ICSP2008 Proceedings

Eigenfaces vs. Fisherfaces vs. ICA for Face Recognition; A Comparative Study

M. Sharkas, M. Abou Elenien
Arab Academy for Science and Technology - AAST, Egypt
msharkas@aast.edu, elenien@aast.edu

Abstract

Face recognition issue gained more interest recently due to its various applications and the demand of high security. Some researches with contradicting results were published concerning this issue. This paper compared three popular face recognition projection methods: (eigenfaces), (fisherfaces), and ICA. We also applied some data transformations: (Discrete Wavelet and cosine Transforms) preceding methods to see their effect. Most researches based their results on the FERET database. AR and AT&T databases were used here to see if the same results apply. We also compared the results of two sets of experiments with the second set using half the training images used in the first to observe if the results may change. Overall conclusion is it can't be stated that specific algorithm outperforms others, though ICA and Eigenfaces respectively showed better results than fisherfaces for both experiments sets and both databases. Preceding algorithms with transformations yield better results for some algorithms.

1. Introduction

Many applications rely on the performance of digital image processing systems like biometrics authentication, multimedia, computer human interaction, security applications, etc. Face recognition as a main application of image processing has been developing rapidly in the past few years. It has been developing not only cause of the problem of recognition itself but also because of the numerous applications where accurate human identification is required. Also security purposes like security requirements in airports gave their share in helping face recognition gain more interest.

Projection methods to be presented are: Eigenfaces using Principal Component Analysis (PCA), Fisherfaces using Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA). We barely found any paper comparing the three algorithms together. The 3 different face recognition algorithms compared lie under the holistic based face recognition method where the whole face region is used as an input to a recognition system. These methods - though some are really old like the PCA - give some very good results. We also introduced some modifications and data transformations to the data when carrying out the algorithms to see their effect on the results as we will discuss later. Briefly the test and training images are projected into a lower-dimensional subspace using different algorithms and then we use distance measures to identify a test image for recognition purposes.

2. Algorithms and previous work

Some papers evaluated each algorithm independently or compared between at most two algorithms. Their results differ significantly and there was no apparent method to be said the best performing one. Some of these papers like Beveridge et al. [13] argued that LDA performed worse than PCA. Martinez and Kak [4] showed that when the training data set is small, PCA can perform better than LDA and also that PCA is less sensitive to different training data sets. Liu and Wechsler [9], and Bartlett et al. [1] found that ICA performs better than PCA, while Baek et al. [11] showed that PCA is better. Delac and Grgic's [10] claimed that there is no algorithm (projection–metric combination) can be considered the best, and that the choice of appropriate algorithm can only be made for a specific task. Moghaddam [12] claimed that there is no significant difference. Navarrete and Ruiz-del-Solar [14] and Belhumeur et al. [8] stated that in the tests they carried out, LDA outperformed PCA.

3. Eigenfaces

For eigenfaces, Principal Component Analysis (PCA) [2] is implemented. PCA is a method that is used a lot for dimensionality reduction. We will use the terms eigenfaces and PCA interchangeably. Consider having a set of face images which can be represented by a g-dimensional original space. PCA aims at getting

978-1-4244-2179-4/08/$25.00 ©2008 IEEE

an h-dimensional subspace having basis vectors which corresponds to the directions of maximum variance in the g-dimensional space. Accordingly h is much smaller than g. The original subspace is mapped into the smaller one through a linear transformation. The *L* training images are normalized then subtracted from the calculated mean image to get mean centered images. Let *W* be a matrix containing the mean centered training images w_i where *i = 1, 2, …, L.* From *W* we get the covariance matrix *D* where

$$D = WW^T$$

The covariance matrix *D* could be too large and it was proved mathematically that we can use $D = W^T W$ instead. Then from *D* we get the eigenvectors e_i and their associated eigenvalues λ_i . Then a matrix *E* is formed where it is composed of the eigenvectors associated with the largest eigenvalues to obtain $z_i = E^T w_i$ where, *i = 1,2,….L*

z_i is the new feature vectors in the new lower dimensional subspace. Therefore the training images are projected this way into the new subspace. A test image is treated in the same manner until projected onto the same subspace. We then use distance measures to test the similarity between the test and training images.

4. Fisherfaces

For fisherfaces, using also the terms fisherfaces and LDA interchangeably, Linear Discriminant Analysis (LDA) [7, 8] tries to differentiate between classes rather than trying to present the data. Therefore, LDA cares about getting feature vectors for class discrimination. We define 2 scatter matrices

$$S_w = \sum_{j=1}^{R} \sum_{i=1}^{M_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T$$

$$S_b = \sum_{j=1}^{R} (\mu_j - \mu)(\mu_j - \mu)^T$$

The first is called the within-class scatter matrix while the second is called the between-class scatter matrix. *j* denotes the class while *i* denotes the image number. μ_j is the mean of class *j* while μ is the mean of all classes. M_j is the number of images in class *j* and *R* is the number of classes. The algorithm aims at maximizing the between-class matrix while minimizing the within-class matrix. This can be done by maximizing the ratio ($\det |S_b| / \det |S_w|$). As with PCA we have a projection matrix *G.* This matrix is used to maximize the mentioned ratio when its columns are the eigenvectors of $S_w^{-1} S_b$. A problem arises when S_w becomes singular. [8, 6] suggested using an intermediate space - before transforming the original space to the final one - which is the PCA space. In other words the original space is projected to the PCA space first then to the LDA space.

5. Independent component analysis (ICA)

ICA is considered as a generalization of PCA. PCA considers image elements as random variables with minimized 2nd order statistics. ICA proposed by [1, 3] minimizes both second-order and higher order dependencies in the input data and tries to get the basis of which the projected data is statistically independent. Also here PCA is used to reduce dimensionality prior to performing ICA.

Two different approaches or architectures are taken by the ICA for face recognition where the first one is called ICA Architecture 1. In this approach according to [1] images are considered as random variables and pixels as trials. So here we care about independence of images or functions of images. In other words ICA arch 1 tries to find a set of statistically independent basis images. The second approach is called ICA Architecture 2. In this approach, pixels are considered as random variables and images as trials. So in this one we care about independence of pixels or functions of pixels. In other words ICA arch 2 uses ICA to get a representation in which the coefficients used for coding images are statistically independent

6. Experimental results

Two database sets were used in the experiments. The AR database is used in this research under permission from A.M. Martinez [5]. We manually localized, cropped, and resized the images of 50 different individuals (25 males and 25 females). For the AT&T database [15], which is already cropped and resized, 40 different individuals are available.

Figure (1): Some images of one individual in the AR database (up) and one individual in AT&T (down).

We made 2 sets of experiments. Same algorithms and transformations were introduced to both experiment sets. The difference is that the 2nd set used only half of the training set used in the 1st. On the other hand more test images are used in the second set. We observed if these changes may affect the results.

Testing was carried out by using the nearest-neighbor algorithm using the standard L2-norm for the Euclidean distance for eigenfaces and fisherfaces algorithms. The cosine distance was used for the ICA Arch1 and Arch2 algorithms. The nearest neighbor is determined by calculating the minimum distance d from a test image projection to all training projections. All algorithms were implemented in MATLAB.

6.1 1st Set experimental results

In the first set, for the AR database, 10 images were randomly selected for each individual to be used in the training set plus one more image for the test set, so we have 500 training images and 50 test images.

For AT&T database, 9 images were selected for each individual to be used in the training set plus one more image to be used in the test set, so we have 360 training images and 40 different test images.

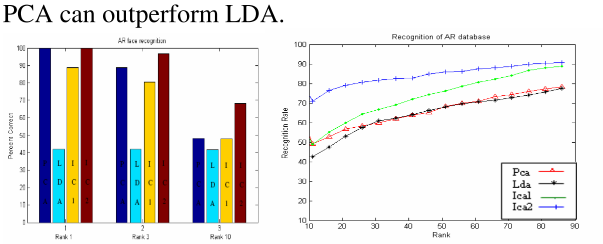
In our experiments we have applied discrete wavelet transform DWT and discrete cosine transform DCT on data before carrying out algorithms. DWT [16, 17] is a transformation which can be used for analyzing the temporal and spectral properties of non-stationary signals. DCT [18] is a technique for converting signal into elementary frequency components.

Several ways were used to present the results. Bar plots show the algorithms performance at different ranks. Rank 1 corresponds to correct recognition rate within the first one match, rank 3 (correct recognition rate within the first 3 matches) and rank 10 (correct recognition rate within the first 10 matches) for all algorithms with and without DWT and DCT. Also Cumulative Match Score (CMS) curves are presented [19], showing the cumulative recognition rates for ranks 10 and higher for all different algorithms combinations as said above. We observed in this set that in most cases the results of these different methods agreed when trying to find the best performing algorithm, meaning that mostly the best algorithm at rank 1, 3 and 10 is the same best at higher ranks. Lower ranks results were good except for fisherfaces due to the great number of varying training images.

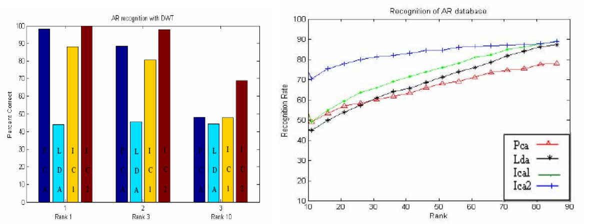
Starting with the AR database, first the 4 algorithms are compared without introducing any modifications as shown in figure 2 (a). It was observed that ICA Arch 2 yielded the best results at all ranks and at the CMS curves. PCA eigenfaces gives same result as ICA arch2 for rank 1 and the 2nd best result for rank 3, but further on its results got much worse. ICA Arch1 gives the 2nd best results for rank 10 and higher. LDA fisherfaces showed the worst results for almost all tests.

On subjecting face images to 1-level DWT, the approximation coefficients are introduced to different algorithms instead of using images vectors directly. Figure 2 (b) demonstrate that ICA Arch2 remained the best algorithm giving the best results. No significant change was observed in the PCA or ICA Arch1. The performance of LDA improved much when preceded with DWT approaching ICA Arch2 results at rank 85.

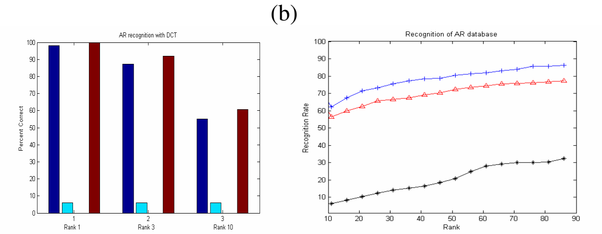
Introducing DCT as shown in figure 2 (c), ICA Arch1 data was distorted in the whitening process, so it is out of the comparison. ICA Arch2 results have been slightly reduced while LDA results have been significantly reduced. On the contrary PCA surprisingly has significantly improved especially for lower ranks. Obviously ICA Arch 2 outperforms other algorithms for all cases. PCA gives the 2nd best results for rank 10 and lower, while ICA Arch 1 yields 2nd best results for ranks 10 and higher. LDA has shown poor performance compared to other algorithms. Some problems in implementing LDA were addressed in detail in Martinez and Kak [4]. where it is concluded that when the training set is small, or when the training data non-uniformly sample the underlying distribution, PCA can outperform LDA.



(a)



(b)



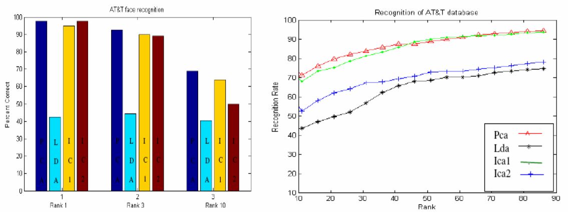
(c)

Figure (2): Exp. Set 1, AR database, bar results (left) and CMS results (right) for PCA, LDA, and ICA (a) without modifications, (b) with DWT, (c) with DCT

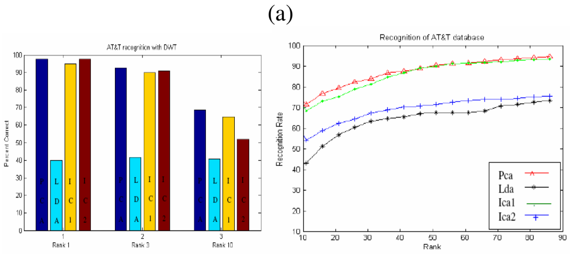
For the AT&T database, as before, the algorithms first are compared without modifications. The results are presented in figure 3 (a). Here PCA outperformed other algorithms for all ranks though ICA Arch1 is comparable giving results that are very close. ICA Arch 2 gave the same result as eigenfaces for rank 1 but the results deteriorated significantly afterwards. Still fisherfaces gave the poorest performance.

Introducing DWT showed the results shown in figure 3 (b). PCA eigenfaces still yielded best results. ICA Arch1 yielded results that are so close to PCA and are almost the same for ranks 40 and higher. ICA Arch2 gave results that are the same as PCA for rank 1 but is reduced significantly afterwards. Introducing DWT slightly improved the results. LDA remained the worst though introducing DWT improved its performance especially for ranks lower than 40.

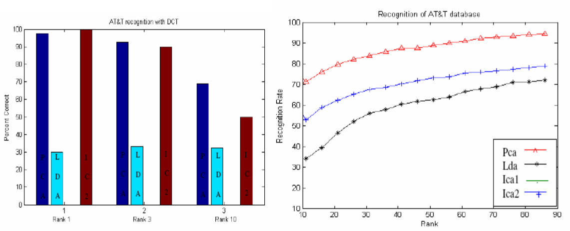
Results of introducing DCT are presented in figure 3 (c). Again ICA arch1 wasn't presented due to distortion. ICA Arch2 results have stayed almost the same except for an improvement in rank 1 after introducing DCT giving best results for all algorithms in this rank. While the LDA results have been reduced significantly, PCA stayed almost the same for all ranks giving best results for ranks other than rank 1.



(a)



(b)



(c)

Figure (3): Exp. Set 1, AT&T database, bar results (left) and CMS results (right) for PCA, LDA, and ICA (a) without modifications, (b) with DWT, (c) with DCT

6.2 2nd Set experimental results

In the second set of experiments, we used different number of training and test images. For the AR database, 5 images were randomly selected for each individual to be used in the training set. 5 different images were used for the test set, so we have 250 training images and 250 test images.

For AT&T database, 5 images were selected for each individual to be used in the training set. 5 different images were used for the test set, so we have 200 training images and 200 test images. So we used in the 2nd set of experiments half the number of training images used in the 1st set. On the other hand we used a greater number of test images.

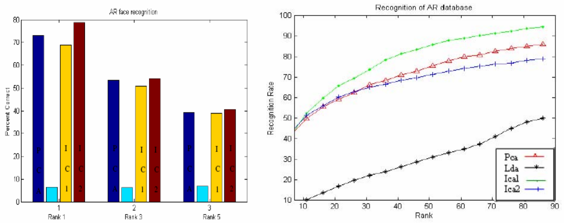
Same ways were used to present the results as well. Bar plots are used with rank 5 this time instead of 10 (correct recognition rate within the first 5 matches) and (CMS) curves, showing the cumulative results for ranks 7 and higher for all algorithms combinations. As in the 1st set we observed that AT&T results of these methods agreed when trying to find the best performing algorithm, meaning that mostly the best algorithm at rank 1, 3 and 5 is the same best at higher ranks. But their results didn't agree this time with the AR database. Also ranks 1 and 3 results were not as good as they were for the 1st experiments which is reasonable because of the fewer training images used.

Again starting with the AR database, the 4 algorithms are compared without introducing any modifications as shown in figure 4 (a). It was observed here too that ICA Arch 2 yielded the best results for ranks 1, 3, and 5 but for higher ranks as shown in the CMS curves its results got worse than eigenfaces and ICA arch 1. PCA eigenfaces gave the 2nd best results for only lower ranks. ICA arch 1 gave the 2nd best results for rank 6 and higher. Also the LDA fisher faces showed the worst results for all ranks.

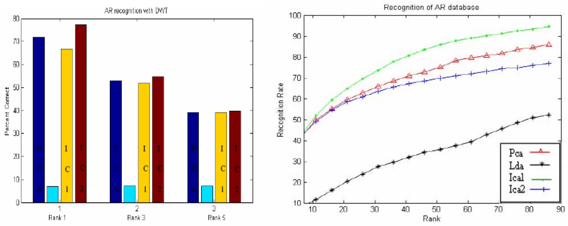
In figure 4 (b), subjecting images to 1-level DWT. ICA Arch2 remained the best algorithm for ranks 1, 3, and 5. For higher ranks its results stayed worse than PCA and ICA arch 1. No significant change was observed in the PCA or ICA Arch1. The performance of LDA improved when preceded with DWT.

On introducing DCT as shown in figure 4 (c), also here ICA Arch1 data was distorted in the whitening process, so it is out of the comparison. ICA Arch2 and fisherfaces results have been reduced but ICA arch2 stayed the best for rank 1. PCA results were reduced for the 1st rank only while they have improved for the rest of the higher ranks.

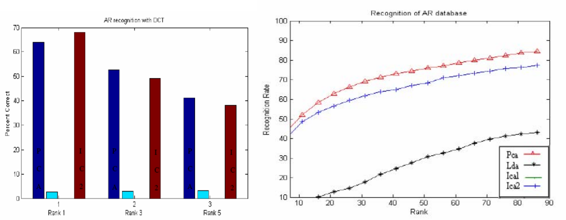
For AR database obviously ICA Arch 2 outperforms other algorithms for lower ranks only here while for the 1st set of experiments it outperformed all algorithms for all ranks which shows that ICA arch 2 here works better with larger training sets. PCA gave the 2nd best results for rank 5 and lower, while ICA Arch 1 yields 2nd best results for ranks 6 and higher. This agrees with the 1st experiments results. In the 2nd set of experiments PCA and Arch 1 gave better recognition rates for higher ranks and worse rates for lower ranks than the 1st set of. Again LDA has shown poor performance compared to other algorithms.



(a)



(b)



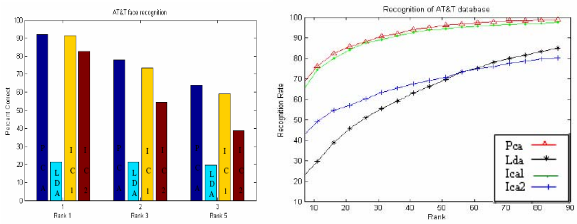
(c)

Figure (4): Exp. Set 2, AR database, bar results (left) and CMS results (right) for PCA, LDA, and ICA (a) without modifications, (b) with DWT, (c) with DCT

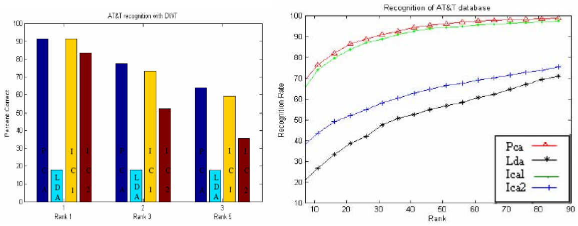
For the AT&T database, in figure 5 (a), the algorithms are compared with no modifications. Here again Eigenfaces outperformed other algorithms for all ranks though ICA Arch1 is comparable giving results that are very close to eigenfaces as seen from the CMS curves and the bar plots. Eigenfaces and ICA arch 1 gave same results for rank 1 only. ICA Arch 2 came in the 3rd position after them and even its results deteriorated more than the fisherfaces starting from rank 55 and higher. Still fisherfaces yielded the poorest performance of them all.

In figure 5 (b), after introducing DWT as before, it was observed that eigenfaces still gave the best performance and even as seen in the CMS curves it showed a slight improvement. ICA Arch1 yielded results that are so close to PCA with the same performance at rank 1. ICA Arch2 results improved slightly for rank 1 while reduced afterwards. LDA results were reduced and remained the worst.

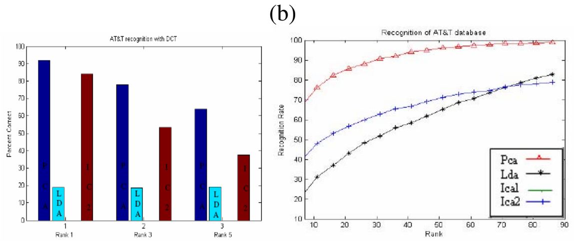
The results of introducing DCT are shown in figure 5 (c). ICA arch 1 is excluded from the comparison again cause of the data distortion in the whitening process. Again PCA eigenfaces outperformed all algorithms for all ranks. With a little improvement in rank 1, ICA Arch2 results have stayed almost the same. LDA results were reduced staying the worst of all other algorithms.



(a)



(b)



(c)

Figure (5): Exp. Set 2, AT&T database, bar results (left) and CMS results (right) for PCA, LDA, and ICA (a) without modifications, (b) with DWT, (c) with DCT

For AT&T we observed that the results of the 2nd set of experiments almost agreed with the 1st set except for an obviously lower performance of ICA arch2 which again shows that ICA arch 2 works better with larger training set. Again in the 2nd set of experiments PCA and Arch 1 gave better performance for higher ranks and worse performance for lower ranks than the 1st set of experiments.

7. Conclusions

In this paper three of the most popular face recognition algorithms are compared. Some transformations are applied on the data while carrying out the algorithms. AR and AT&T databases are used in the study. And all the results were compared one time with a large set of training images and the other time with a much smaller set of training images. The overall conclusions are: (1) In general none of the algorithms can be said to outperform the other. (2) For rank 1, ICA arch 2 yielded the best results of all algorithms for both databases in most cases (3) All algorithms yield better recognition performance for lower ranks as observed with larger training sets, especially ICA Arch 2 which results improve significantly for all ranks (4) On the other hand the orders of the best performing algorithms do not interchange significantly with larger or smaller training sets. (5) For all other ranks ICA and PCA yielded best results interchangeably showing the close performance of them though we can say in general that ICA is slightly better than PCA. This drives us to the question whether it is really worth using ICA with all its computational complexity and its implementation time. (6) Results vary significantly between both databases which show the high dependency of the results on the databases used and the way they are collected. (7) LDA yields the worst performance of all algorithms which shows its high sensitivity to any variations in the training images within each class. (8) Introducing DWT has slightly improved the results especially with the LDA algorithm. On the other hand introducing DCT did not work that well with ICA and LDA but has in general improved PCA, and sometimes with ICA arch2 it improved the results of rank 1 only. (9) More attention should be driven to introducing some data transformations to each algorithm independently and observing their effect on the overall performance. Further researches are required for different algorithms-metric-transforms combinations.

8. References

[1] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans Neural Networks 13 (2002), 1450–1464.

[2] M. Turk and A. Pentland, Eigenfaces for recognition, J Cogn Neurosci 3 (1991), 71–86. W.

[3] B. Draper, K. Baek, M.S. Bartlett, and J.R. Beveridge, Recognizing faces with PCA and ICA, Comput Vis Image Understand (Special Issue on Face Recognition) 91 (2003).

[4] A. Martinez and A. Kak, PCA versus LDA, IEEE Trans Pattern Anal Machine Intell 23 (2001), 228–233.

[5] A.M. Martinez and R. Benavente, “The AR-face database” CVC Technical Report 24, June 1998.

[6] D. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," IEEE Transactions on Patt Analysis and Mach Intell, vol. 18, pp. 831-836, 1996.

[7] Zhao, R. Chellappa, A. Krishnaswamy, Discriminant analysis of principal components for face recognition, Proc Third IEEE Int Conf Automatic Face and Gesture Recogn, Nara, Japan, 14–16 April, 1998, pp. 336– 341.

[8] P. Belhumeur, J. Hespanha, and D. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, Proc Fourth Eur Conf Computer Vision, Vol. 1, 14–18 April 1996, Cambridge, UK, pp. 45–58.

[9] C. Liu and H. Wechsler, Comparative assessment of (ICA) for face recognition, 2nd International Conference on Audio and Video-based Biometric Person Authentication, Washington DC, 22–23 March, 1999.

[10] K. Delac, M. Grgic, S. Grgic, Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set, International Journal of Imaging Systems and Technology, Vol. 15, Issue 5, pp. 252-260

[11] K. Baek, B. Draper, J.R. Beveridge, and K. She, PCA vs. ICA: A comparison on the FERET data set, Proc Fourth Int Conf Computer Vision, Pattern Recogn and Image Process, Durham, NC, 8–14 March, 2002, pp. 824– 827.

[12] B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition, IEEE Trans Pattern Anal ysis and Machine Intelligence 24 (2002).

[13] J.R. Beveridge, K. She, B. Draper, G.H. Givens, A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition, Proc IEEE Conf Comp Vision and Patt Recog 2001, pp. 535–542.

[14] P. Navarrete,J. Ruiz-del-Solar, Analysis and comparison of Eigenspace-based face recognition approaches, Int J Pattern Recogn Artif Intell 16 (2002). 817–830

[15] “AT&T database” AT&T Laboratories Cambridge. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase

[16] Graps, Amara, Summer, “An Introduction to Wavelets,” IEEE Computational Sciences and Engineering.

[17] Strang, G., “Wavelet Transforms vs. Fourier Transforms,” Bulletin of AMS, 28(2), pp. 288-305 1993.

[18] F. A. Kamangar and K.R. Rao, “Fast Algorithms for the 2-D Discrete Cosine Transform,” IEEE Transactions on Computers, v C-31 p. 899.

[19] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, The FERET evaluation methodology for face recognition algorithms, IEEE Trans Pattern Anal Machine Intell (2000)