Extracting Discriminative Information from Medical Images: A Multivariate Linear Approach

Carlos E. Thomaz[1], Nelson A.O. Aguiar[1], Sergio H.A. Oliveira[1], Fabio L.S. Duran[2], Geraldo F. Busatto[2], Duncan F. Gillies[3], and Daniel Rueckert[3]
[1]*Department of Electrical Engineering, Centro Universitario da FEI, São Paulo, Brazil*
[2]*Departments of Psychiatry and Radiology, Faculty of Medicine, University of São Paulo, Brazil*
[3]*Department of Computing, Imperial College, London, UK*
cet@fei.edu.br

Abstract

Statistical discrimination methods are suitable not only for classification but also for characterisation of differences between a reference group of patterns and the population under investigation. In the last years, statistical methods have been proposed to classify and analyse morphological and anatomical structures of medical images. Most of these techniques work in high-dimensional spaces of particular features such as shapes or statistical parametric maps and have overcome the difficulty of dealing with the inherent high dimensionality of medical images by analysing segmented structures individually or performing hypothesis tests on each feature separately. In this paper, we present a general multivariate linear framework to identify and analyse the most discriminating hyper-plane separating two populations. The goal is to analyse all the intensity features simultaneously rather than segmented versions of the data separately or feature-by-feature. The conceptual and mathematical simplicity of the approach, which pivotal step is spatial normalisation, involves the same operations irrespective of the complexity of the experiment or nature of the data, giving multivariate results that are easy to interpret. To demonstrate its performance we present experimental results on artificially generated data set and real medical data.

1. Introduction

In the generic discrimination problem, where the training sample consists of the class membership and observations for N patterns, the outcome of interest fall into g classes and we wish to build a rule for predicting the class membership of an observation based on n variables or features. However, statistical discrimination methods are suitable not only for classification but also for characterisation of differences between a reference group of patterns and the population under investigation. For example, in clinical diagnosis we might want to understand underlying causes of medical data by exploring the discriminating hyper-plane found by a statistical classifier using image samples of patients and controls.

In the last years, statistical pattern recognition methods have been proposed to classify and analyse morphological and anatomical structures of magnetic resonance (MR) images [4, 6, 8]. Most of these techniques work in high-dimensional spaces of particular features such as shapes or statistical parametric maps and have overcome the difficulty of dealing with the inherent high dimensionality of medical data by analysing segmented structures individually or performing hypothesis tests on each feature separately. Unfortunately, in such approaches changes that are relatively more distributed and involve simultaneously several structures of the pattern of interest (i.e., ventricles and corpus callosum of the brain) might be difficult to detect, despite the possibility of some methods [6, 8] of extracting statistically multivariate differences between image samples of patients and controls.

In this work, we present a general multivariate statistical framework to identify and analyse the most discriminating hyper-plane separating two populations. The goal is to analyse all the intensity features simultaneously rather than segmented versions of the data separately or feature-by-feature. We use a novel method proposed recently [10], called Maximum uncertainty Linear Discriminant Analysis (MLDA), to overcome the well-known instability of the within-class scatter matrix in limited sample size problems and to increase the computational efficiency of the approach. The approach is not restricted to any particular set of

features and describes a simple and straightforward way of mapping multivariate classification results of the whole images back into the original image domain for further interpretation.

The remainder of this paper is divided as follows. In section 2 we describe the main parts of the multivariate linear framework and its design. This section includes a brief review of Principal Component Analysis (PCA) and the novel MLDA method used. Section 3 presents experimental results of the approach and demonstrates its effectiveness on a simple artificially generated data set and on a real medical data. In the last section, section 4, the paper concludes with a short summary of functionalities that form the basis for this methodology of discriminating and analysing the patterns of interest.

2. A Multivariate Linear Approach

Our main concern here is to describe a multivariate framework that highlights the most discriminating differences between two populations when the number of examples per class is much less than the dimension of the original feature space. This problem is indeed quite common nowadays, especially in medical image analysis. For instance, patients and controls are classes defined commonly by a small number of MR images but the features used for recognition may be millions of voxels or hundreds of pre-processed image attributes.

2.1. Principal Component Analysis (PCA)

There are a number of reasons for using PCA to reduce the dimensionality of the original images. PCA is a linear transformation that is not only simple to compute and analytically tractable but also extracts a set of features that is optimal with respect to representing the data back into the original domain. Moreover, using PCA as an intermediate step will reduce dramatically the computational and storage requirements for the subsequent LDA-based covariance method. Since in our applications of interest the number of training patterns N (or images) is much smaller than the number of features n (or instance: voxels), it is possible to transform data in a way that patterns occupy as compact regions in a lower dimensional feature space as possible with far fewer degrees of freedom to estimate.

Although much of the sample variability can be accounted for by a smaller number of principal components, and consequently a further dimensionality reduction can be accomplished by selecting the principal components with the largest eigenvalues, there is no guarantee that such additional dimensionality reduction will not add artefacts on the images when mapped back into the original image space. Our aim is to map the classification results back to the image domain for further visual interpretation. For that reason, we must be certain that any modification on the images, such as blurring or subtle differences, is not related to an "incomplete" or perhaps "misleading" feature extraction intermediate procedure. For example, Figure 1 illustrates on the top a reference image (shown on the left) reconstructed using several principal components and on the bottom the corresponding differences between these reconstructions to the original image. The values in parentheses represent the number of principal components used and corresponding total variance explained. We can see clearly that even when we use a set of principal components that represents more than 90% of the total sample variance we still have subtle differences between the reconstructed image and the original one.

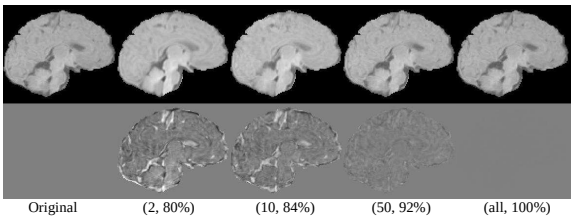

Fig. 1. Reconstruction of a reference image (shown on the top left) using several principal components. The row on the bottom illustrates the corresponding differences between the reconstructions to the reference image. The number of components retained and the corresponding total sample variance explained are shown in parentheses. We can see modifications on the reconstructed images where all principal components with non-zero eigenvalues are not selected.

Therefore, in order to reproduce the total variability of the samples we have composed the PCA transformation matrix by selecting all principal components with non-zero eigenvalues. To avoid the high memory rank computation of the possibly large total covariance matrix and because the MLDA approach deal with the singularity of the within-class scatter matrix, we have assumed that all the N training patterns are linearly independent. In other words, we have assumed that the rank of the total covariance matrix is $N-1$ and the number of PCAs selected is $m = N-1$.

2.2. Maximum Uncertainty LDA (MLDA)

The primary purpose of LDA is to separate samples of distinct groups by maximising their between-class separability while minimising their within-class variability. LDA's main objective is to find a projection matrix P_{lda} that maximizes the following ratio (Fisher's criterion):

$$P_{lda} = \arg\max_{P} \frac{\left| P^T S_b P \right|}{\left| P^T S_w P \right|}, \quad \textbf{(1)}$$

where S_b is the between-class scatter matrix defined as

$$S_b = \sum_{i=1}^{g} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad \textbf{(2)}$$

and S_w is the within-class scatter matrix defined as

$$S_w = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T . \quad \textbf{(3)}$$

The vector $x_{i,j}$ is the *n*-dimensional pattern j from class π_i, N_i is the number of training patterns from class π_i, and g is the total number of classes or groups. The vector $\bar{x}_i$ and matrix S_i are respectively the unbiased sample mean and sample covariance matrix of class π_i [5]. The grand mean vector $\bar{x}$ is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{g} N_i \bar{x}_i = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} x_{i,j} , \quad \textbf{(4)}$$

where N is the total number of samples, that is, $N = N_1 + N_2 + \cdots + N_g$. The Fisher's criterion described in equation (1) is maximised when the projection matrix P_{lda} is composed of the eigenvectors of $S_w^{-1} S_b$ with at most $(g-1)$ nonzero corresponding eigenvalues. This is the standard LDA procedure.

It is well known, however, that the performance of the standard LDA can be seriously degraded if there are only a limited number of total training observations N compared to the dimension of the feature space m. Since the within-class scatter matrix S_w is a function of $(N-g)$ or less linearly independent vectors, where g is the number of groups, its rank is $(N-g)$ or less. Therefore in recognition problems where the number of training patterns is comparable to the number of features, S_w might be singular or mathematically unstable and the standard LDA cannot be used to perform the task of the classification stage.

In order to avoid both the singularity and instability critical issues of the within-class scatter matrix S_w when LDA is used in limited sample and high dimensional problems, we have used a maximum uncertainty LDA-based approach (MLDA) based on a straightforward covariance selection method for the S_w matrix. In an earlier study [10], Thomaz and Gillies compared the performance of MLDA with other recent LDA-based methods, such as Chen et al.'s LDA [2], direct LDA [14], and Optimal Fisher Linear Discriminant [13], with application to the face recognition problem. Since the face recognition problem involves small training sets, a large number of features, and a large number of groups, it has become the most used application to evaluate such limited sample size approaches. The experimental results carried out have shown that the MLDA method improved the LDA classification performance with or without an intermediate dimensionality reduction and using less linear discriminant features.

The MLDA algorithm can be shortly described as follows:

i. Find the Φ eigenvectors and Λ eigenvalues of S_p, where $S_p = S_w / [N-g]$;

ii. Calculate the S_p average eigenvalue $\bar{\lambda}$, that is,

$$\bar{\lambda} = \frac{1}{n} \sum_{j=1}^{n} \lambda_j = \frac{trace(S_p)}{n}; \quad \textbf{(5a)}$$

iii. Form a new matrix of eigenvalues based on the following largest dispersion values

$$\Lambda^* = diag[\max(\lambda_1, \bar{\lambda}), ..., \max(\lambda_n, \bar{\lambda})]; \quad \textbf{(5b)}$$

iv. Form the modified within-class scatter matrix

$$S_w^* = S_p^* (N-g) = (\Phi \Lambda^* \Phi^T)(N-g) . \quad \textbf{(5c)}$$

The maximum uncertainty LDA (MLDA) is constructed by replacing S_w with S_w^* in the Fisher's criterion formula described in equation (1). As pointed out by Thomaz and Gillies [TG05], it is based on a maximum entropy covariance selection idea developed to improve the performance of Bayesian classifiers on limited sample size problems [11].

2.3. Framework Design

We can divide the design of the PCA+MLDA multivariate framework into two main tasks: classification (training and test stages) and visual analysis.

In the classification task the principal components and the maximum uncertainty linear discriminant vector are generated. As illustrated in Figure 2, first a training set is selected and the average image vector of all the training images is calculated and subtracted from each pre-processed image vector. Then the descending order. Recall, from section 2.1, that we have retained all the PCA eigenvectors with non-zero eigenvalues. The zero mean image vectors are projected on the principal components and reduced to *m*-dimensional vectors representing the most expressive features of each one of the pre-processed *n*-

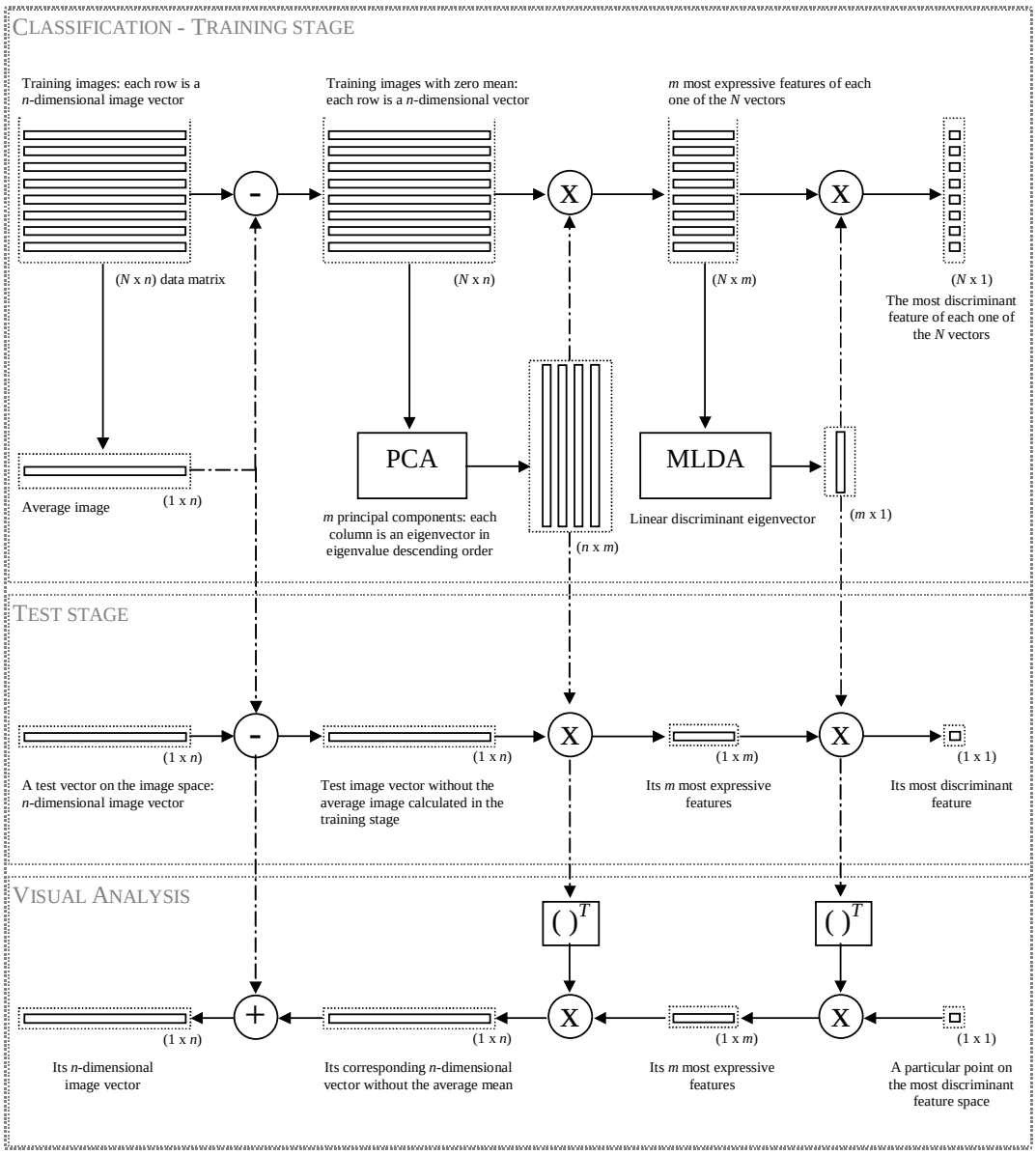

Fig. 2. Design of the multivariate linear framework.

training matrix composed of zero mean image vectors is used as input to compute the PCA transformation matrix. The columns of this *n* x *m* transformation matrix are eigenvectors, not necessarily in eigenvalues

dimensional image vector. Afterwards, the *N* x *m* data matrix is used as input to calculate the MLDA discriminant eigenvector. Since we are assuming only two classes to separate, there is only one MLDA

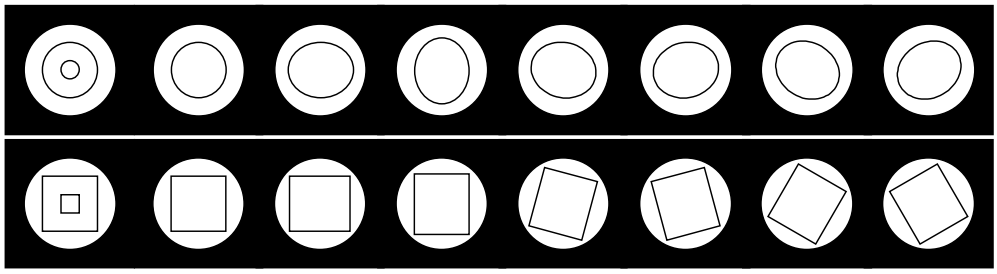

Fig. 3. A synthetic data set.

discriminant eigenvector. The most discriminant feature of each one of the *m*-dimensional vectors is obtained by multiplying the $N \times m$ most expressive features matrix by the $m \times 1$ MLDA linear discriminant eigenvector. Thus, the initial pre-processed training set consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on only 1 most discriminant feature.

The other main task that can be implemented by this two-stage multivariate statistical approach is to visually analyse the most discriminant feature found by the maximum uncertainty method. According to Figure 2, more specifically from right to left in its Visual Analysis frame, any point on the most discriminant feature space can be converted to its corresponding *n*-dimensional image vector by simply: (1) multiplying that particular point by the transpose of the linear discriminant vector previously computed; (2) multiplying its *m* most expressive features by the transpose of the principal components matrix; and (3) adding the average image calculated in the training stage to the *n*-dimensional image vector. Therefore, assuming that the clouds of the classes follow a multidimensional Gaussian distribution and applying limits to the variance of each cloud, such as $\pm 3s$, where s is the standard deviation of each group, we can move along this most discriminant feature and map the result back into the image domain. This mapping procedure provides an intuitive interpretation of the classification experiments and, as we will show in the experimental results of real medical data, biologically plausible results that are often not detectable simultaneously.

3. Experimental Results

To illustrate the performance of the multivariate linear approach we present in this section experimental results of the framework based on a simple artificially generated data set and on a real medical data.

3.1. A Synthetic Data Example

We have chosen a very simple artificial data set composed of 8 binary images of circles (ellipses) and 8 binary images of squares (rectangles). Figure 3 shows both samples of images composed of 70 x 70 pixels.

As described in the previous section, we use such training examples (without any spatial normalisation) to construct the multivariate linear classifier for labelling new examples and identifying the most discriminating hyper-plane separating circles (or ellipses) from squares (or rectangles). Since those samples are very simple and easily separable, the classifier achieved 100% of leave-one-out accuracy. Figure 4 presents the PCA+MLDA most discriminant feature of the synthetic database using all the 16 examples as training images. It displays the image regions captured by the classifier that change when we move from one side (squares or rectangles) of the

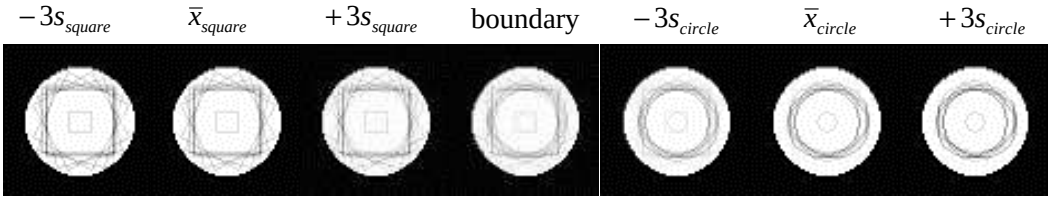

Fig.4. Image display of the regions captured by the classifier that change when we move from one side (squares or rectangles) of the dividing PCA+MLDA hyper-plane to the other (circles or ellipes), following limits of ± 3 standard deviations for each sample group.

dividing hyper-plane to the other (circles or ellipes), following limits to the variance ($\pm 3s$ standard deviations) of each sample group.

Despite the changes due to misalignments of the images, Figure 4 shows clearly that the statistical mapping effectively extracts the group differences. It is important to note that these differences could be very subtle on samples that are very close to the dividing boundary and consequently difficult to characterise as belonging to one of the groups.

3.2. A Real Data Example

In order to demonstrate the effectiveness of the methodology on medical data, we have used an Alzheimer MR brain data set that contains images of 14 patients and 14 healthy controls. All these images were acquired using a 1.5T Philips Gyroscan S15-ACS MRI scanner (Philips Medical Systems, Eindhoven, The Netherlands), including a series of contiguous 1.2mm thick coronal images across the entire brain, using a T1-weighted fast field echo sequence (TE = 9ms, TR = 30ms, flip angle 30^o, field of view = 240mm, 256 x 256 matrix). All images were reviewed by a MR neuro-radiologist. Ethical permission for this study was granted by the Ethics Committee of the Clinical Hospital, University of Sao Paulo Medical School, Sao Paulo, Brazil.

3.2.1. Mass-univariate Statistical Analysis

For comparison purpose, Statistical Parametric Mapping (SPM, version SPM2) [4] analyses were conducted using an optimised Voxel-based Morphometry (VBM) protocol [7]. In contrast to the multivariate approach, SPM has been designed to enable voxel-by-voxel inferences about localised differences between the groups and, consequently, does not characterise interregional dependencies between the structures of the brain [3].

A standard template set selected by the psychiatrists was created specifically for this study, consisting of a mean T1-weighted image, and *a priori* gray matter, white matter and CSF templates based on the images of all AD (Alzheimer Disease) and healthy control subjects. Initially, images were spatially normalized to the standard SPM T1-MRI template [9], using linear 12-parameter affine transformations. Spatially normalized images were then segmented into gray matter, white matter and cerebrospinal fluid (CSF) compartments, using a modified mixture model cluster analysis technique [7]. The segmentation method also included: an automated brain extraction procedure to remove non-brain tissue and an algorithm to correct for image intensity non-uniformity. Finally, images were smoothed with an isotropic Gaussian kernel (8mm FWHM), and averaged to provide the gray, white matter and CSF templates in stereotactic space.

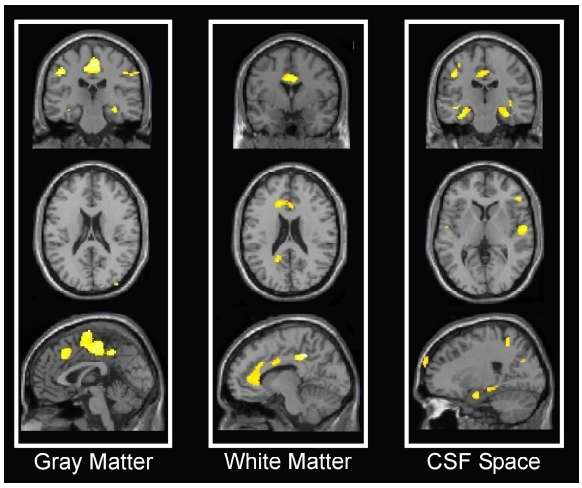

Fig. 5. Brain regions where significant differences in Alzheimer patients relatively to controls were detected by the SPM voxel-wise statistical tests at $p < 0.01$. We can see between-group differences in the occipital, parietal and frontal lobes, inter-hemispheric fissure, and corpus callosum.

To boost the signal-to-noise ratio, the image processing of the original images from all AD patients and controls was then carried out, beginning by image segmentation. The segmented images were spatially normalized to the customized templates previously created by using 12-parameter linear as well as non-linear (7 x 9 x 7 basis functions) transformations. The parameters resulting from this spatial normalization step were reapplied to the original structural images. These fully normalized images were re-sliced using tri-linear interpolation to a final voxel size of 2 x 2 x 2 mm^3, and segmented into gray matter, white matter and CSF partitions. Voxel values were modulated by the Jacobian determinants derived from the spatial normalisation, thus allowing brain structures that had their volumes reduced after spatial normalisation to have their total counts decreased by an amount proportional to the degree of volume shrinkage [7]. Finally, images from AD patients and controls were smoothed using a 12mm Gaussian kernel and compared statistically between the two groups using unpaired Student's t-tests at $p < 0.01$ (level of significance).

Figure 5 illustrates the locations where significant differences between the groups were detected. The underlying image is the reference template used in the

spatial normalisation of all MR images. As can be seen, there are some localised differences in the Alzheimer patients relatively to controls in the occipital, parietal and frontal lobes, in the inter-hemispheric fissure, and corpus callosum. These structures, especially where significant gray matter changes were observed, are among the regions thought to be the most prominently affected by atrophic changes in Alzheimer disease [1].

3.2.2. Multivariate Statistical Analysis

Evaluating the classifier's performance. In order to evaluate the PCA+MLDA classification's rule, we have used the Bhattacharyya bound to estimate the error probability of the multivariate statistical framework.

For two-class problems, the upper bound of the error probability e_u is defined as [5]

$$e_u = (p_1 p_2)^{1/2} \exp(-d), \qquad \textbf{(6)}$$

where p_1 and p_2 are the prior probabilities of classes π_1 and π_2 respectively, and d is the Bhattacharyya distance between the two classes defined as

$$d = \frac{1}{8}(\bar{x}_1 - \bar{x}_2)^T \left(\frac{S_1 + S_2}{2}\right)^{-1} (\bar{x}_1 - \bar{x}_2) + \frac{1}{2}\ln \frac{\left|\frac{S_1 + S_2}{2}\right|}{\sqrt{|S_1||S_2|}}, \qquad \textbf{(7)}$$

where the notation "|.|" denotes the determinant of a matrix. As described previously, the vector $\bar{x}_i$ and matrix S_i are respectively the unbiased sample mean and sample covariance matrix of class π_i ($i = 1,2$).

Since the dataset under investigation comes with the same proportion of patient images relatively to controls, we have assumed that the prior probabilities of both groups are equal. Thus, assuming $p_1 = p_2 = 0.5$ and calculating the Bhattacharyya distance d using all the patient and control samples, the multivariate statistical classifier achieves the value of 1.56%. This result confirms the classifier's ability of discriminating the brains of controls from those of patients with a successful classification rate of 98.44%, using the closed-form method for the error probability.

Visual Analysis of discriminative information. The visual analysis of the linear discriminant feature found by the multivariate approach is summarised in Figure 6. As mentioned earlier, the one-dimensional vector found by the PCA+MLDA approach corresponds to a hyper-plane on the original image space which direction describes statistically the most discriminant differences between the control and patient images used for training.

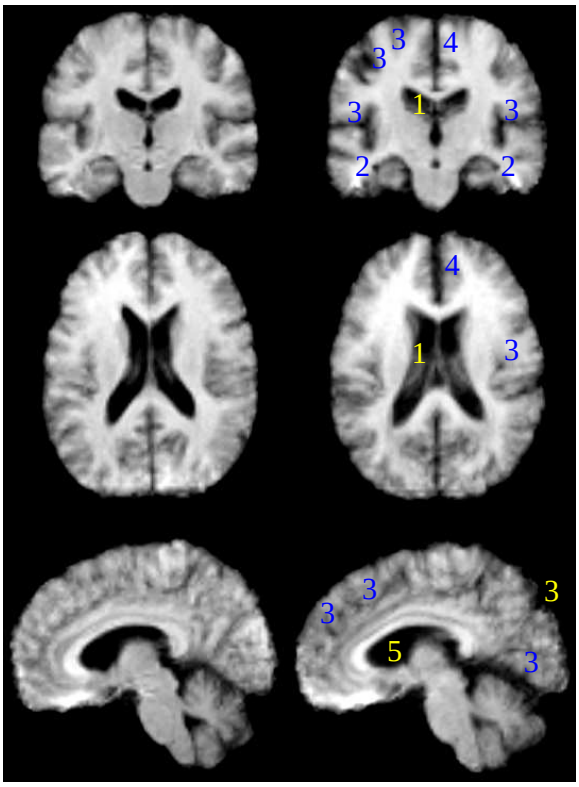

Fig. 6. Statistical differences between the control (on the left) and Alzheimer patient (on the right) images captured by the multivariate statistical classifier. We can see the following brain differences in the Alzheimer patients relatively to the controls: (1) enlargement of the ventricular system, (2) atrophy of the hippocampus, (3) cortical degeneration of the occipital, parietal, and frontal lobes, (4) enlargement of the inter-hemispheric fissure, and (5) atrophy of corpus callosum.

Figure 6 shows the differences between the control (on the left column) and patient (on the right column) images captured by the multivariate statistical classifier using MR intensity features as inputs. These images correspond to one-dimensional points on the PCA+MLDA space projected back into the image domain and located at 3 standard deviations of each sample group. We can understand this mapping procedure as a way of defining intensity changes that come from "definitely control" and "definitely patient" samples captured by the statistical classifier. We can see the following brain differences in the Alzheimer patients relatively to the controls: (1) enlargement of the ventricular system, (2) atrophy of the hippocampus, (3) cortical degeneration of the occipital, parietal, and frontal lobes, (4) enlargement of the inter-hemispheric

fissure, and (5) atrophy of the corpus callosum. These multivariate results are consistent with the SPM between-group differences presented previously and with other common findings of patients who have developed the pathology [12], such as the enlargement of the ventricular system. Therefore, the use of the multivariate approach has allowed not only the simultaneous identification of localised between-group differences but also distributed ones that are often measured separately in the voxel-wise statistical approaches.

4. Conclusion

We have presented a general PCA+MLDA multivariate linear framework to identify and analyse the most discriminating hyper-plane separating two populations. The statistical analysis generates a detailed description of the neuroanatomical changes due to diseases and can facilitate the studies of the brain disorders, such as Alzheimer, through understanding of the captured anatomical changes.

The idea of using PCA plus an LDA-based approach to discriminate patterns of interest is not new. In this paper we have added to the functionality of this approach the following important points for medical image analysis. The use of full rank version of PCA transformation matrix that allows valuable low representation of high dimensional data, providing optimal reconstruction of the most discriminant intensity features without adding any artefacts on the patterns when mapped back into the original image space. By selecting a slightly biased within-class scatter matrix composed of the most informative dispersions we resolve not only the LDA singularity problem but also we stabilise the maximisation of the Fisher's criterion on limited sample size problems. The conceptual and mathematical simplicity of the approach, which pivotal step is spatial normalisation, involves the same operations irrespective of the complexity of the experiment or nature of the data, giving multivariate results that are easy to interpret.

Although the approach has been demonstrated in two-class problems, it is extensible to several classes. The only difference is the visual analysis of the discriminant features, which might be performed pair-wisely.

References

[1] G. F. Busatto, G. E. J. Garrido, O. P. Almeida, C. C. Castro, C. H. P. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and C. M. Bottino, "A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease", *Neurobiology of Aging*, 24, pp. 221-231, 2003.

[2] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem", *Pattern Recognition*, 33 (10), pp. 1713-1726, 2000.

[3] K. J. Friston and J. Ashburner, "Generative and recognition models for neuroanatomy", *NeuroImage*, vol. 23, pp. 21–24, 2004.

[4] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical Parametric Maps in Functional Imaging: A General Linear Approach", *Human Brain Mapping* 2, pp. 189–210, 1995.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second edition. Boston: Academic Press, 1990.

[6] P. Golland, W. Grimson, M. Shenton, and R. Kikinis, "Detection and Analysis of Statistical Differences in Anatomical Shape", *Medical Image Analysis*, vol. 9, pp. 69-86, 2005.

[7] C. D. Good, I. S. Johnsrued, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains, *NeuroImage*, vol. 14, pp. 21-36, 2001.

[8] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods", *NeuroImage*, vol. 21, pp. 46-57, 2004.

[9] J.C. Mazziotta, A.W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: Theory and rationale for its development", *NeuroImage*, vol. 2, pp. 89-101, 1995.

[10] C. E. Thomaz and D. F. Gillies, "A Maximum Uncertainty LDA-based approach for Limited Sample Size problems - with application to Face Recognition", in Proceedings of SIBGRAPI'05, IEEE CS Press, pp. 89-96, 2005.

[11] C. E. Thomaz, D. F. Gillies and R. Q. Feitosa, "A New Covariance Estimate for Bayesian Classifiers in Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 214-223, 2004.

[12] P. M. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, A. Khan, M.S. Mega, G. Small, J. Cummings, and A. W. Toga, "Cortical variability and asymmetry in normal aging and Alzheimer's disease", *Cerebral Cortex*, vol. 8, pp. 492-509, 1998.

[13] J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space? ", *Pattern Recognition*, vol. 36, pp. 563-566, 2003.

[14] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data – with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067-2070, 2001.