Learning high-dimensional image statistics for abnormality detection on medical images

Guray Erus[1,*], Evangelia I. Zacharaki[1,2,*], Nick Bryan[1,*], and Christos Davatzikos[1,*]

[1]Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, 19104 Philadelphia, PA, USA

[2]Biosignal Processing Group, University of Patras, School of Medicine, 26504 Rio, Greece

[*]{Guray.Erus, Eva.Zacharaki, R.Nick.Bryan, Christos.Davatzikos}@uphs.upenn.edu

Abstract

We present a general methodology that aims to learn multi-variate statistics of high dimensional images, in order to capture the inter-individual variability of imaging data from a limited number of training images. The statistical learning procedure is used for identifying abnormalities as deviations from the normal variation. In most practical applications, learning an accurate statistical model of the observed data is a very challenging task due to the very high dimensionality of the images, and the limited number of available training samples. We attempt to overcome this problem by capturing the statistics of a large number of lower dimensional subspaces, which can be estimated more reliably. The subspaces are derived in a multi-scale fashion, and capture image characteristics ranging from fine and localized to coarser and relatively more global. The main premise is that an imaging pattern that is consistent with the statistics of a large number of subspaces, each reflecting a marginal probability density function (pdf), is likely to be consistent with the overall pdf, which hasn't been explicitly estimated. Abnormalities in a new image are identified as significant deviations from the normal variation captured by the learned subspace models, and are determined via iterative projections on these subspaces.

1. Introduction

In medical image analysis, detection of abnormalities, i.e. imaging patterns that do not conform to the normal behavior of anatomy, is the major problem in many applications [1, 14, 8, 18]. A common approach for detecting abnormalities is to use supervised learning techniques, assuming that labeled training data for the normal and abnormal classes are available. These techniques learn a predictive model from the positive and negative training samples, which is then used for classifying a new sample as normal or abnormal. Recent kernel-based methods, in particular Support Vector Machines (SVMs), are known to obtain a high classification performance on several clinical problems [17].

The supervised learning approach has two major limitations. First, the learned model is specific to a given type of pathology. This means that one might need to train classifiers for many possible types of abnormalities, which is often not practical or even possible, particularly when the pathology is variable and/or unknown in advance. Second, these methods are generally trained on local intensity-based features, and they disregard the spatial information that might be very informative. For example, a brain lesion might have certain image intensity characteristics, but its spatial/shape properties also carry important information.

In this paper, we present a more general framework in order to overcome the limitations of the aforementioned approaches. The proposed method aims to learn the inter-individual variability of the healthy anatomy, and to detect abnormalities as deviations from the normality. It doesn't make any assumption about the characteristics of the abnormality, however it assumes that normal anatomy is relatively consistent and its statistics can be estimated from a number of representative examples. We focus on overcoming limitations of such direct estimation of statistical variation, which is particularly challenging in high-dimensional spaces and from limited sample sizes. As in [4] the learning model is trained exclusively using images of healthy subjects (referred as "normal images" through the paper).

The idea of learning a statistical model that describes an object category from only the normal samples has been successfully applied in a number of approaches, including the popular shape and appearance learning models. In Active

Shape Models method [2], an object is represented by a set of landmark points, and a shape model that estimates the locations of these landmark points is learned via Principal Component Analysis (PCA). However, learning multivariate statistics of very high dimensional data is a very challenging task. Accurately estimating the *pdf* of a 3D image of the entire brain could require tens of thousands of training images, if not more. In practical applications, the number of training samples available to be used for estimating the *pdf* is very small relative to the dimensionality of the image domain. Consequently, the direct application of classical subspace projection methods on the whole image domain performs poorly and can only capture the global variations of the data. Partitioning the image domain into lower dimensional subspaces is a common approach against the limitation of PCA in representing finer image details. In block-PCA (also called modular PCA) method an image is divided by a regular grid into smaller blocks and PCA is applied on each block independently [5]. A similar approach is 2D PCA [16], where PCA is applied on a 2D image matrix rather than a 1D image vector. In Wavelet Block-PCA approach [3, 15], the image is decomposed into multiple frequency subbands on which PCA is applied. In [7] a spectral graph partitioning technique is applied for partitioning the wavelet coefficients into correlated clusters.

Instead of imposing an ad-hoc partitioning scheme on the image domain, we propose to extract a large number of lower dimensional subspaces, each of which herein represent image patches, albeit they don't necessarily have to. We impose conditions that allow the statistical variations of these subspaces to be estimated from the available training data. For each subspace, a statistical model, which is used for estimating the marginal *pdf*, is learned from the training samples. Through an iterative procedure, which we call Statistical Model (SM)-constrained reconstruction, a new test image is modified such that it will be consistent with all subspace models. This procedure allows us to detect image regions that deviate significantly from the normal variation, and provides us a general framework to detect any kind of abnormalities on high dimensional images.

The remainder of the paper is organized as follows: Section 2 describes the proposed method. The experimental results are given and discussed in section 3. Section 4 summarizes and concludes with additional discussions and future perspectives.

2. Method

2.1. Formulation of the Problem

Consider images of a specific anatomic part coregistered to the template image domain Ω as realizations of a d-dimensional random vector I, consisting of d scalar random variables $[x_1, x_2, ..., x_d]$ corresponding to image voxels. The joint probability density function *pdf* of I

$$\phi(I) = \phi(x_1, x_2, ..., x_d) \quad (1)$$

describes the relative likelihood for I to be observed. In the d dimensional space, the set of normal images, i.e. images for which $\phi(I) >= t$ (where t is a predetermined likelihood value above which an image is considered normal), constitutes a hypervolume. An image I^0 that contains abnormalities is expected to have a low likelihood. By moving I^0 in the direction that maximizes its likelihood, an estimated normal image I^* for which $\phi(I^*) = t$ can be obtained (Figure 1.a). Moving I^0 towards I^* should have the effect of removing the abnormalities, and the voxelwise difference between I^0 and I^* would correspond to a probability map of the abnormalities on I^0.

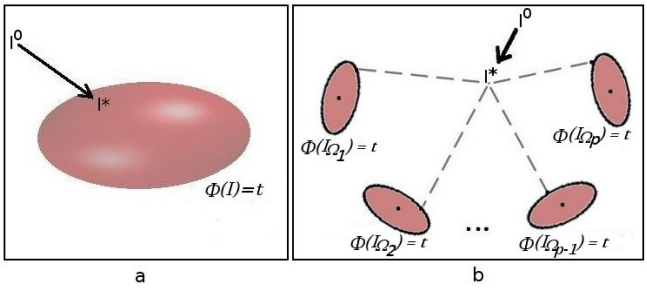

Figure 1. Illustration of the SM-constrained reconstruction of the healthy anatomy from an image containing abnormalities. a. for the case where the *pdf* of the healthy anatomy is known. b. for the case where the *pdf* is estimated through subspace models

However, in most cases, the *pdf* of the healthy anatomy is unknown, while only a small set of training samples is available for estimating it. In most practical applications, the number of the training samples n is very small compared to d. This is an important limitation for estimating the *pdf* of high dimensional data accurately.

While $\phi(I)$ can not be estimated accurately, *pdf*s of lower dimensional subspaces can be estimated more reliably (if the subspace is small enough, variation of the image projection on it can be estimated from a given dataset). Thus, in order to overcome the dimensionality problem, we propose to sample a large number of subspaces $\Omega_1, ..., \Omega_p$ from Ω. Our main premise is that an imaging pattern that is consistent with the statistics of a large number of subspaces, each reflecting a marginal *pdf*, is likely to be consistent with the overall *pdf*, which hasn't been explicitly estimated.

An image I^0 containing pathologies can be constrained to be normal by moving it in the d dimensional space such that the likelihood of the new image I^* will be high when projected to each of the subspaces (Figure 1.b). The likelihood within a subspace Ω_i can be calculated by estimating $\phi(I^*_{\Omega_i})$, the marginal *pdf* corresponding to the image patch with voxels within Ω_i, using the set of n training samples.

This reconstruction procedure can be formulated as an optimization problem in which the following energy function is minimized:

$$E(I^*) = E_{SIM}(I^*, I) + \lambda E_{SM}(I^*) \quad (2)$$

where the energy term E_{SIM} reflects the image similarity between the test image and the SM-reconstructed image, and

$$E_{SM} = \sum_{i=1}^{p} L(I^*_{\Omega_i}) \quad (3)$$

is the energy term reflecting the overall likelihood of the SM-reconstructed image according to the statistical model of each subspace. Here L is the likelihood function that returns the likelihood of $I^*_{\Omega_i}$ according to the statistical model of the i^{th} subspace.

2.2. Extraction of Subspaces

Our main objective while extracting lower dimensional subspaces from the image domain is to estimate the *pdf* more reliably. The estimability of a subspace can be considered as a measure of its intrinsic dimensionality, which is determined by the size of the subspace, in conjunction with the variation of the image projection onto that subspace. Due to the high correlation among neighboring voxels in natural images, the intrinsic dimensionality of image patches consisting of a set of neighboring voxels is generally much lower than their actual dimensionality.

Accordingly, we select subspaces as image patches consisting of voxels in a neighborhood $s = [s_x, s_y, s_z]$ around a voxel at position $\{x, y, z\}$, for different values of x, y, z and s. In that way, the set of all subspaces capture image characteristics ranging from fine and localized to coarser and relatively more global at different spatial locations of the image domain.

A subspace with a low estimability means that the statistical model learned from the training samples is not capable to capture the variation of the imaging pattern on this subspace. An estimability score is calculated for each subspace in order to exclude non-estimable subspaces. This score depends on the selected modeling approach and is defined in section 2.4.

2.3. Iterative Reconstruction Method

As the number of all estimable subspaces may be very large for high dimensional images, we propose an iterative procedure that provides an approximate solution to the optimization problem defined in equation 2. The principal idea is to constrain I to be consistent with the statistical model of one subspace at a time, and to iterate over the subspaces until the algorithm converges to a stable state where the SM-reconstructed image jointly satisfies the constraints over all subspaces. Algorithm 1 briefly presents an overview of the algorithmic procedure.

Algorithm 1 Global algorithm of the iterative reconstruction method

$I^* \leftarrow I$
while I^* has not converged **do**
  Randomly select a subspace Ω_{Sel}
  while Ω_{Sel} is non-estimable **do**
    Select a new subspace Ω'_{Sel}
    $\Omega_{Sel} \leftarrow \Omega'_{Sel}$
  end while
  Learn the statistical model for Ω_{Sel}
  Modify $I^*_{\Omega_{Sel}}$ such that it becomes consistent with the learned model
  Update I^*
end while

Return I^*

When a selected subspace is not estimable, two different strategies are applied for selecting a new subspace:

1. Ignore the subspace and select a new one randomly

2. Decrease the size of the subspace until the new subspace becomes estimable. A number of different ways can be pursued here, including reducing the size of a patch, randomly sampling voxels from the patch, or applying a low-pass filter to estimate only lower-frequency characteristics of the patch. We have applied the first approach that provides a more adaptive selection strategy where the largest estimable neighborhood around $\{x, y, z\}$ is detected.

2.4. PCA Model Within a Subspace

Various modeling approaches may be used for estimating the *pdf* within a subspace, and for the reconstruction of an image patch. We applied PCA, a popular data analysis method which calculates a number of principal components that are frequently called principal modes of variation, and which is based on the assumption that the data follow a Gaussian distribution.

Let $\mathbf{Y_i} \in \Re^{k \times n}$ be the data matrix containing k dimensional data vectors $I^j_{\Omega_i}, j = 1, ..., n$, extracted from the i^{th} subspace of n training images, where $n < k$. Let $\mathbf{C_i}$ be the covariance matrix of $\mathbf{Y_i}$. By applying PCA on $\mathbf{Y_i}$, we obtain $\bar{I}_{\Omega_i}$, the mean of n data vectors, Λ_i, the vector of $n-1$ non-zero eigenvalues of $\mathbf{C_i}$ sorted in descending order, and $\mathbf{Q_i} \in \Re^{(n-1) \times k}$, the matrix that keeps the $n-1$ principal components of $\mathbf{Y_i}$, i.e. the first $n-1$ eigenvectors of $\mathbf{C_i}$.

Estimability of a Subspace

We consider that a subspace is estimable if a significant fraction γ_1 of the overall variance of the data can be explained by a small fraction γ_2 of eigenvectors. The thresholds γ_1 and γ_2 are determined empirically. Normalized eigenvectors

$$\hat{\lambda}_{ik} = \lambda_{ik} / \sum_{j=1}^{n-1} \lambda_{jk} \tag{4}$$

represents the fraction of variance contributed by each eigenvector. We calculate

$$n_\lambda = \underset{x \in \{1,\ldots,n-1\}}{\arg\min} \sum_{j=1}^{x} \hat{\lambda}_{ik} \geq \gamma_1 \tag{5}$$

If $n_\lambda < \gamma_2$ the block is considered estimable.

SM-constrained Reconstruction

When projected to the space spanned by $\mathbf{Q_i}$, the data vector I_{Ω_i} can be represented by its projection vector (or feature vector) V_i:

$$V_i = \mathbf{Q_i}^T (I_{\Omega_i} - \bar{I}_{\Omega_i}) \; . \tag{6}$$

As the application of PCA diagonalizes the covariance matrix $\mathbf{C_i}$, the pdf of I_{Ω_i} can be calculated by

$$\phi(I_{\Omega_i}) = \phi(V_i) = c_i \;\; exp\{-\frac{1}{2} \sum_{k=1}^{n-1} \frac{v_{ik}^2}{\lambda_{ik}}\} \; , \tag{7}$$

where c_i is the normalization coefficient, and v_{ik} and λ_{ik} are the k^{th} elements of V_i and Λ_i respectively. In fact, the likelihood of I_{Ω_i} increases as the normalized Euclidian norm of V_i decreases. Consequently, an image patch that has a low likelihood can be constrained to have a desired likelihood t by scaling down its projection coefficients by a scalar factor a

$$V_i{}' = a\, V_i \;\; s.t. \;\; \phi(V_i') = t \; . \tag{8}$$

The SM-reconstructed image patch $I^*_{\Omega_i}$ can be obtained by projecting V_i' back to the original space

$$I^*_{\Omega_i} = \mathbf{Q_i}\, V_i' + \bar{I}_{\Omega_i} \; . \tag{9}$$

The method used for the reconstruction is based on the assumption that the normal variation lies within a low dimensional subspace (the PCA subspace). The image patch is decomposed into two components: one within the PCA subspace, and the other orthogonal to PCA space. By projecting an image patch into the PCA space, the orthogonal component is considered as a residual error and discarded.

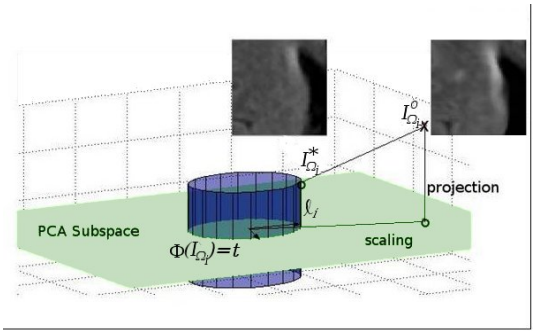

Figure 2. Illustration of the reconstruction of an image patch

However, depending on the complexity of the imaging patterns within a subspace Ω_i, the PCA subspace may not capture the normal variation. For this reason, we propose a more conservative reconstruction strategy where the orthogonal component is partially preserved, by projecting an image patch to an orthogonal distance ℓ_i from the PCA space, which is determined adaptively. If available, a spatial prior that reflects image complexity may be used for calculating ℓ_i. As this is not the case in our work, we learn ℓ_i from the data, by calculating the average residual error of the training samples through a leave-one-out (LOO) cross-validation:

$$\ell_i^j = ||I^j_{\Omega_i} - \mathbf{Q_i^j} \mathbf{Q_i^j}^T I^j_{\Omega_i}||_2 \; , \tag{10}$$

$$\ell_i = \frac{1}{n} \sum_{j=1}^{n} \ell_i^j \; . \tag{11}$$

Here $\mathbf{Q_i^j}$ is the PCA basis learned from the training set excluding the j^{th} sample, and $||.||$ denotes the L_2 norm.

In that way, the normal variation within a subspace Ω_i is modeled as the volume that lies within a hypercylinder. The reconstruction moves a given image patch to the closest point on this hypercylinder. Figure 2 illustrates the reconstruction procedure.

3. Experimental Results

We evaluated the proposed method on segmentation of white matter lesions (WMLs) and infarcts on brain MR images. WMLs are common abnormalities of the brain, which may be the result of different brain diseases, such as multiple sclerosis and cerebrovascular disease, or may appear in normal elderly subjects. MR imaging is widely used for diagnosing such diseases clinically. Manual lesion segmentation by trained experts is extremely time consuming, and suffers from high intra-observer and inter-observer variability. On FLAIR images WMLs show up as hyperintensities with respect to surrounding healthy white matter tissue. However, their intensity range also overlaps with normal

gray matter (GM), causing the failure of automatic segmentation methods based solely on image intensity. An infarct is generally the result of a stroke that occurs when the blood supply to the brain is interrupted.

On FLAIR images, infarcts have usually a hyperintense area surrounding a necrotic part that has intensity similar to the cerebrospinal fluid (CSF). Due to intensity similarities with both CSF and GM, the automatic segmentation of infarcts is a challenging problem that was seldomly addressed [11].

The training samples consist of FLAIR images without pathologies belonging to 72 healthy subjects. Image preprocessing involves skull stripping using the BET algorithm implemented in FSL software library [12], and bias correction done using N3 [6]. The images are registered to a common template using the HAMMER registration algorithm [10]. From all images, a predetermined axial slice is extracted and the experiments are performed on 2D images.

A simulated test set is created by selecting 5 images from the training set and adding simulated lesions and infarcts on these images. The remaining 68 images are used as the new training set. Simulated lesions have a circular-like shape, and a hyperintense Gaussian profile with intensities varying in a predefined interval. Simulated images are created by inserting lesions independently in 3 different positions, and with 3 different intensity ranges ([90,110], [100,120] and [110,130]). Images containing infarcts are created in a similar way. A hypointense area with intensity value 0 is inserted into the center of simulated lesions. Figure 3 shows examples of simulated lesions and infarcts. Binary ground truth masks are generated in order to be used in the assessment of the method.

Three variations of the proposed method, with increasing complexity, are used for generating the SM-reconstructed image I^* from a given test image I^0. In the baseline method (*SMR1*) a non-estimable subspace is replaced by a randomly selected subspace, and the orthogonal distance to PCA space is discarded, by setting $\ell_i = 0$. In the second method (*SMR2*) the selection strategy presented in section 2.3 is used in case of a non-estimable subspace. The third method (*SMR3*) is based on *SMR2*, but the orthogonal distance to PCA space is preserved adaptively by calculating ℓ_i for each subspace i. The performance of the proposed method is compared with a voxelwise t-test method, a standard univariate statistical analysis technique.

An abnormality map is obtained by calculating the voxelwise difference $|I^0 - I^*|$ between the test image and the SM-reconstructed image. For the t-test method, the abnormality map consists of the t-score of each voxel. Receiver operating characteristic (ROC) analysis is applied using the abnormality map and the ground truth mask. The area under the curve (AUC) score is calculated as a qualitative measure of the segmentation performance.

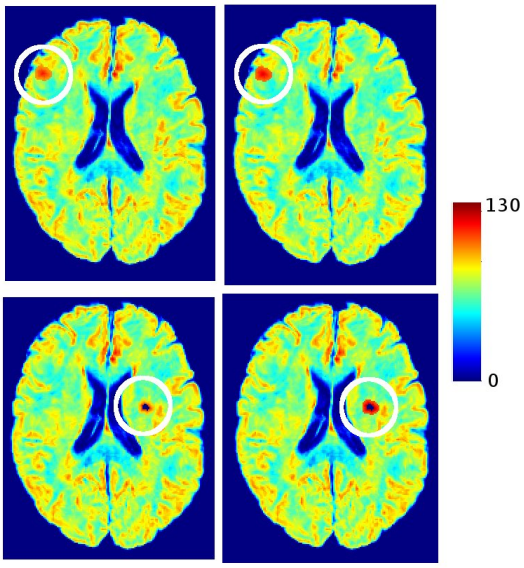

Figure 3. Top raw: simulated lesions with intensity range [100,120] and [110,130] from left to right, bottom: simulated infarcts with intensity ranges [100,120] and [110,130] on the bright area , from left to right.

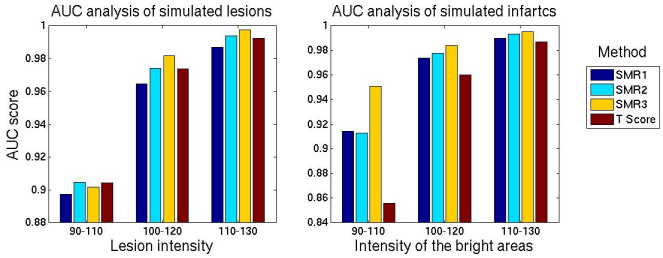

Figure 4. Average AUC scores for the segmentation of simulated lesions (left) and simulated infarcts (right).

Figure 4 presents the average AUC scores obtained by each method in segmentation of lesions and infarcts, grouped by the intensity profiles of the simulated pathologies. *SMR3* method performed better than the other methods in all test cases, with the exception of lesions with low intensity range. A qualitative analysis shows that the reconstruction using the *SMR3*method succeeds in preserving the original anatomy of the brain in healthy regions (Figure 5).

A test set with 33 images containing WMLs, and corresponding manual lesion masks extracted by an expert radiologist, is also provided. We applied the *SMR3* method on 15 of these images that have the highest lesion load on the selected slice. We obtained an average AUC score of 0.9816. Figure 6 shows an example reconstruction together with the original test image and the corresponding manual mask.

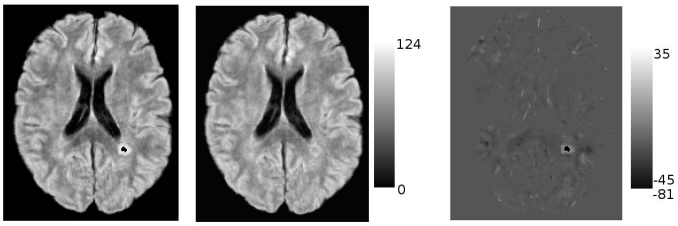

Figure 5. Sample results using method *SMR3*. Left to right: Test image with simulated periventricular infarct, with intensity range [100,120] on the bright area . SM-reconstructed image. Voxelwise difference between the initial and reconstructed images $(I^0 - I^*)$.

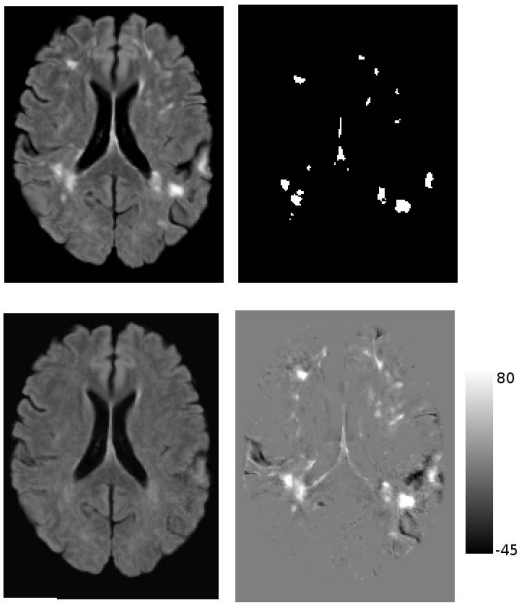

Figure 6. Sample results using *SMR3* method. Top row, left to right: Test image containing WMLs. The manual lesion mask. Bottom row, left to right: SM-reconstructed image. Voxelwise difference between the initial and reconstructed images $(I^0 - I^*)$.

4. Conclusion

We aim to learn the normal variation of the healthy anatomy, in order to detect abnormalities on a new image as significant deviations from the normal variation. Learning multivariate statistics of high dimensional images from a limited number of training samples is a challenging task. In this paper, we address the dimensionality problem by capturing the statistics of a large number of subspaces, which are more reliably estimable. We present an iterative strategy for selecting subspaces in a multi-scale fashion. We also present a model-constrained reconstruction method applied within each subspace, for obtaining images consistent with the learned subspace models. The final method is evaluated on segmentation of lesions and infarcts on brain MR images, and obtained promising results. The proposed framework is generic, and may also be applied to abnormality detection problems in other domains.

In the future, non-linear dimensionality reduction methods like Kernel PCA [9] or Isomap [13] will be investigated, in order to capture non-linearities in data while learning the subspace models. Another perspective is to use other basis like wavelets or discrete cosine transform for extracting subspaces.

5. References

References

[1] P. Anbeek, K. Vincken, M. van Osch, R. Bisschops, and J. van der Grond. Probabilistic Segmentation of White Matter Lesions in MR Imaging. *NeuroImage*, 21(3):1037–1044, March 2004. 1

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995. 2

[3] G. C. Feng, P. C. Yuen, and D. Q. Dai. Human face recognition using PCA on wavelet subband. *Journal of Electronic Imaging*, 9(2):226–233, 2000. 2

[4] D. T. Gering. *Recognizing deviations from normalcy for brain tumor segmentation*. PhD thesis, Massachusetts Institute of Technology, 2003. Supervisor-Grimson, W. Eric. 1

[5] R. Gottumukkal and V. K. Asari. An improved face recognition technique based on modular PCA approach. *Pattern Recogn. Lett.*, 25(4):429–436, 2004. 2

[6] A. E. J. Sled, A. Zijdenbos. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*, 17:87–97, 1998. 5

[7] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3-D shape representation and segmentation using spherical wavelets. *MedImg*, 26(4):598–618, April 2007. 2

[8] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig. A brain tumor segmentation framework based on outlier detection. *NeuroImage*, 21(3):1037–1044, March 2004. 1

[9] B. Schlkopf, A. J. Smola, and K. R. Müller. Kernel principal component analysis. *Advances in kernel methods: support vector learning*, pages 327–352, 1999. 6

[10] D. Shen and C. Davatzikos. HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. *IEEE Trans. Med. Imaging*, 21:1421–1439, 2002. 5

[11] S. Shen, A. Szameitat, and A. Sterr. Detection of infarct lesions from single MRI modality using inconsistency between voxel intensity and spatial location–a 3-D automatic approach. *IEEE Trans Inf Technol Biomed.*, 12(4), 2008. 5

[12] S. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17:143–155, 2002. 5

[13] J. B. Tenenbaum, V. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000. 6

[14] K. Van Leemput, F. Maes, D. Vandermeulen, C. A., and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677 – 688, 2001. 1

[15] Z. Xue, D. Shen, B. Karaçalı, J. Stern, D. Rottenberg, and C. Davatzikos. Simulating deformations of MR brain images for validation of atlas-based segmentation and registration algorithms. *NeuroImage*, 33:855–866, 2006. 2

[16] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:131–137, 2004. 2

[17] E. I. Zacharaki, S. Kanterakis, R. N. Bryan, and C. Davatzikos. Measuring brain lesion progression with a supervised tissue classification system. In *MICCAI '08*, pages 620–627, Berlin, Heidelberg, 2008. Springer-Verlag. 1

[18] Y. Zheng, S. Englander, S. Baloch, E. I. Zacharaki, Y. Fan, M. D. Schnall, and D. Shen. STEP: Spatio-Temporal Enhancement Pattern, for MR-based breast tumor diagnosis. *Medical Physics*, 36(7):3192–3204, 2009. 1