Your Project Report Title Here  
CS7IS2 Project (2019-2020)

Student Name (Multiple names separated by comma)

email\_address\_1 (multiple email addresses separated by comma)

**Abstract.** This paper compares Principle Component Analysis (PCA), Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF) based on their accuracy of predication on facial recognition and time it takes for the SVM classifier to train on them. This paper also tries to explain the underlying differences in each of the three-dimension reduction methods. In the comparison PCA outperforms ICA slightly and NMF performed the worst when it comes to prediction accuracy.

1 Introduction

In the recent years, within the larger filed of computer vision face recognition has become a niche application area. There are now commercial systems which can in real time perform face detection, image recognition and matching. Even these sophisticated systems like most of the simple yet canonical scenarios are based on the two-step process of subspace projection and then classification [34].

There are numerous techniques applied by face recognition systems for subspace selection. This paper tries to compare the accuracy of three subspace projection methods namely PCA, ICA and NMF when used in conjunction with SVM classifier to recognize faces. These algorithms are compared based on accuracy of prediction and the time it took for the classifier to run on the subspace projections produced by them.

The choice of dimensionality reduction techniques is based on the contradicition available in the literature about the performance of PCA and ICA. Bartlett [4, 6], Liu and Wechsler [30], and Yuen and Lai [41] in their work claim that when it comes to face recognition ICA outperforms PCA whereas Baek in his work claim that [1] suggests that ICA is outperformed by PCA and Moghaddam in his work [32] reports that there is no as such difference between the performance of the two [32]. These works suggest that the performance comparison of these two algorithms is still an open question.

There are numerous factors while implementing PCA and ICA which are to be taken care of, for PCA these are task statement, distance metric for subspace and the number of subspace dimensions retained and for ICA these are the task, approximation algorithm used for ICA, and retained number of subspace dimensions. There are two varied ICA architectures, one treats images as random variables and the other treats pixels as random variables. In this paper we use the one which uses images as random variables, which is called ICA architecture I. The ICA architecture architecture under consideration produces features which are spatially localized i.e. they are only influenced by small part of images.

Even though it has been argued that since ICA under consideration implements facial recognition by parts, it will produce better object recognition [27].This paper will show that the performance of PCA and ICA are almost similar whereas NMF lags way behind when it comes to prediction accuracy and it even takes the most time for the SVM classifier to be trained.

This paper is further broken down into 4 more sections. Section 2 of the paper reviews the literature available in the field of application of dimension reduction method in face recognition. Section 3 provides the basic theoretical knowledge of PCA, ICA, NMF and the SVM classifier. Section 4 contains the steps conducted to complete the research. Section 5 concludes with practical recommendations.

2 Related Work

Research in the field of face recognition started way back in 1960’s [13]. With the emergence of appearance- based most current face recognition techniques emerged in 1980’s and 1990’s. PCA was first applied to face images by Kirby and Sirovich. They showed that using it images can be compressed to their reconstructions with minimum mean squared error between them, i.e. it is an optimal compression scheme [26, 36]. PCA for face recognition was also used by Turk and Pentland [40]. In fact, they popularized its use. In their work compression to compressed subspace (eigenspace) was performed from a database of face images by computing a set of subspace basis vectors (eigenfaces) using PCA. The success of this method leads to the popularization of matching images in the compressed subspace. PCA produces spatially global feature vectors.

Researchers also put in efforts to create techniques that create spatially localized feature vectors, in the hopes that they would implement recognition by parts. ICA is the most general way to generate spatially localized features, which does it by producing statistically independent basis vectors [2]. One other method to generate localize feature vector is Non-negative matrix factorization (NMF) [27]. ICA can also be used to create feature vectors that uniformly distribute data samples in subspace [4, 5]. This is conceptually different from what ICA is believed to do, i.e. rather than producing features vectors which are spatially localized, it produces feature vectors which produce very fine distinction between images in order to spread the samples in subspace. To Keep up with the terminology, we refer to former as architecture I, and the latter as architecture II.

There are certain techniques which are formed by the combination of local linear subspaces. For instance, mix local PCA which is used to compress face data by Kambhatla and Leen [24], and a mixture of factor analyzers produced by Frey [19]. Although they have not yet been applied to face recognition, Tipping and Bishop, and Lee provide an algorithm for optimizing mixture models of PCA and ICA subspaces respectively [39, 28].

Altenative to these algorithms is a supervised learning algorithm knows as Fisher’s linear discriminant analysis (LDA, a.k.a. “fisherfaces”) [38]. Its goal is to produce N-1 basis vector in order to maximize and minimize the intra and inter class distances between N-clsses of N-class problem. Even though LDA differ PCA in terms that one is supervised learning algorithm and the other is unsupervised but when the data is labeled then either of the two can be used and in these cases LDA has been compared to PCA [7, 10, 31, 37]. One common characteristic of both PCA and LDA is that they produce spatially global feature vectors.

3 Problem Definition and Algorithm

This paper tries to implement numerous dimension reduction methods in conjunction with SVM classifier to aid facial recognition process. The dimension reduction methods used are PCA (Principle Component Analysis), ICA (Independent Component Analysis) and NMF (Non-negative Matrix Factoriztaion).

3.1 Principle Component Analysis (PCA)

This was introduced by Karl Pearson in 1901 and is mostly used for EDA and in predictive models. It is the eigenvector based multivariate analysis and is the simplest in its kind. It is basically reducing the dimensionality of the data in a way that it explains the variation (major features) in the best possible way to make it more understandable.

Say there are n\*n dimensional images i.e. n2 dimensions (pixels). When such a multivariate dataset is visualized as set of coordinates in high dimensional data space, then PCA can provide with the lower dimensional picture (eigen faces) of the object when viewed from its most informative viewpoint.

Mathematically by definition PCA is a procedure that converts M correlated into a set of K uncorrelated variables called principal components using orthogonal transformations (where K<=M).

Since PCA selects the principle components which show the direction of data and each proceeding component shows less direction and more noise. Hence first few principle components are enough to represent the original data.

Once we have found the ‘k’ principle components then image in the dataset or incoming new data can be represented as the linear combination (weighted sum) of k components, i.e. W\*K (W is a matrix of weight and k is a matrix of eigenfaces).

When we represent the data this way it reduces the number of values needed to recognize it. This makes the process faster and more error free because this discards the noise in the dataset. It is done by eigenvalue decomposition of a data covariance matrix. Results are usually discussed in form of principle components (how much of each of the k principle components) and loadings (weight).

3.2 Independent Component Analysis (ICA)

PCA is about finding correlation and the way that it does this is by maximizing variance, which in turn gives us the ability to reconstruct. ICA on the other hand tries to maximize independence. This when simply put means that it tries to find linear transformation of feature space into new feature space such that each of the individual new features are mutually independent. The new components produced are perpendicular to each other and there is no information loss between moving from observables to independent components.

There are certain hidden variables (independent components) which are random and mutually independent (value of one doesn’t tell us anything about the value of other). Then there are known observable variables which are given rise to by the linear combination of these hidden variables. The job of unobservable learner in this case is to find the hidden variables given the observables (under the assumption that hidden variables are independent of one another).

The classical example of this is the Blind Source Separation problem.

|  |  |  |
| --- | --- | --- |
|  | PCA | ICA |
| Mutually Orthogonal | Yes | No |
| Mutually Independent | No | Yes |
| Maximal Variance | Yes | No |
| Maximal Mutual Interaction | No | Yes |
| Ordered Features | Yes | No |
| Bag of Features | Yes | Yes |

*Table 1: Differences between PCA and ICA*

PCA is directional, and when it comes to face recognition problem it captures brightness, avg. (Eigen) face. Since PCA is based on global orthogonality hence it captures global features.

ICA on the other hand in face recognition problem captures nose, eye selectors, mouth selectors, hair selectors. Since ICA is based on local search hence it finds parts of face.

3.3 Non-negative Matrix Factorization (NMF)

Basic idea of NMF is to reduce a given matrix into two matrices which are both easier to work with and which when multiplied produce the original matrix. For a matrix V with n\*m dimensions the NMF will try and decompose it into two matrices W and H with n\*r and r\*m dimensions respectively, such that:

W\*H=V and Wij, Hij, Vij >=0

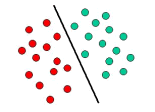
This problem cannot be solved analytically so it is generally solved numerically. NMF is relatively a new way of reducing dimensionality of our data into a linear combination of bases which in turn is necessary for machine learning algorithms for ease of computation. In the image example it;’s very difficult to consider all the pixels each time the image is handled so it is better to reduce the image into few representative pixels.

Due to fact that NMF has this non-negative constraint it can be used to can be used to depict data with non-negative features. It is quite similar to PCA but weights are only allowed to have positive values. This is compatible with the intuitive notion of combining parts to form the whole. Sparse bases and spares weightings are constructed by NMF assuming that there is an underlying structure in the data.

When it comes face recognition NMF forms bases that are parts of the face. The bases in this case are mostly empty and weighting matrix is also sparse, i.e. all parts are not used to form the image. On the contrary PCA has both its matrices densely populated and forms bases of positive and negative pixels and the weight blends them together.

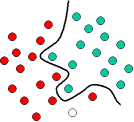
3.4 Support Vector Machine

Support Vector Machines work on the very idea of decision planes which states decision boundaries. Members belonging to different classes are separated by using decision planes. An example is shown below. The objects belong either to class red or they belong to the green one in this example. Both the classes are separated by the separating line such that all objects to the right are red and to the left are green. The new object falling to the right will be labelled red and to the left will be labelled green (or classified as RED should it fall to the left of the separating line).



Fig

The example shown above is of linear classifier i.e., it separates the set of objects into red and green classes with a line. Most of the classification tasks we come across are complex than this one and often more complex decision structures are required to classify them properly i.e., correctly classify new objects (test cases) based on the examples that are available (train cases). This is shown in the image below. AS compared to the last example the complete separation of red and green classes is not possible unless a curve is used. These type of classification tasks which use drawing separating lines are called hyperplanes. Support Vector Machines are particularly suited to handle such tasks.



Fig

Reasons for choosing these classifiers:

* Previously used.
* Gives the option to compare with previous models.
* These use different techniques and so it is reasonable to detect whether different defects are detected by each and whether the prediction consistency is different among the classifiers.

4 Experimental Results

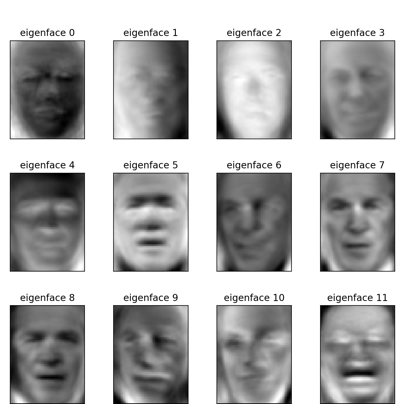
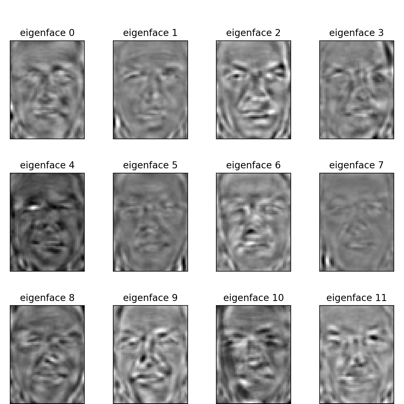
4.1 Methodology

**Evaluation Criteria**: Results are being evaluated based on the accuracy of prediction on training as well as testing data using each of the dimension reduction techniques, namely PCA, ICA and NMF when used in conjunction with the SVM classifier.

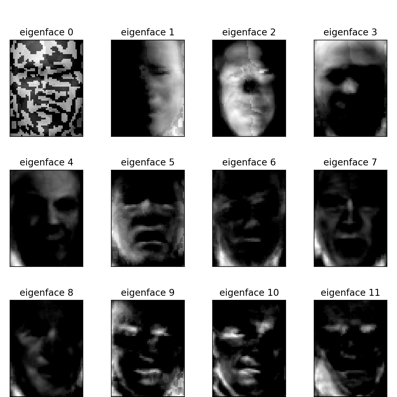
**Data**: Labelled dataset named ‘labelled faces in the wild’ is used which consists of 5 users with more than 100 faces each. Test data is of 12 images.

**Modus Operandi:**

* The training data available at hand was already greyscale. So, it didn’t require any tweaking.
* Dimensionality is reduced using PCA, ICA and NMF. Twelve eigenfaces are created using each of the algorithms on same data.

PCA ICA



NMF

*Fig. Eigenfaces of (a) NMF, (b) PCA and (c) ICA*

As expected, it is quite evident that the eigenfaces generated by PCA captures global features like brightness, avg. eigenfaces and that each next eigen face has less information and more noise. ICA in accordance with theory captured various facial elements in each of its eigenfaces captured the localized features like nose, eye selectors etc. NMF also performed as expected and produced sparsely information rich eigenfaces as per it inherent property due to sparsely positively populated decomposed matrices.

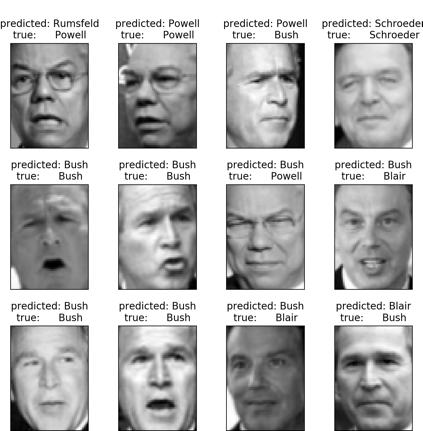
* SVM classifier is fit to this training data using each set of eigenfaces and then accuracy of SVM classifier is compared on the training data and test data based on the accuracy of prediction and time taken for a SVM classifier to be trained on it.

4.2 Results

Dataset of 500 images at hand after being reduced to 12 eigen vectors using PCA, ICA and NMF were compared for their accuracy of prediction on training as well as test data and for the time it took for the SVM classifier to run on them.

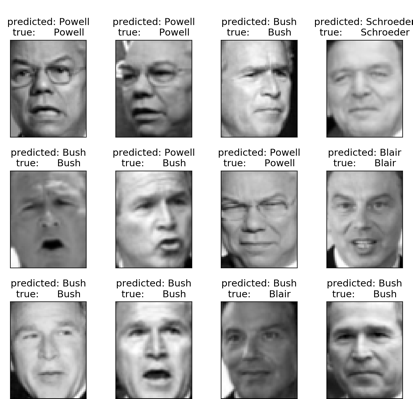
|  |  |
| --- | --- |
| Dimensionality Reduction | Accuracy |
| PCA | 0.90 |
| ICA | 0.87 |
| NMF | 0.57 |

*Table 2: Accuracy on training data*

 A picture containing photo, text, showing, newspaper

Description automatically generated

NMF PCA



ICA

*Fig. Prediction of test images by (a) NMF, (b) PCA and (c) ICA*

|  |  |
| --- | --- |
| Dimensionality Reduction | Accuracy |
| PCA | 0.83 |
| ICA | 0.87 |
| NMF | 0.50 |

*Table: Accuracy on testing data*

|  |  |
| --- | --- |
| Dimensionality Reduction | Time taken by SVM classifier |
| PCA | 17.068 seconds |
| ICA | 14.173 seconds |
| NMF | 25.999 seconds |

*Table: Time taken to fit SVM classifier to training data*

4.3 Discussion

The results show that performance of SVM classifier on training data when trained with dimensionally reduced data using PCA slightly edges ICA which are followed by NMF whereas when it comes to test data PCA and ICA has the same accuracy which is followed by NMF.

These results show that NMF is not a good enough algorithm to be used for facial recognition. PCA and ICA perform almost equally well and can be used to interchangeably according to results, but this second deduction is not in sync with the data in [3, 17]. Also, it took the most time for the SVM classifier to train on NMF generated eigenfaces.

5 Conclusions

Comparison between PCA, ICA and NMF is complex because of differences in underlying tasks, architectures, ICA algorithms and distance metrics must be considered. This paper explores the performance of PCA, ICA and NMF based on their accuracy of prediction on training and testing data when used in conjunction with SVM classifier and suggests that PCA and ICA performs equally well rather PCA edges ICA slightly which is in contrast with the literature which say that ICA edges PCA slightly when it comes to facial identity recognition.

This paper is limited in its scope because it compares the algorithms only on the basis of prediction accuracy and the time it took for the SVM classifier to be trained on the test data. This can be extended in future by finding the factors which influence the performance of algorithms and the time it takes for the classifiers to be trained on these eigenfaces. Numerous variations of these algorithms which are based on varied architectures (in ICA) and the distance matrices (in PCA) can be considered. These can further be explored and compared based on underlying pattern of the data distribution.

References

[1] K. Baek, B. A. Draper, J. R. Beveridge, and K. She, "PCA vs ICA: A comparison on the FERET data set,"

presented at Joint Conference on Information Sciences, Durham, N.C., 2002.

[2] M. S. Bartlett, Face Image Analysis by Unsupervised Learning: Kluwer Academic, 2001.

[3] M. S. Bartlett, G. Donato, J. R. Movellan, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Image

representations for facial expression coding," in Advances in Neural Information Processing Systems, vol.

12. Cambridge, MA: MIT Press, 2000, pp. 886-892.

[4] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face

recognition," presented at SPIE Symposium on Electronic Imaging: Science and Technology; Conference

on Human Vision and Electronic Imaging III, San Jose, CA, 1998.

[5] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component

Analysis," IEEE Transaction on Neural Networks, Vol 13, pp. 1450-1464, 2002.

[6] M. S. Bartlett and T. J. Sejnowski, "Viewpoint invariant face recognition using independent component

analysis and attractor networks," in Neural Information Processing Systems - Natural and Synthetic, vol. 9,

M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 817-823.

[7] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class

Specific Linear Projection," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, pp.

711-720, 1997.

[8] A. J. Bell and T. J. Sejnowski, "An information-maximization Approach to Blind Separation and Blind

Deconvolution," Neural Computation, vol. 7, pp. 1129-1159, 1995.

[9] A. J. Bell and T. J. Sejnowski, "The 'Independent Components' of Natural Scenes are Edge Filters," Vision

Research, vol. 37, pp. 3327-3338, 1997.

[10] J. R. Beveridge, "The Geometry of LDA and PCA Classifiers Illustrated with 3D Examples.," Colorado

State University, web page 2001.

[11] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, "A Nonparametric Statistical Comparison of

Principal Component and Linear Discriminant Subspaces for Face Recognition," presented at IEEE

Conference on Computer Vision and Pattern Recognition, Kauai, HI, 2001.

[12] I. Biederman and P. Kalocsai, "Neurocomputational Bases of Object and Face Recognition," Philosophical

Transactions of the Royal Society: Biological Sciences, vol. 352, pp. 1203-1219, 1997.

[13] W. W. Bledsoe, "The model method in facial recognition," Panoramic Research, Inc., Palo Alto, CA

PRI:15, August 1966.

[14] J.-F. Cardoso, "Infomax and Maximum Likelihood for Source Separation," IEEE Letters on Signal

Processing, vol. 4, pp. 112-114, 1997.

Revised copy

25

[15] X. Chen, L. Gu, S. Z. Li, and H.-J. Zhang, "Learning Representative Local Features for Face Detection,"

presented at IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, 2001.

[16] G. W. Cottrell and M. K. Fleming, "Face recognition using unsupervised feature extraction," presented at

International Neural Network Conference, Dordrecht, 1990.

[17] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. Sejnowski, "Classifying Facial Actions," IEEE

Transactions on Pattern Analysis and Machine Intelligence, vol. 21, pp. 974-989, 1999.

[18] P. Ekman and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial

Movement. Palo Alto, CA: Consulting Psychologists Press, 1978.

[19] B. J. Frey, A. Colmenarez, and T. S. Huang, "Mixtures of Local Linear Subspaces for Face Recognition,"

presented at IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, 1998.

[20] D. Guillamet, M. Bressan, and J. Vitrià, "A Weighted Non-negative Matrix Factorization for Local

Representations," presented at IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI,

2001.

[21] A. Hyvärinen, "The Fixed-point Algorithm and Maximum Likelihood Estimation for Independent

Component Analysis," Neural Processing Letters, vol. 10, pp. 1-5, 1999.

[22] A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis. New York: John Wiley & Sons,

2001.

[23] P. Kalocsai, H. Neven, and J. Steffens, "Statistical Analysis of Gabor-filter Representation," presented at

IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

[24] N. Kambhatla and T. K. Leen, "Dimension Reduction by Local PCA," Neural Computation, vol. 9, pp.

1493-1516, 1997.

[25] J. Karvanen, J. Eriksson, and V. Koivunen, "Maximum Likelihood Estimation of ICA-model for Wide

Class of Source Distributions," presented at Neural Networks in Signal Processing, Sydney, 2000.

[26] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of

Human Faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 103-107,

1990.

[27] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature,

vol. 401, pp. 788-791, 1999.

[28] T.-W. Lee, T. Wachtler, and T. J. Sejnowski, "Color oppency is an efficient representation of spectral

properties in natural scenes," Vision Research, vol. 42, pp. 2095-2103, 2002.

[29] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning Spatially Localized, Parts-Based Representation,"

presented at IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, 2001.

[30] C. Liu and H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face

Recognition," presented at International Conference on Audio and Video Based Biometric Person

Authentication, Washington, D.C., 1999.

[31] A. M. Martinez and A. C. Kak, "PCA versus LDA," IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 23, pp. 228-233, 2001.

[32] B. Moghaddam, "Principal Manifolds and Bayesian Subspaces for Visual Recognition," presented at

International Conference on Computer Vision, Corfu, Greece, 1999.

[33] B. Moghaddam and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition,"

presented at International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

[34] H. Moon and J. Phillips, "Analysis of PCA-based Face Recognition Algorithms," in Empirical Evaluation

Techniques in Computer Vision, K. Boyer and J. Phillips, Eds. Los Alamitos, CA: IEEE Computer Society

Press, 1998.

[35] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for FaceRecognition Algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp.

1090-1104, 2000.

[36] L. Sirovich and M. Kirby, "A Low-dimensional Procedure for the Characterization of Human Faces,"

Journal of the Optical Society of America, vol. 4, pp. 519-524, 1987.

[37] D. Socolinsky and A. Selinger, "A Comparative Analysis of Face Recognition Performance with Visible

and Thermal Infrared Imagery," presented at International Conference on Pattern Recognition, Quebec

City, 2002.

[38] D. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," IEEE Transactions on

Pattern Analysis and Machine Intelligence, vol. 18, pp. 831-836, 1996.

[39] M. E. Tipping and C. M. Bishop, "Mixtures of Probabilistic Principal Component Analysers," Neural

Computation, vol. 11, pp. 443-482, 1999.

Revised copy

26

[40] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol. 3, pp. 71-

86, 1991.

[41] P. C. Yuen and J. H. Lai, "Independent Component Analysis of Face Images," presented at IEEE Workshop

on Biologically Motivated Computer Vision, Seoul, 2000.

[42] M. Zibulevsky and B. A. Pearlmutter, "Blind Separation of Sources with Sparse Representation in a Given

Signal Dictionary," presented at International Workshop on Independent Component Analysis and Blind

Source Separation, Helsinki, 2000.

[43] A. Ziehe, G. Nolte, T. Sander, K.-R. Müller, and G. Curio, "A Comparison of ICA-based Artifact

Reduction Methods for MEG," presented at 12th International Conference on Biomagnetism, Espoo,

Finland, 2000