

Better than best: matching score based face registration

Luuk Spreeuwers

University of Twente

Fac. EEMCS, Signals and Systems Group

Hogekamp Building, 7522 NB Enschede

The Netherlands

`l.j.spreeuwers@ewi.utwente.nl`

Bas Boom

University of Twente

Fac. EEMCS, Signals and Systems Group

Hogekamp Building, 7522 NB Enschede

The Netherlands

`b.j.boom@ewi.utwente.nl`

Raymond Veldhuis

University of Twente

Fac. EEMCS, Signals and Systems Group

Hogekamp Building, 7522 NB Enschede

The Netherlands

`r.n.j.veldhuis@ewi.utwente.nl`

Abstract

For most face recognition systems, proper registration of the faces to a common coordinate system is of great importance to obtain acceptable performance. Generally, for this purpose, landmarks are selected that can be located reliably, like the centres of the eyes, the tip of the nose and the centre or corners of the mouth. Many of the published results of face recognition methods are based on registration using manually selected landmarks, which is often regarded as the "gold standard". In this paper we show that using a matching score based registration approach, we can significantly improve upon face identification and verification results obtained using registration with manual landmarks. For recognition using PCA/Mahalanobis Cosine, we obtained up to 9% improvement in identification rate on the FERET data base.

1 Introduction

In most face recognition systems the following stages can be distinguished:

1. Preprocessing
2. Feature extraction
3. Calculation of a matching score
4. Classification

The first step, preprocessing, generally consists of pose normalisation or registration, some form of normalisation of illumination and selection of a region of interest (ROI). The registration casts a facial image into a reference coordinate system by rotating, scaling and translating the original image. Once the face is in a reference coordinate system, a ROI is defined to exclude the background. Typically, an elliptical ROI is used that contains the face from forehead to chin and from cheek to cheek. The second step, feature extraction, is used to obtain a meaningful and compact description of a facial image. Well known methods are PCA [7] (principle component analysis),

PCA+LDA [5, 2](linear discriminant analysis), LBP [1] (local binary patterns) and wavelets in EBG_M (elastic bunch graph matching) [8]. The third step, calculation of a matching score, is used to compare the features of a probe image with those of a reference image. The outcome can be a distance measure, e.g. Euclidian distance or the Mahalanobis distance, or a similarity measure like the likelihood ratio [2]. The final step is the classification, which is done by e.g. thresholding the matching score to determine if it is a match or by e.g. selecting the best score of all reference individuals in a list.

In order to be able to compare face recognition systems, often a fixed method is chosen for the preprocessing, based on registration using manually located landmarks. On the one hand this allows one to compare only the recognition part of the systems independently of the quality of the preprocessing stage. On the other hand, if the landmarks are located carefully, accurate registration is possible and the results obtained with such a registration method might be considered an upper-bound of what is attainable with a certain face recognition system. However, the accuracy of the manually located landmarks cannot easily be assessed and the influence of these errors on the recognition results is unclear.

The basic idea of matching score based registration is straightforward: instead of determining the registration parameters (rotation, translation and scaling) using landmarks, they are determined by maximising or minimising an object function. For matching score based registration the object function is the matching score of the face recognition system. Our first results on matching based registration were published in [4], where significant improvements in verification rates are reported relative to an automatic landmark based registration approach [3].

The objectives of this paper are to:

- investigate if matching score based registration results in improvement in recognition rates for both verification as well as identification relative to landmark based registration using manually located landmarks
- provide more insight into the results of matching score based registration

The remainder of this paper is structured as follows: in section 2 first the matching score based method and the face recognition methods used are described in more detail. Next the evaluation methods are described. Subsequently, in section 3 the experiments and results are presented. Finally, section 4 contains conclusions.

2 Methods

2.1 Matching score based registration

As mentioned in the introduction, the basic idea of matching score based registration is simple: vary the registration parameters rotation, scaling and translation until the best matching score is obtained, where the matching score is a measure of similarity or a distance measure between two facial images. This process is illustrated in fig.1.

It is to be expected that the best matching score results if the registration is correct. For small deviations from the correct registration parameters, we would normally expect a smooth degradation of the matching score. We performed experiments to test this and they confirmed this assumption. For larger deviations from the correct registration parameters we may expect to find local optima. This was also confirmed by our experiments. However, the matching score also depends on other factors like differences in illumination, facial expression and recording time between the input and

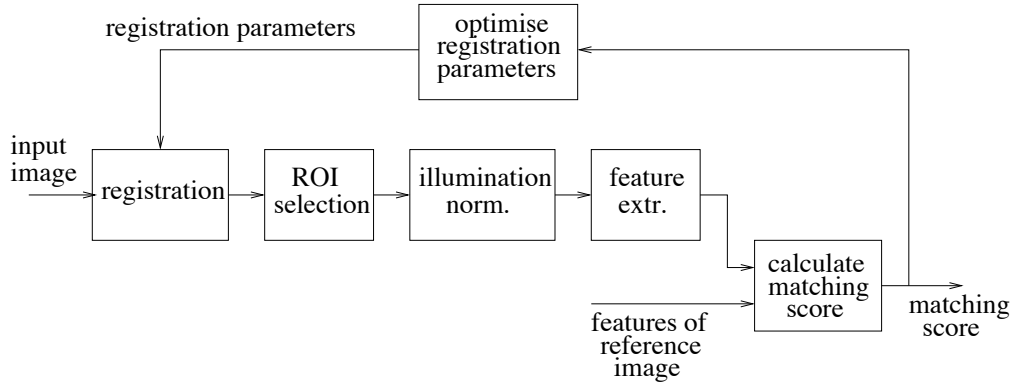


Figure 1: Box diagram of the matching score based registration approach

the reference images. These factors that cause variations of images of the same individual are called *within class variations* as opposed to the differences between images of different individuals which are called *between class variations*. If within class variations result in only slightly worse matching scores relative to errors in registration, then the global optimum remains approximately the same. If, however, within class variations have a large impact on the matching score, the global optimum may change significantly for images of the same individual ("genuine"). Also, it becomes more likely, that the comparison of certain images of different individuals ("imposters") yields a better matching score than comparison of images of the same individual. To conclude, matching score based registration will likely work well with face recognition methods, which are relatively insensitive to within class variations while for methods that are sensitive for within class variations, it can even make results.

The implementation of the matching score based face registration is straightforward. Since we want to find out if matching score based registration yields better results than landmark based registration and we may assume that the manual landmarks result in a registration close to the optimum, we only have to search the parameter space for registration in a small area around the parameters defined by the manual landmarks. This search is in our case done by an iterative direction search for each of the 4 parameters (scaling, rotation, translation in x and y). In each direction a fixed number of parameter values was chosen and for each the matching score was calculated and the best score selected. For the face recognition we chose feature extraction based on PCA, where all principle components were retained. For the matching score the Euclidian distance and the Mahalanobis Cosine distance were selected. The first is known for its sensitivity to illumination and other within class variations while the latter is known to perform better in these circumstances. Since the optimisation procedure is rather slow, it is not feasible to compute a full score matrix (i.e. compare each input image with all images in the reference or gallery set). Instead, the best n candidates were selected based on the matching score using the landmark based registration and those were registered using matching score based registration. In our case we set $n = 10$. Note that this is likely to give a pessimistic view of the results, because sometimes the correct match may not be between the 10 selected candidates and is hence excluded from improvement by matching score based registration.

2.2 Evaluation methods

For evaluation of the identification performance, rank curves were generated. In rank curves the recognition rate is plotted as a function of the rank in a sorted list of scores. E.g. rank 1 means only the best score is considered, rank 2 means the 2 best scores are

considered etc. If matching score based registration improves scores, then the curves should lie above the curves for landmark based registration. Since only the 10 best candidates were registered using matching score based registration, we can only expect differences between the landmark based and matching score based registration results for rank 1-10.

Verification performance was evaluated by generating ROC graphs. In ROC graphs, the recognition rate (RR) is plotted as a function of the FAR. The better the curve, the more it lies in the top left. If matching score based registration improves scores, then the curves should lie above the curves for landmark based registration.

To obtain insight in what happens to the matching scores by using matching score based registration, normalised histograms were generated of the genuine scores, the imposter scores and of the best score of the imposters. Here we expect the distributions to move to the left (matching score based registration minimises the distance measures used) and for improvement, the distributions of the genuine scores should move more to the left than the other distributions.

3 Experiments and results

3.1 Data

For evaluation we used the well-known FERET data set [6]. The training set used consists of 428 images, the gallery (reference or enrolled set) consists of 1196 images and there are 4 test sets: fafb, fafc, dup1 and dup2. The set fafb (1195 images) consists of samples with mainly variation of facial expression relative to the gallery, set fafc (194 images) contains images with variation in illumination, dup1 (722 images) contains images that were taken between 1 minute and 1031 days after their respective gallery matches and dup2 (234 images) is a subset of dup1 with only images with time difference of at least 18 months.

3.2 Parameter choices

For feature extraction PCA was used and all principle components were retained (428=size of training set). The search range for the registration parameters, all relative to the parameters obtained using landmark based registration, was as given in the following table:

parameter	min	max	step	remarks
x-translation	-0.05	0.05	0.005	as fraction of distance between eyes
y-translation	-0.05	0.05	0.005	as fraction of distance between eyes
scaling	0.95	1.05	0.005	
rotation	-0.05	0.05	0.005	radians

The number of iterations in the 4 directions was set to 25, so for each registration $25*4*10 = 1000$ matching scores had to be calculated. As mentioned before, matching based registration was only performed on the 10 images from the gallery with the highest scores for landmark based registration.

3.3 Identification results

Figure 2 shows the rank curves for the 4 test sets. For the Euclidian distance measure, there is very little difference between results of the landmark based and the matching score based registration methods, except for dup2 which shows a clear improvement.

For fafb the matching score based results are even slightly worse. For the Mahanalobis Cosine distance measure, however, all rank curves for matching score based registration lie above those for landmark based registration. For dup1 and dup2 a rank 1 improvement of 9% is obtained. Because only the 10 best images are used for matching score based registration, above rank 10 the curves for manual and matching score based registration coincide.

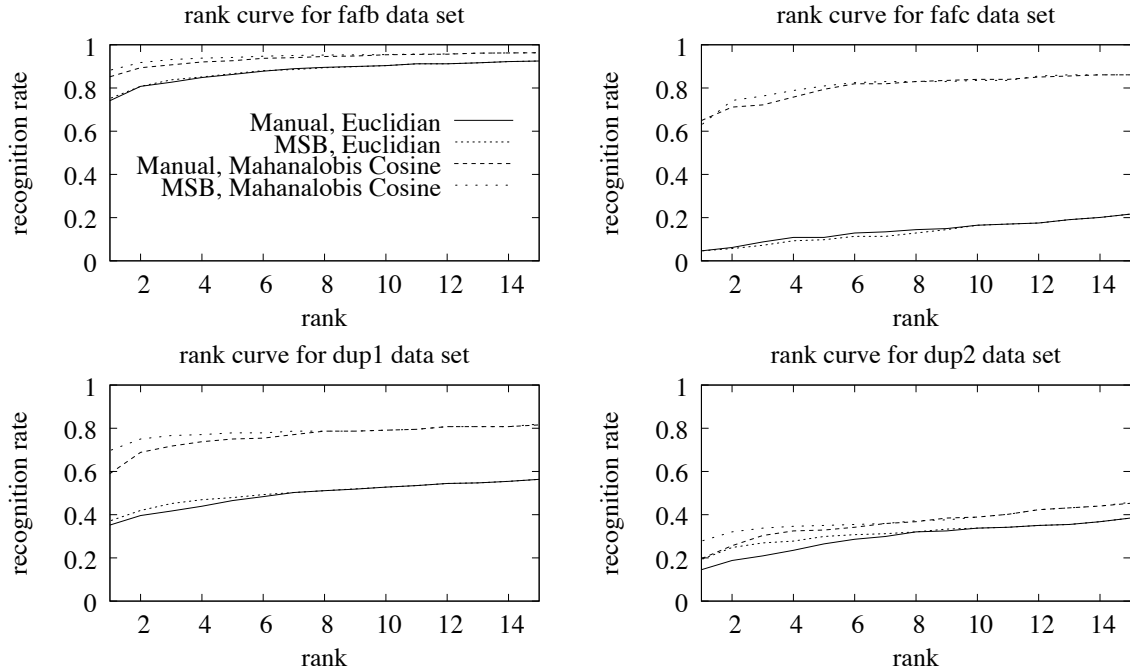


Figure 2: Rank curves for matching score based and landmark based registration

The results are in accordance with our prediction that better face recognition methods (i.e. less sensitive to within class variances) give better results with matching score based registration.

3.4 Verification results

Figure 3 shows the ROC curves for the 4 data sets. For all data sets matching score based registration gives a clear improvement for both the Euclidian as well as the Mahanalobis distance measure (the curves for matching score based registration lie well above those for landmark based registration). For the fafc data set, improvement for the Euclidian distance measure is marginal and the results are quite bad anyway. This again supports the hypothesis that matching score based registration works better well for "good" distance measures. Improvements can especially be observed for FAR below 0.01. On the average, the improvements for the Mahanalobis Cosine distance measure are larger than those for the Euclidian distance.

3.5 Histograms of matching scores

In order to study the differences of the distributions of the matching scores between landmark based registration and matching score based registration, the normalised histograms of the matching scores for the genuine, imposters and best imposters were plotted. The best imposter is the imposter with the best matching score. The overlap of the genuine and imposter distributions give a measure of the verification performance

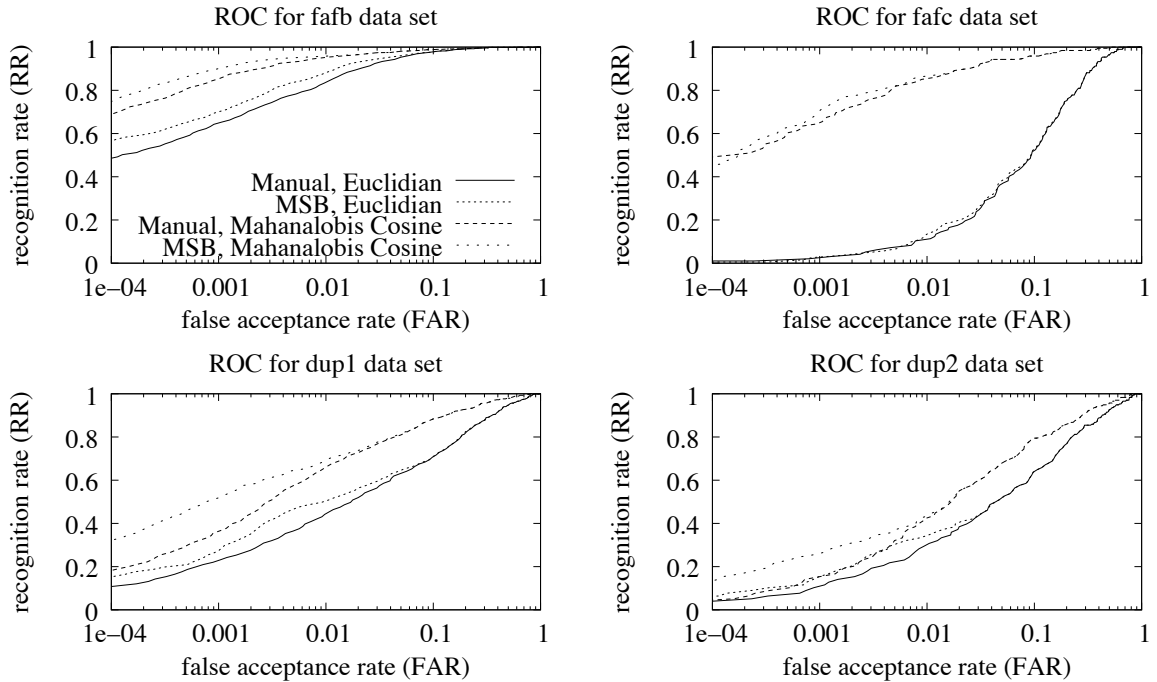


Figure 3: ROC curves for matching score based and landmark based registration

while the overlap of the genuine and the best imposter score distributions give a measure for the identification performance.

In the graphs on the left of fig. 4 we observe that for the Euclidian distance measure the distributions are more or less Gaussian and that the distributions of the genuine and best imposter scores are moved to the left by matching score based registration. Because these are distance measures this means the matching scores improve. The distribution of the imposters remains the same, because to only 10 out of approximately 1195 imposters matching score based registration is applied. Except for the fafc data set, the overlap of the genuine and imposter and the genuine and best imposter distributions becomes slightly smaller by matching score based registration, indicating a slight improvement in recognition for both verification as well as for identification. For the fafc data set, the identification becomes worse, because the best imposter score distribution is shifted left more than the genuine score distribution. For the Mahalanobis Cosine distance measure (right graphs), the improvements are stronger and even present for the fafc data set. Furthermore, we can observe that the distributions for this distance measure are not Gaussian. The asymmetry seems to be beneficial for both identification as well as verification, because the right tail of the genuine distribution goes to 0 faster for matching score based registration scores.

3.6 Further experiments

Some further experiments were performed to investigate the improvements by matching score based registration of genuine samples which were not between the 10 candidates selected using landmark based registration. It appeared that the scores of some of these samples was improved dramatically. E.g. in the fafb data set for a certain test sample, the genuine sample from the gallery set was originally at rank 72 with Mahalanobis Cosine distance of -0.10. After matching score based registration it surpassed rank 1 (-0.479) and got a matching score of -0.622. The range of the Mahalanobis Cosine distance measure is $[-1, \dots, 1]$. Visual inspection of the matching score based and landmark based registered images revealed hardly any differences. Apparently, very slight

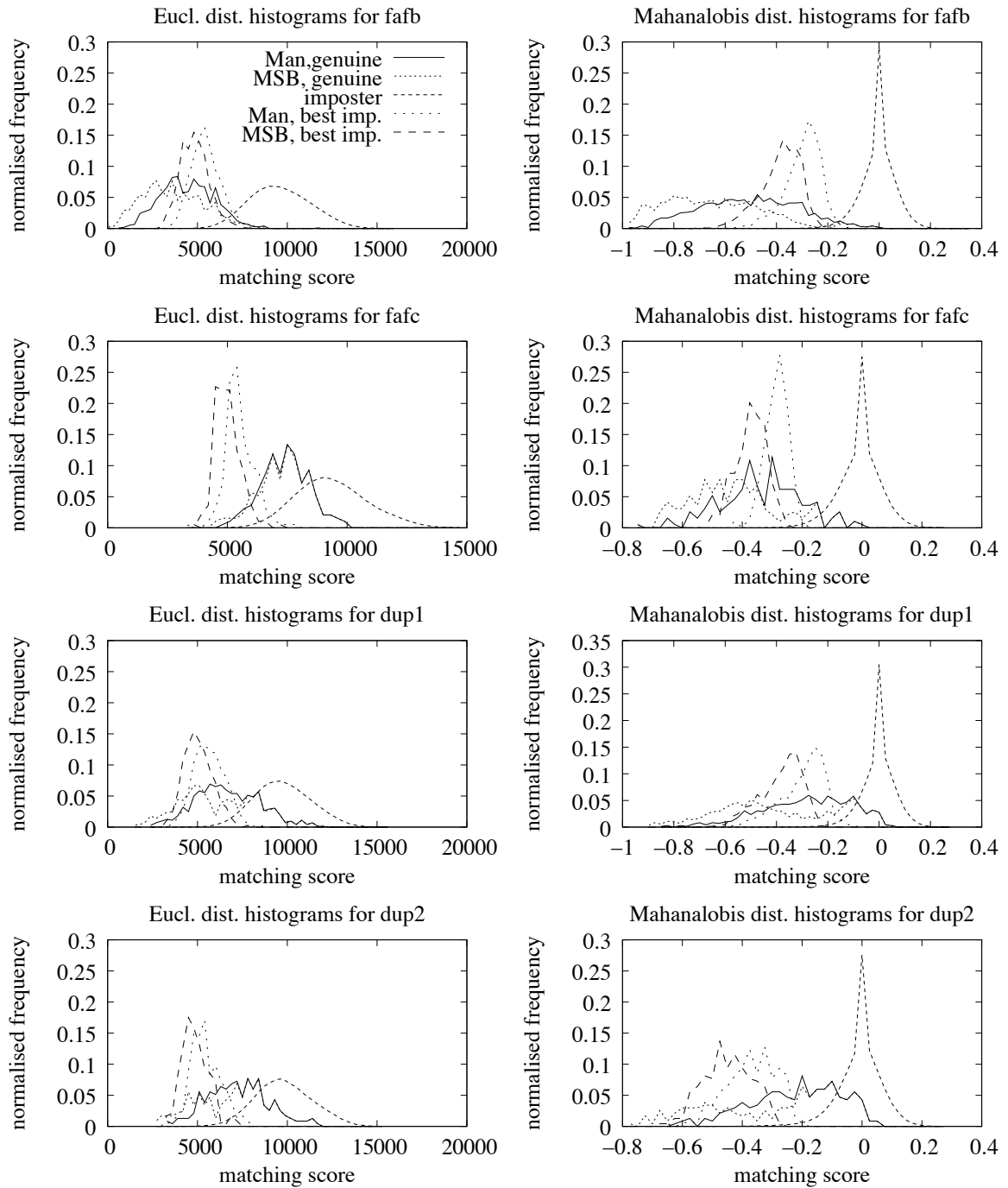


Figure 4: Normalised histograms of genuine and imposter scores for matching score based and landmark based registration

changes in registration parameters can result in dramatic changes in the matching score. This is a reason why landmark based registration may fail in some occasions.

4 Conclusions

The results of this research show that matching score based registration can result in significantly better identification and verification rates even relative to the "gold standard" of landmark based registration using manually located landmarks. Improvements of rank 1 scores of up to 9% were obtained in our experiments. For verification, especially the recognition rates for low FAR are improved significantly. Our hypothesis that matching scores that are relatively insensitive to within class variations (illumination, facial expression, aging etc.) improve more by matching score based registration and that matching scores that are very sensitive to these variations might even give worse results is supported by our experiments with the Mahalanobis Cosine and Euclidian distance measures. We also found that small variations in the registration parameters can have a large effect on the matching score. This is one of the reasons why landmark based registration sometimes fails, while matching score based registration is better able to handle these cases. By studying normalised histograms of the matching scores for landmark based and matching score based registration, we gained more insight into the distribution of the matching scores. We will continue our research into matching score based registration in combination with different recognition algorithms, different data etc. and further explore its optimal use.

References

- [1] T. Ahonen, A. Hadid, , and M. Pietikäinen. Face recognition with local binary patterns. In *Proceedings of the 8th European Conference on Computer Vision - ECCV 2004*, pages 469–481, Prague, Czech Republic, May 11-14 2004.
- [2] A. Bazen and R. Veldhuis. Likelihood-ratio-based biometric verification. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:86–94, Jan. 2004.
- [3] G. Beumer, Q. Tao, A. Bazen, and R. Veldhuis. A landmark paper in face recognition. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, Southampton, UK, April 2006.
- [4] B. J. Boom, G. M. Beumer, L. J. Spreeuwiers, and R. N. J. Veldhuis. Matching score based face recognition. In *Proceedings of ProRISC the 17th Annual Workshop on Circuits, Systems and Signal Processing, Veldhoven, The Netherlands*, pages 1–4, November 2006.
- [5] J. Lu, K. Plataniotis, and A. Venetsanopoulos. Face recognition using LDA based algorithms. *IEEE Transactions on Neural Networks*, 14(1):195–200, 2003.
- [6] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [7] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71 – 86, 1999.
- [8] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.