

The effect of position sources on estimated eigenvalues in intensity modeled data

Anne Hendrikse Raymond Veldhuis Luuk Spreeuwens
University of Twente
Fac. EEMCS, Signals and Systems Group
de veldmaat 10, 7522 NB Enschede, The Netherlands
a.j.hendrikse@ewi.utwente.nl

Abstract

In biometrics, often models are used in which the data distributions are approximated with normal distributions. In particular, the eigenface method models facial data as a mixture of fixed-position intensity signals with a normal distribution. The model parameters, a mean value and a covariance matrix, need to be estimated from a training set. Scree plots showing the eigenvalues of the estimated covariance matrices have two very typical characteristics when facial data is used: firstly, most of the curve can be approximated by a straight line on a double logarithmic plot, and secondly, if the number of samples used for the estimation is smaller than the dimensionality of these samples, using more samples for the estimation results in more intensity sources being estimated and a larger part of the scree plot curve is accurately modeled by a straight line.

One explanation for this behaviour is that the fixed-position intensity model is an inaccurate model of facial data. This is further supported by previous experiments in which synthetic data with the same second order statistics as facial data gives a much higher performance of biometric systems. We hypothesize that some of the sources in face data are better modeled as position sources, and therefore the fixed-position intensity sources model should be extended with position sources. Examples of features in the face which might change position between either images of different people or images of the same person are the eyes, the pupils within the eyes and the corners of the mouth.

We show experimentally that when data containing a limit number of position sources is used in a system based on the fixed-position intensity sources model, the resulting scree plots have similar characteristics as the scree plots of facial data, thus supporting our claim that facial data at least contains sources inaccurately modeled by the fixed position intensity sources model, and position sources might provide a better model for these sources.

1 Introduction

In biometrics the objective is to automate the task of either identifying a client or verifying a client's identity claim. This could be done by measuring some physical properties of the body and compare these measurements with enrolled templates. However, in practice the measurements are a mixture of signals, in which we distinguish three classes: a class of signals which are related with the body and contain discriminative information, a class of signals which can also be related with the body but contain no discriminative information and a noise signals class.

To illustrate these classes, consider face recognition. In face recognition, the physical properties are usually measured by taking photo's of the subjects face. In these images, examples of the first class of signals are the size of the mouth and the color of

the iris. Examples of the second class of signals are the position of the pupils or many of the changes in the face due to expression.

Given this setting, a major aspect of biometrics is to extract the signals containing information on the identity from the mixture of signals and noise and decode the identity information from these signals. To perform this task, a description of the data generating process is required, which both formulates a model of how the data is generated and gives the parameters of that model. But commonly the only information available for this task is a set of examples, denoted as the training set.

In several biometric modalities, for example face, the samples are images. A classical method used in face recognition is the eigenface method [13]: after some preprocessing the examples are arranged to column vectors. These samples are then modeled as being drawn from a random process, where it is assumed that the underlying distribution is reasonable approximated by a multivariate normal distribution, so the samples can be described by a random variable \mathbf{x} with a distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \mathcal{E}\{\mathbf{x}\}$, $\boldsymbol{\Sigma} = \mathcal{E}\left\{(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})\right\}$ and $\mathcal{E}\{\cdot\}$ is the expectancy operator.

The covariance matrix $\boldsymbol{\Sigma}$ can be decomposed via:

$$\boldsymbol{\Sigma} = \mathbf{E} \mathbf{D} \mathbf{E}^T \quad (1)$$

where \mathbf{E} is an orthogonal matrix, where every column is an eigenvector of $\boldsymbol{\Sigma}$, and \mathbf{D} is a diagonal matrix, with the eigenvalues corresponding to the eigenvectors in \mathbf{E} on the diagonal. The idea is that any sample, after the preprocessing, can be projected on these eigenvectors, which result in a set of uncorrelated signals.

The implicit model behind the eigenface method is that the eigenvectors can be transformed back to the original image space, where they identify the source signals in the face. This model assumes that face information is encoded as a weighted set of intensity sources at fixed positions in the face, the eigenfaces.

Usually the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and have to be estimated from the training set. From now on we will denote the parameters from the model as the population parameters and their estimates as sample parameters. Let \mathbf{X} denote a matrix where each column contains a training sample after preprocessing. Then an estimate of the mean can be obtained from the sample mean: $\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{k=1}^N \mathbf{X}_{:,k}$, where $\mathbf{X}_{:,k}$ denotes the k^{th} column of \mathbf{X} , and an estimate of the covariance matrix can be obtained from the sample covariance matrix:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{X}_{:,k} - \hat{\boldsymbol{\mu}})^T (\mathbf{X}_{:,k} - \hat{\boldsymbol{\mu}}) \quad (2)$$

The sample covariance matrix can be decomposed into sample eigenvectors and sample eigenvalues. The so-called scree plot of the sample eigenvalues has some remarkable properties as we will show in Section 2: if plotted in a double log plot, the curve becomes an almost straight line. Further more, if the number of samples is below the dimensionality of the samples, adding new samples results in an increase of the number of estimated intensity sources.

These characteristics are surprising because if the eigenface model is a good approximation and there is only a limit number of signals present, then we would expect the scree plot to show two groups of eigenvalues: a group with a fixed number of eigenvalues of relatively large value (although there may be a large variability within the group) and a group of very small eigenvalues. The scree plots, however, show no grouping of eigenvalues: the scree plots follow a smooth curve in which no separation point can be distinguished.

To give another explanation of these characteristics, we hypothesize that at least some of the sources are not accurately modeled by fixed position intensity sources,

but instead some of the information is actually encoded in the position of features. In Section 3 we demonstrate with a synthetic data experiment that position sources can explain the scree plot characteristics rather accurately.

Based on these observations we derive some conclusions in section 4 and discuss how biometric systems could be modified to take this effect into account.

2 Eigenvalue estimation with facial data

To illustrate the previously described characteristics of the eigenvalue scree plots of the eigenface method, we did an experiment with an eigenface system. As training data we used a subset of the FRGC2 database [11]. We selected, using the supplied meta data database, images containing faces with frontal view, neutral expression and no glasses, beards and moustaches. For the training set we used a standard training set selection procedure in face recognition: first the number of individuals to be in the training set is chosen, after which most of their samples are added to the training set. We chose 4 configurations: in the first configuration we used 7047 samples from 400 individuals, in the second 5386 samples from 300 individuals, in the third 1889 samples from 100 individuals and in the last 503 samples from 30 individuals. After some preprocessing the samples had a dimensionality of 8762, so all configurations were undersampled.

Figure 1 shows the scree plots corresponding to the different configurations, with both the horizontal axis (the eigenvalue index) and the vertical axis (the eigenvalue) with a logarithmic scale. A large part of most curves can be described by a straight line. The first 3 eigenvalues differ from this line. The last part also differs, but this can be some border effect (Note that the eigenvalues with an index higher than the number of samples are necessary zero valued, which drags the scree plot to minus infinity on a logarithmic scale). In [7] it was suggested to model the scree plot by

$$\lambda_k = \frac{\alpha}{k + \beta} \quad (3)$$

where α and β are two constants to be determined. But in our experiments such a model still gives a rather poor description of the first few eigenvalues and so we only model the straight part, which is better described by

$$\lambda_k = 10^b \cdot k^a \quad (4)$$

where a is the slope coefficient and b the offset of the line parts in the log log scree plots. The constant a is close to -1 , which is in accordance with the results of [10] which named this model the 1 over f model.

Figure 1 shows a second remarkable characteristic when the number of samples is increased. As long as the training set is undersampled, it is to be expected that every new sample introduces a new source, if the number of (noise) sources is larger than the number samples, but what is remarkable is that with the increase of the samples, a larger part of the of the plot follows the log log relation.

One of the problems with facial data is that the number of samples is of the same order as the dimensionality of the samples, which results in a bias of the sample eigenvalues (see [9] or [12] for example). In [3], Karoui performed several eigenvalue estimation experiments, where one of the experiments resembled our expected set of eigenvalues as described in the introduction: in the experiment half of the population eigenvalues are equal to 2 and the other half equal to 1. In the corresponding sample eigenvalues a large spread occurred where no clear separation point is visible in the scree plot, which is similar in our facial data scree plot. Note as well that the experiments by Karoui were done with 5 times as many samples as dimensions, so in our facial data eigenvalue estimation the spreading effect should be stronger. Also the Marčenko Pastur rule ([9],

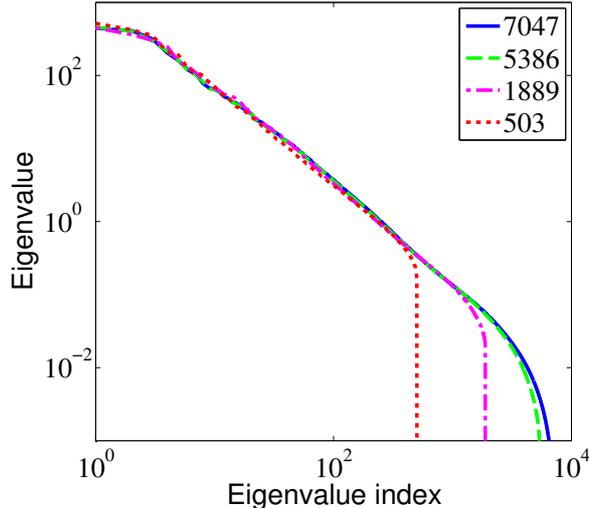


Figure 1: Example of the behaviour of facial data scree plots if plotted with both axis logarithmic. The legend shows the number of samples used to estimate a scree plot.

[1] and [8]) describes a large set of very small sample eigenvalues and a wide spread among the larger ones for close to undersampled eigenvalue estimation problems. However in our experiments, only the smaller eigenvalues were changed by bias correction methods, so bias does not (fully) explain the characteristics described before.

As stated earlier the implicit model in the eigenface method is that the information is encoded in a weighted combination of fixed position intensity sources. If we combine the characteristics of the scree plots with this model it would mean that there is a large number of sources, as adding new sample introduces new sources and there is no clear distinction between the two signal classes related to physical properties of the body and the noise class as they are described accurately by equation 4. The bias did not fully explain the characteristics of the scree plots, so there should be another explanation.

It seems plausible that the model assumed is inaccurate in modeling face data. This is also supported by our previous observations that if synthetic data is generated with the parameters as estimated from the facial data, and this data is used in a biometric verification system, the performance of these systems increases considerably. In [6] we used a very small training set, but still the error rates were a factor 10 lower than with real facial data.

To explain the characteristics of the scree plots, we hypothesize that instead of having only fixed position sources, at least some of the sources encode their information in position of features instead of in intensities of pixels. Practical examples of position sources in face data are the position of the eyes, the position of the pupils within the eyes, corners of the mouth and so on. As we will demonstrate later on, a very limited number of position sources can generate scree plots with similar characteristics as we have seen with facial data.

3 Eigenvalue estimation with position sources

We demonstrate the effect of position sources in the eigenface method with synthetic data, where features are positioned randomly in the image. As feature we use a black square while the remainder of the image is kept white. The coordinates of the square are determined by drawing samples from normal distributions. If a pixel is only partially covered by a black square, its intensity is determined by bilinear interpolation. We performed two experiments: a low resolution and a high resolution experiment.

In the low resolution experiment we generated images of 10 pixels height and width. The square has a height and a width of 1 pixel. The horizontal position and the vertical position are both drawn from a distribution $\mathcal{N}(5.5, 2.5^2)$. An example of a synthetic image is given in figure 2.

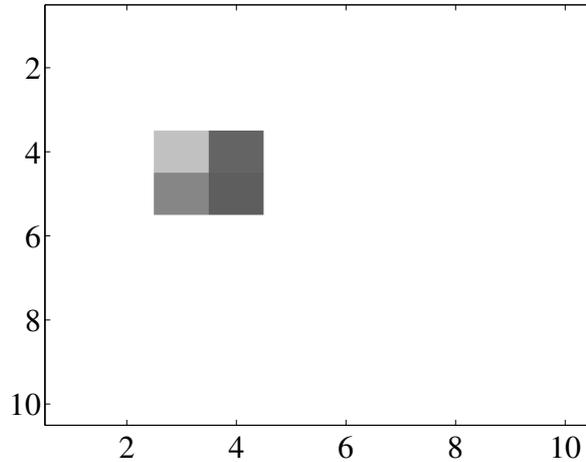


Figure 2: Example of a low resolution sample with interpolation.

We generated a set of 1000 training samples and estimated the eigenvalues resulting from the eigenface method. The resulting scree plot is given in figure 3. The scree plot is plotted with only the horizontal axis logarithmic, instead of both horizontal and vertical axis logarithmic, like in Figure 1. So clearly, a single position feature can not explain the characteristics of the scree plot found when using facial data, but there is a similarity in some of the characteristics: firstly, the number of sources appears to be large, since the number of non zero eigenvalues equals the dimensionality of the samples (note that we have more samples than the number of dimensions, so no undersampling). Secondly, the scree plot of the synthetic data is also accurately described by a very smooth function, although different from the facial data scree plot. Because this smooth function is a straight line on a logarithmic plot, again no separation point can be identified, so no distinction between source classes can be made.

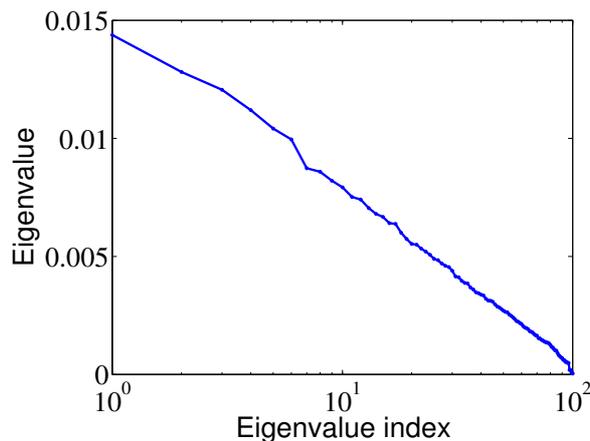


Figure 3: Log log screeplot of the synthetic square low resolution data.

To explain how a position source results in a large number of estimated intensity

sources, consider figure 4. In the figure we consider two pixels (the squares x_1 and x_2) from an image containing a 1 by 1 pixel wide black square. The black square is moved from left to right over the image. If its position is not exactly the same as one of the pixels, the intensity of the pixels is determined by linear interpolation. The vertical position of the black square is equal to the two pixels considered and the horizontal position of the square changes from a position before the two pixels to a position after the two pixels. The curve on the right shows a plot of the intensity of pixel x_1 versus the intensity of pixel x_2 . At the start, both pixels are empty, so we start at position 1. As the square moves to the right, it starts to overlap pixel x_1 which results via linear interpolation in an increase of intensity of pixel x_1 , so we move from position 1 to position 2. After the black square overlaps pixel x_1 completely, its overlap reduces with x_1 and it start to overlap x_2 , so we move to position 3 and so on.

Note that in the end we end up in position 1 again. So even though the original source has a completely different value, in the curve it ends up at the same position. This is a severe problem for classification systems that assume that samples from different classes form separate clusters (for example LDA, described in [4]): clusters which are far apart in position space, are mapped close together in intensity space.

Note as well that the curve forms a triangle in the 2D intensity space. This leads to 2 non zero eigenvalue estimates if the covariance matrix of the images is determined, so we estimate 2 intensity sources, even though there is only one position source. The fixed position intensity model seems therefore ill equipped for modeling this kind of data. However, we have considered a small object with large position variations. Note that if the objects movements are limited, so that the intensity curve stays between positions 1 and 2 of figure 4, then position variations are translated to intensity variations, so for larger objects with low frequency textures and low position variations, the intensity model may still be adequate.

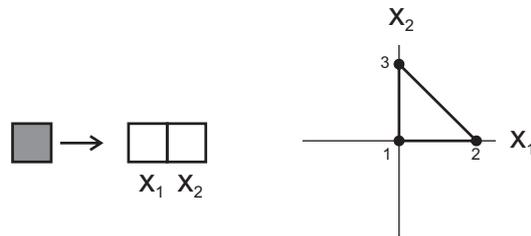


Figure 4: Example showing that a single position source can create multiple sources if the data is intensity modeled. The position source is univariate. However, if the manifold is drawn in 2D intensity space, the intensity shows a triangle as path. This has two directions of variations, thus gives two non zero eigenvalues.

In the low resolution experiment we used images of 10 by 10 pixels, which results in samples with a dimensionality of 100, far less than the 8762 dimensions of the facial data. Therefore we repeated the experiment, but now with images with width and height of 80 pixels and we used 5000 samples for training. We increased the square size to 20 by 20 pixels and the number of squares to 4. The mean positions of the squares are distributed equally over the diagonal of the images from the top left corner to the lower right corner. The position of the squares is distributed normally around these means with a variance of 6^2 in both the horizontal and vertical direction.

The resulting eigenvalue estimates are represented in the scree plot in figure 5. If we compare the curve with figure 1 we see a similar shape, although the curve of the synthetic data does not match a straight line as close as the real face data. However, the gradient of the part that fits the straight line is close to -1, similar to the facial data scree plot.

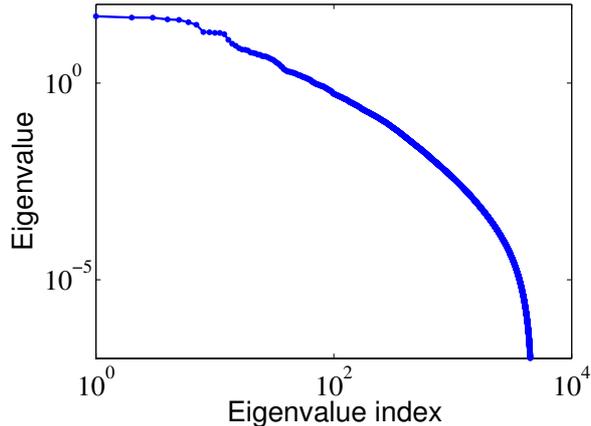


Figure 5: Log log screeplot of the synthetic square high resolution data.

4 Conclusion and discussion

We considered the scree plot of eigenvalues in face data and showed that it has several remarkable characteristics: firstly, a large part of the curve follows a 1 over f model, as was already suggested by others. Secondly adding samples to an undersampled training set leads to the estimation of additional intensity sources, and it extends the part of the curve which matches the 1 over f model.

While these observations could partially be explained by a complicated configuration of intensity sources, we consider it more likely that the implicit model of fixed position intensity sources is not an accurate model of facial data and propose a model that assumes that some of the information is encoded in the position of features. The scree plot of synthetic data with the signals of 8 sources positioning 4 squares showed similar characteristics as the scree plot of facial data.

If some of the information is indeed encoded in the position of features, it has some implications for biometric systems. For example, in face recognition, if the resolution of the face images is increased, the effect of position sources is likely to become worse, for position changes of large size objects with low frequency textures may still be accurately modeled by fixed position intensity sources, but for smaller objects or objects with high frequency textures this approximation is no longer possible. In [2] it was discovered that there is an optimum resolution for an eigenface system, which might be explained by our findings. A solution to use higher resolution images may therefore lie in detecting moving features first as is done for example in [14] and then reduce their influence in the eigenface approach.

The effect of position sources in eigenface model based systems may also explain some of the results in bias correction experiments. In eigenvalue estimation, the largest eigenvalues tend to be overestimated and the smallest eigenvalues tend to be underestimated (see [5], [7] or [3] for example). Bias correction will therefore increase the smallest sample eigenvalues and decrease the largest sample eigenvalues. If small objects with large position changes are present in the image, the object will divide its signal power over many small intensity sources as shown in section 3. Therefore bias correction might increase the influence of sources incorrectly modeled by the eigenface model, causing a deterioration instead of improving performance. This will be a subject of our future studies.

References

- [1] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. In *Statistica Sinica*, 9, pages 611–677. National University of Singapore, 1999.
- [2] B. J. Boom, G. M. Beumer, L. J. Spreeuwers and R. N. J. Veldhuis. The effect of image resolution on the performance of a face recognition system. In *Proceedings of the Ninth International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, Malaysia*, pages 409–414. December 2006. ISBN 1-4244-0342-1.
- [3] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *ArXiv Mathematics e-prints*, september 2006.
- [4] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990. ISBN 0-12-269851-7.
- [5] A. Hendrikse, L. Spreeuwers and R. Veldhuis. A bootstrap approach to eigenvalue correction. In *International Conference on Data Mining*, page 1. 2009.
- [6] A. J. Hendrikse, L. J. Spreeuwers and R. N. J. Veldhuis. Eigenvalue correction results in face recognition. In *Twenty-ninth Symposium on Information Theory in the Benelux*, pages 27–35. 2008.
- [7] X. Jiang, B. Mandal and A. Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):383–394, 2008. ISSN 0162-8828.
- [8] I. M. Johnstone. On the distribution of the largest principle component. Technical report, Dep. of Statistics, Stanford University, 2000.
- [9] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR - Sbornik*, 1(4):457–483, 1967.
- [10] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:696–710, 1997. ISSN 0162-8828.
- [11] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. Overview of the face recognition grand challenge. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 947–954. IEEE Computer Society, Washington, DC, USA, 2005. ISBN 0-7695-2372-2.
- [12] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivar. Anal.*, 55(2):331–339, 1995. ISSN 0047-259X.
- [13] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991. ISSN 0898-929X.
- [14] V. Vezhnevets and A. Degtiareva. Robust and accurate eye contour extraction. In *Proc. Graphicon-2003*, pages 81–84. 2003.