On the Impact of Outliers on High-Dimensional Data Analysis Methods for Face Recognition

Sid-Ahmed Berrani
France Telecom R&D – TECH/IRIS
35512 Cesson Sévigné Cedex, France
SidAhmed.Berrani@francetelecom.com

Christophe Garcia
France Telecom R&D – TECH/IRIS
35512 Cesson Sévigné Cedex, France
Christophe.Garcia@francetelecom.com

ABSTRACT
In this paper, the impact of outliers on the performance of high-dimensional data analysis methods is studied in the context of face recognition. Most of the existing face recognition methods are based on PCA-like methods: Faces are projected into a lower dimensional space in which similarity between faces is more easily evaluated. These methods are, however, very sensitive to the quality of face images used in the training and the recognition phases. Their performance significantly degrades when faces are not well centered or taken under variable illumination conditions. In this paper, we study this phenomenon for two face recognition methods (PCA and LDA2D) and we propose a filtering process that allows an automatic isolation of noisy faces which are responsible for the performance degradation. This process is performed during the training phase as well as the recognition phase. It is based on the recently proposed robust high-dimensional data analysis method RobPCA. Experiments show that this filtering process improves the recognition rate by 10 to 20%.

Keywords
High-dimensional data analysis, outliers, face recognition, dimensionality curse.

1. INTRODUCTION
High-dimensional data analysis is a very important research topic for different domains (indexing, data mining, pattern recognition, etc.). The aim is to develop methods that can extract knowledge and explore high-dimensional datasets.

Depending on the domain, the subsequent effect of increasing the dimension can be an exponential growth of computing time, an important degradation of precision or a greater sensitivity to outliers. All these problems are usually referred to as the dimensionality curse phenomenon, an expression introduced by Bellman in [3]. Many papers have studied this phenomenon with respect to specific domains: machine learning [10], similarity searches [1, 4], neural networks [11]...

In this paper, we study a specific aspect of this problem in the context of face recognition. In particular, we deal with the sensitivity to outliers of the high-dimensional data analysis methods used in face recognition. These methods are used to reduce the dimension of feature vectors associated with faces. The objective is to project face images into a lower dimensional space, while keeping most of the information expressed by the images. The similarity between faces is then evaluated in this projection space. Examples of such methods are Principal component analysis (PCA) [9], PCA2D [13], Linear Discriminant Analysis (LDA) and LDA2D [12].

The first contribution of this paper is to study the sensitivity of these methods when faces have not been very precisely cropped, are not well-centered or come from images taken under variable illumination conditions. We will study this problem from a statistical point of view and we will show that “noisy” face images could be viewed as outliers in feature space.

The second contribution of the paper is a filtering method which is able to automatically isolate face images which are responsible for the performance degradation of the face recognition methods. The idea is to filter out feature vectors that affect the accuracy of internal statistical entities used in PCA-like methods (typically the mean and the covariance matrix). To achieve that, we propose to use robust statistics and in particular RobPCA [7].

Throughout this paper, we will illustrate the introduced concepts using two face databases and two statistical methods for face recognition (PCA and LDA2D).

The rest of the paper is organized as follows. Section 1 gives a brief overview of face recognition methods and focuses on statistical approaches. Section 2 studies the sensitivity of statistical approaches to outliers, explains its origin and presents two experiments corroborating the analysis. The proposed solution for isolating noisy face images is presented in Section 3 and the experimental results showing its effectiveness are presented in section 4. Section 5 concludes the paper and proposes future extensions.

2. THE FACE RECOGNITION PROBLEM
The face recognition problem has been extensively studied due to its importance in different domains (biometrics, video surveillances, multimedia indexing, etc.). The basic objec-
tive of a face recognition system is to identify an unknown face using a database of faces. A detailed survey of existing methods is given in [14]. In this section, we will limit our description to the class we are concerned with, that is, the statistical methods.

The most famous statistical method for face recognition is the method known as eigenfaces [9]. The principle of this method is to make use of the statistical properties of feature vectors associated with faces to compute a projection space. Faces are projected in this space and their similarity is evaluated as an Euclidean distance. In this method, the feature vector is the vector obtained by concatenating the rows or the columns of the face image. The projection space is the space defined by the eigenvectors of the covariance matrix of the feature vectors. The identification of an unknown face is achieved by finding the face in the database whose projection vector is the closest to the projection vector of the unknown face (nearest-neighbor classification).

Another way to perform face recognition when a set of faces is available for each person stored in the database, is to use Linear Discriminant Analysis (LDA), also known as Fisher Discriminant Analysis. In this case, the projection space is built so that the classes are as compact as possible and their centers are as far away from each others as possible (here, a class refers to the set of feature vectors extracted from the face images of a person). LDA is hence carried out via a scatter matrix analysis. Once again, the identification of an unknown face is usually performed via a nearest-neighbor classification.

Extensions of the PCA and LDA have been recently proposed respectively in [13] (PCA2D) and [12] (LDA2D). These extensions use the same statistical principles as PCA and LDA. The main difference with respect to the original methods is that faces are represented by matrices, that is, each face is not transformed into a vector but the whole matrix associated with the face image is projected. The advantages are an important gain in storage, a better numerical stability and an increased recognition rate.

In this paper, we perform experiments using the PCA and LDA2D methods.

3. IMPACT OF OUTLIERS

As introduced in the previous section, statistical methods are mainly based on the analysis of the covariance matrix of feature vectors. They are therefore likely to be sensitive to outliers as they are mainly based on the first and second order moments. For example, in the case of the PCA, only the few first components are kept. These are supposed to encode most of the information expressed by the data. They correspond to directions within which the variance is at its maximum. However, if the dataset contains too many noisy vectors, the first components will encode only the variation due to the noise and not the variation containing the necessary information to differentiate faces (e.g. attribute and shape of the face).

In the case of LDA, in addition to the impact of outliers on the scatter matrices involved in the computation of the projected space, noisy vectors introduce overlapping between classes, and therefore, reduce the performance of the classifier.

One of the main objectives of face recognition methods is to deal with variations that might affect the face images. These variations could be due to the acquisition conditions or to the changes in expression. A face recognition method is therefore supposed to recognize a face despite these variations. In practice, this is far from being the case. The presence of noisy face images significantly decreases the performance of face recognition methods. We consider that a face image is noisy if it presents a variation that the face recognition method cannot handle. In this paper, we propose to classify these variations into three categories:

1. Important illumination changes,
2. Imprecise face cropping,
3. Non-frontal poses.

Some face images presenting these kinds of variations are shown in figure 1. Face images from the second and the third category arise in general from an automatic extraction of faces from still images or video sequences. They are due to the imprecision inherent to face detection methods.

To illustrate the impact of these variations on statistical methods for face recognition, we have evaluated the recognition rates using PCA and LDA2D on two databases: PF01 to study the impact of important illumination changes and FDB15 for imprecise face cropping and non-frontal poses.

PF01 is a database of 107 people containing 17 different views per person. Among the 17 views, there are four images with important illumination changes (e.g. the two first images of the first row of figure 1). Two of them have been used for training and the two others for testing. The 13 remaining images have been divided into two subsets: 9 for training and 4 for testing. The recognition rates have been evaluated twice: The first time using all the images and the second time after having removed the four noisy images of each person from the training and the test sets. The obtained results are shown in figure 2.(a) for PCA and in figure 2.(b) for LDA2D.

FDB15 contains 15 people. 21 face images per person have been used for training. Among them, there are 6 non-centered or non-frontal face images. 5 well-framed face images and 5 noisy face images per person have been used for

PF01 is available following the URL:
http://nova.postech.ac.kr/.

FDB15 is a database we have created specially for the needs of this paper. These two databases are further described in section 5.
evaluation. As for the PF01 database, a second evaluation has been carried out without including noisy images. The obtained results are shown in figure 2.(c) for PCA and in figure 2.(d) for LDA2D.

Figure 2 clearly shows that the recognition rates of LDA2D and PCA decrease by about 10% when all the faces including noisy ones are used. We can also notice that this impact is independent of the number of components retained during the analysis. We recall that this number of components corresponds to the dimension of the projection space in the case of the PCA, and the number of the row oriented discriminant (ROD) components in the case of LDA2D. The dimension of the LDA2D projection space is hence the number of ROD components multiplied by the height of the face images.

On the other hand, if we study the three first eigenvalues from PCA, we notice that the ratio of their sum w.r.t the sum of all the eigenvalues is much more important when noisy faces are included during the analysis. We recall that this ratio corresponds to the proportion of the variation expressed by the corresponding principal component.

Table 1 gives the ratio of the first three eigenvalues w.r.t. the sum of all eigenvalues for PF01 and FDB15. This table clearly shows that, despite their relatively small number, noisy faces in the training sets significantly modify PCA results, and hence decrease the recognition rates as they affect the discriminant information encoded by the first principal components.
Table 1: Ratio of the first three eigenvalues w.r.t. the sum of all eigenvalues.

<table>
<thead>
<tr>
<th></th>
<th>PF01</th>
<th>FDB15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using all faces</td>
<td>35.51</td>
<td>35.02</td>
</tr>
<tr>
<td>Without noisy faces</td>
<td>25.68</td>
<td>29.13</td>
</tr>
</tbody>
</table>

4. THE PROPOSED SOLUTION

In the previous section, we have shown that a few noisy face images can reduce the performance of statistical face recognition methods. A way to overcome this problem is to filter out these noisy face images. This approach is of course well-adapted to the case where recognition is performed using video sequences: many faces are available for each person. It is, however, also an interesting approach even if the recognition is performed using still images as it allows the recognition system to avoid answering if the query face is of poor quality.

A first solution for performing this filtering process would be to analyze the face image content and to find a way to distinguish between a noisy face image and a well-framed one. This solution requires a fine analysis of the image content and more importantly, an inventory of all of the characteristics of noisy face images. Therefore, this solution is very difficult to use in practice.

It is however more interesting to consider the problem from a statistical point of view. If we assume that the proportion of noisy face images is relatively low, we can consider the problem as an outlier detection problem: A face is considered as noisy if the associated feature vector is isolated as an outlier.

The outlier detection problem has been very deeply studied and applied in a wide range of applications. The interested reader can refer to [6] for a complete description of the problem, its applications and the state of the art of existing methods. In this section, we only give a brief overview of the most recent outlier detection methods. These methods can be divided into three classes: (1) The statistical methods that use vector distribution models to characterize the proximity or the belonging of a vector to a set of vectors; (2) The neural network-based methods that use neural training to isolate outliers and (3) Machine learning methods that make use of decision trees and clustering techniques and that, unlike the two first classes, can handle categorical data.

In our case, no prior knowledge of the data is available. Therefore, methods that use supervised learning or a predefined model of data cannot be applied. Moreover, the notion of outliers is slightly different from the common definition\(^2\): Our main concern is the isolation of vectors that affect the accuracy of the first and second order moments involved in the process of face recognition.

We therefore propose the use of the RobPCA method introduced by Hubert et al. [7]. It is a statistical method that aims at performing a robust principal component analysis, i.e. finding principal components that are not too much influenced by outliers. It also provides a useful method to flag outliers. The main idea of RobPCA is to find the subset of vectors that allows to compute reliable basic statistics. These are then used to compute robust principal components and to isolate outliers.

4.1 Outlier detection using RobPCA

RobPCA [7] combines the ideas of two different approaches for robust estimation of principal components. The first approach aims at finding a subset of vectors whose covariance matrix has the smallest determinant, that is, the most compact in the space. The mean and the covariance matrix are computed on this subset. The second approach uses Projection Pursuit techniques. The idea is to maximize a robust measure of spread to sequentially find the principal axes.

To estimate the robust mean ($\hat{\mu}$) and the robust covariance matrix ($\hat{C}$) of a dataset $X_{n,d}$ of $n$ $d$-dimensional vectors, the RobPCA proceeds in three steps:

1. Data vectors are processed using a classical PCA. The objective is not to reduce the dimension but only to remove superfluous dimensions.

2. The $h$ “least outlying” vectors are searched, where $h < n$ and $h = n$ is the maximum expected number of outliers. To do that, a measure of “outlyingness” is used. This measure is computed by projecting all the vectors on a set of lines and by measuring the degree of “outlyingness” of each vector w.r.t. the spread of projections. A PCA is then performed on the found $h$ vectors and the dimension is reduced.

3. The final $\hat{\mu}$ and $\hat{C}$ are estimated using an MCD estimator, i.e. based on the $h$ vectors whose covariance matrix has the smallest determinant. To find these vectors, a FAST-MCD algorithm [8] is used. The principle of FAST-MCD is to draw a set of random subsets and to refine them iteratively:
   - Compute the mean ($m$) and covariance matrix ($C$) of the $h$ vectors,
   - Compute the $C$-Mahalanobis distances of all the vectors to $m$,
   - Choose a new set composed of the $h$ vectors with the smallest Mahalanobis distances. The determinant of the covariance matrix of these new $h$ vectors is smaller than the determinant of $C$.

This procedure is repeated until convergence, i.e. no further improvements are obtained.

Once $\hat{\mu}$ and $\hat{C}$ have been estimated, the vectors are projected into a lower dimensional space defined by the eigenvector of $\hat{C}$. Let $Y_{n,k}$ be the new data matrix:

$$Y_{n,k} = (X_{n,d} - l_d \hat{\mu}) P_{d,k},$$

where $l_d$ is a $d$-dimensional vector of all components equal to 1 and $P_{d,k}$ is the projection matrix. $P_{d,k}$ is computed from a spectral decomposition of $\hat{C}$:

$$\hat{C} = P_{d,k} L_{d,k} P_{d,k}^T,$$

where $L_{d,k}$ is the diagonal matrix of eigenvalues $l_1, ..., l_k$.

\(^2\)Definition (outlier): An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs [2].
The outliers are then determined by analyzing the distribution of the two following distances (computed for the vector $i$):

$$D_1 = \sum_{j=1}^{k} \frac{y_{ij}^2}{l_j},$$

(3)

and

$$D_2 = ||x_i - \hat{\mu} - P_{d,k} y_i||.$$  

(4)

The first distance is the distance to the robust center of the vectors. It evaluates the proximity of $x_i$ to the cloud of vectors in the projection space. The second distance is the orthogonal distance to the projection space. Two thresholds are then derived from the distribution of these distances. If a vector has at least one of the two distances greater than the associated threshold then it is considered as an outlier.

The distribution of $D_1$ can be approximated by a $\chi^2_k$ distribution because it is a Mahalanobis distance of normal vectors. Therefore, the associated threshold can merely be $\sqrt{\chi^2_{k,0.975}}$ for example. However, the distribution of $D_2$ is not exactly known. Therefore, we use the approximation proposed in [7], i.e. $D_2$ to the power of 2/3 is approximately normally distributed. The associated threshold is hence $(m + \sigma z_{0.975})^{2/3}$, where $m$ and $\sigma$ are respectively the robust estimations of the mean and the standard deviation and $z_{0.975}$ is the 97.50% quantile of the normal distribution.

### 4.2 Application to face recognition

Filtering out noisy faces is needed twice in the whole recognition process: (1) off-line, to filter out the face images of the training set, and (2) on-line, to decide whether to perform recognition with a query face or not.

In order to filter face images of the training set, all that is needed is to transform these images into vectors and apply the outlier detection procedure described in the section 4.1. This can be done simply by concatenating the rows or the columns of each face image. However, as we are only concerned with the head pose and the global illumination of the face in the image, we suggest reducing the resolution of the images. This allows the acceleration of the filtering process and avoids taking small details of the images into account.

This filtering procedure is applied per person if there are enough face images per person. It can also consider all of the face images in the training dataset simultaneously. The only important condition is to make sure that the set of images contains a majority of well-framed face images. This condition is inherent to the principle of our approach which considers noisy face images as outliers, and therefore makes the assumption that noisy face images are less frequent than well-framed face images.

However, the situation is different for query faces as we cannot control their quality and their number. The proposed solution is as follows:

- Insert all or a subset of the query faces into a set of well-framed face images from the training set,

- Apply the filtering procedure on that set and only keep query faces that have not been isolated as outliers.

Once again, the only condition that needs to be fulfilled is to have a majority of well-framed images in the set in which query faces are inserted.
5. EXPERIMENTAL RESULTS

This section presents some experimental results that aim at assessing the proposed method. This evaluation has been conducted on the two previously introduced databases, PF01 and FDB15:

- The Asian Face Image Database PF01\(^3\) contains 17 different views of each person: 1 normal face, 4 illumination variations, 8 pose variations and 4 expression variations (In section 3, we have focused on the illumination variations).

- FDB15 is a database of 15 people. It has been created by automatically extracting faces from video sequences using the CFF face detector [5]. In the training set, we used 15 well-framed images and 6 noisy faces per person. In the test set, we used 5 well-framed images and 5 noisy faces per person. In this database, the considered variations concern the position of the face in the image and the head pose. Training and test sets have been created manually.

The size of face images of the two databases has been set to 65 x 75. However, it has been reduced to 32 x 37 during the filtering process.

The recognition evaluation has been performed using the PCA and the LDA2D. For each one, we have varied the number of components kept when computing the projection space.

The filtering procedure of noisy face images has been carried out separately on each person for the training sets. The selection rates of face images is presented in table 2 for the training and testing sets. We have conducted an evaluation on the two previously introduced databases, PF01 and FDB15:

- PF01 contains 15 faces per person, and FDB15 contains 15 faces per person.

The training set contains 15 well-framed images and 6 noisy faces per person. In the test set, we used 5 well-framed images and 5 noisy faces per person. In this database, the considered variations concern the position of the face in the image and the head pose. Training and test sets have been created manually.

The size of face images of the two databases has been set to 65 x 75. However, it has been reduced to 32 x 37 during the filtering process.

The recognition evaluation has been performed using the PCA and the LDA2D. For each one, we have varied the number of components kept when computing the projection space.

The filtering procedure of noisy face images has been carried out separately on each person for the training sets. The selection rates of face images is presented in table 2 for the training sets and also for the test sets.

<table>
<thead>
<tr>
<th></th>
<th>PF01</th>
<th>FDB15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>65.25</td>
<td>63.81</td>
</tr>
<tr>
<td>Test</td>
<td>40.03</td>
<td>36.00</td>
</tr>
</tbody>
</table>

Table 2: The selection rates of face images (%).

To illustrate the filtering procedure, we present a set of face images in figure 3 that have been automatically extracted from a video sequence. The number of face images we considered is 21, among them there are 6 noisy images. These images correspond to the training images associated with a person stored in the FDB15 database.

The 7 images in the right side of the figure are those that had been isolated as noisy. This example clearly shows that the filtering method is able to isolate atypical face images, among which are the 6 non-frontal and not well-centered face images.

To assess the ability of this procedure to improve the recognition rates, we have then evaluated the recognition rates considering only the selected face images (for both training and testing) and we have compared the obtained results with those obtained without filtering. The results are summarized in figure 4.

We can notice that, overall, the recognition rates have been improved by the filtering procedure by 10 to 20%. What is important to notice is that these improvements are greater than the improvements shown in figure 2. We recall that the results reported in figure 2 have been obtained after a manual filtering of noisy faces. The reason is that in the experiment presented in section 3, the selection concerned only some specific variations. For example, we have filtered out only faces presenting important illumination changes from PF01. The proposed filtering procedure has however isolated all the noisy faces including those presenting important expression changes or a non-frontal poses. This shows that the proposed filtering method is able to handle different changes simultaneously.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have shown the sensitivity to outliers of statistical methods for face recognition. We have analyzed this problem and proposed a filtering method to isolate noisy faces that are responsible for performance degradation.

In our future work, we will extend this study to other kinds of variations (e.g. other kinds of illumination changes) and other face recognition methods (e.g. based on neural networks). We will also consider a way of handling the noisy face images that are rejected in the proposed solution. They can be used, when classified and clustered, to develop multiple specific projection spaces for recognition.

7. REFERENCES


\(^3\)Available following the URL: http://nova.postech.ac.kr/
Figure 4: Recognition capabilities of PCA and LDA2D with and without an automatic selection of face images.


