gave the best improvement of error rates in both the synthetic data experiment and the real facial data experiment. Although the performance of the correction was the same as the synthetic correction, the scree plots did differ. The corrections of the other algorithms did also significantly alter the eigenvalues, but this had little effect on the error rates for most of these corrections. Apparently the actual values of the eigenvalues do not have to be estimated very accurately.

The relative error reduction in the facial data is much lower as in the synthetic data by the Two Subset correction. Also the no correction error rates differ more than an order between the real facial data and the synthetic data. This suggests that the bias in the eigenvalue bias is only a moderate error factor in the real facial data.

References

1. Fukunaga, K.: Introduction to statistical pattern recognition, 2nd edn. Academic Press Professional, Inc., San Diego (1990)
2. Girko, V.: Theory of Random Determinants. Kluwer, Dordrecht (1990)
3. Muirhead, R.J.: Aspects of multivariate statistical theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., Chichester (1982)
4. Stein, C.: Lectures on the theory of estimation of many parameters. Journal of Mathematical Sciences 34(1), 1371–1403 (1986)
5. El Karoui, N.: Spectrum estimation for large dimensional covariance matrices using random matrix theory. ArXiv Mathematics e-prints (September 2006)
6. Silverstein, J.W.: Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivar. Anal. 55(2), 331–339 (1995)
7. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
8. Hendrikse, A.J., Spreeuwers, L.J., Veldhuis, R.N.J.: Eigenvalue correction results in face recognition. In: Twenty-ninth Symposium on Information Theory in the Benelux, pp. 27–35 (2008)