"1st International Symposium on Computing in Informatics and Mathematics (ISCIM 2011)"
in Collabaration between EPOKA University and "Aleksandër Moisiu" University of Durrës
on June 2-4 2011, Tirana-Durres, ALBANIA.

PCA Based Bayesian Approach for Automatic Multiple Sclerosis Lesion Detection

Evgin GÖÇERİ

Department of Computer Engineering, Pamukkale University, Denizli-TURKEY
Email: egoceri@pau.edu.tr

ABSTRACT

The classical Bayes rule plays very important role in the field of lesion identification. However, the Bayesian approach is very difficult in high dimensional spaces for lesion detection. An alternative approach is Principle Component Analysis (PCA) for automatic multiple sclerosis lesion detection problems in high dimensional spaces. In this study, PCA based Bayesian approach is explained for automatic multiple sclerosis lesion detection using Markov Random Fields (MRF) and Singular Value Decomposition (SVD). It is shown that PCA approach provides better understanding of data. Although Bayesian approach gives effective results, it is not easy to use in high dimensional spaces. Therefore, PCA based Bayesian detection will give much more accurate results for automatic multiple sclerosis (MS) lesion detection.

INTRODUCTION

MS is a progressive disease with lesions evolving over time. Lesions appear in the Central Nervous System (CNS) which consist of brain, spinal cord and optic nerve. Usually lesions are due to a demyelinization with a replacement of cerebra-spinal fluid instead of myelin. It is believed that a lot of myelin is the result of mistaken attack of immune cells. Immune cells protect our body against farm substances such as bacteria, viruses. But in MS, the immune cells attack the myelin so inflammation and tissue damage occurs. This causes unpredictable symptoms that can change person to person and time to time in the same person.

Manual identification of lesions by experts is extremely time consuming and subjective. This leads to high outcome variability. Therefore, automatic lesion detection is needed. Tissue and lesion intensities vary depending on spatial location in the brain, rendering both manual and automatic classification difficult. Also, overlapping tissue intensities causes intensity based automatic detection problems. Because of this reasons effective lesion detection approaches must be developed. Figure 1 shows an example brain MR image with lesions.

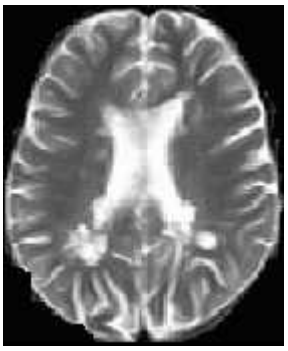


Figure 1. Brain MR image with lesions

BAYESIAN APPROACH

The Bayesian approach provides the means to incorporate prior knowledge in data analysis. Bayes's rule (1) states that the posterior probability is proportional to the product of the likelihood and the prior probability.

$$p(B \mid A) = \frac{p(A \mid B)\, p(B)}{p(A)} \tag{1}$$

Tissue classes of brain MRI voxels are background, grey matter, white matter, cerebro-spinal fluid and lesions for multiple sclerosis disease. Bayesian approaches [1-3] assign a probability to each tissue class for each voxel.

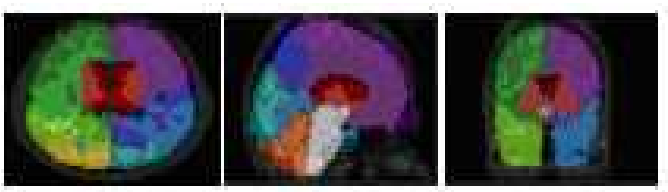
The posterior probability density is the probability of class according to given intensity value for each voxel and location information so term is $p(C_v \mid I_v, x_v, y_v, z_v)$. The prior probability density is the probability of class by given location information so the term is $p(C_v \mid x_v, y_v, z_v)$. The likelihood is the probability of intensity value for each voxel according to the given class and location information so the term is $p(I_v \mid C_v, x_v, y_v, z_v)$. Therefore, we obtain the following equation (2) from the Bayes rule,

$$p(C_v \mid I_v, x_v, y_v, z_v) = \frac{p(I_v \mid C_v, x_v, y_v, z_v)\, p(C_v \mid x_v, y_v, z_v)}{p(I_v \mid x_v, y_v, z_v)} \tag{2}$$

The normalization factor is the same for all classes given the same voxel so can be ignored for the comparision. Intensity information *Iv* for each voxel *v* located at position *(x,y,z)* is represented by a 3D vector consisting of intensities from T1-weighted (T1w), T2-weighted (T2w) and Proton Density (PD) MR images.

Tissue class intensities are found to vary significantly depending on the location in the brain. In particular, tissue and lesion intensity distributions are different in the posterior fossa as compared to the rest of the brain. For this reason,

Rola Harmouch et al. [4] explains a different likelihood distribution, which models the intensity spread for each tissue class, for each of the following regions (Figure 2): Brain center, frontal lobe, parietal lobe, occipital lobe, temporal lobe, and posterior fossa.



(a) Sagittal slice (b) Axial slice (c) Coronal slice

Figure 2. Segmented regions of brain [4]

In [4], the region based likelihood is identified as $p(I_v \mid C_v, x_v, y_v, z_v) = p(I_v \mid C_v, R_v)$. The term $p(I_v \mid C_v, R_v)$ is a 3-dimensional Gaussian distribution, and *R* is the region to which the voxel at location (*x*,*y*,*z*) belongs.

$$p(I_v \mid C_v, R_v) = \frac{e^{-\frac{1}{2}(I_v - M_{C,R})^T \operatorname{cov}_{C,R}^{-1}(I_v - M_{C,R})}}{2® \sqrt[3]{\frac{2}{|\operatorname{cov}_{C,R}|}}} \quad (3)$$

The term $M_{C,R}$ is a vector of Gaussian means $\left[«c,T1w,R \quad «c,T2w,R \quad «c,PD,R \right]^T$ consisting of intensity values for T1w-MRI, T2w-MRI and PD-MRI which are specific to tissue class information *C* and region information *R*, $\operatorname{cov}_{C,R}$ is the covariance matrix.

Prior probabilities are obtained at every voxel using statistical frequencies of tissue classes from pre-labeled training data. The prior probabilities of both lesions are adjusted at each voxel location by adding a fraction of white matter and grey matter prior probability values at that location.

Physically close voxels are more likely to come from the same tissue class. Therefore, spatial neighborhoods can provide useful information in the classification. Neighboring spatial information can be modeled using Markov Random Field (MRF) techniques.

MRF is a statistical model that provides a model to impose spatial constraints on the processed images and it proves to be a robust and accurate model [5]. The general modeling steps of spatial constraints in an MRF model are as follows: A neighborhood structure $N_{i,j}$, which contains neighboring pixels of site (*i, j*) ((*i, j*) is not included), is first defined. Then a clique is defined on the neighborhood

structure $N_{i,j}$. A set of pixel sites *c* in $N_{i,j}$ is a clique if all pairs of sites in *c* are neighbors.

The posterior probability with neighborhood is,

$$p(C_v \mid I_v, x_v, y_v, z_v, C_{neighbors}) \sim p(I_v \mid C_v, x_v, y_v, z_v, C_{neighbors})\, p(C_v \mid x_v, y_v, z_v, C_{neighbors}) \quad (4)$$

A function *Vc* called potential function defines the interactions of pixel sites in clique *c*. Spatial constraints can be imposed on the processed image through the formulation of function *Vc*. The potential function is related to the energy function as $U(w) = \sum_{c \in C} V_c(w)$. It is clear that lower energy clique potentials imply a more stable or more likely configuration that means the labels of the neighbors are likely to co-exist. To find the lowest energy configuration Iterated Conditional Model (ICM) algorithm is used.

PRINCIPLE COMPONENT ANALYSIS (PCA)

PCA, which is mostly used in many areas of applied statistics, is a powerful tool for analyzing data. The main principle of this approach is to represent multidimensional data by using fewer number of variables. PCA performs a rotation of the data that maximizes the variance in the new axes. The main reason to choose PCA technique is that visualization and computation in a fewer dimensional space is always easier and gives higher performance than in many dimensional space.

The properties of the PCA technique can be summarized as follows: 1) It maximizes the variance of the extracted features; 2) The extracted features are uncorrelated; 3) It finds the best linear approximation in the mean-squares sense; 4) It maximizes the information contained in the extracted features.

The first principle component that is the eigenvectors, which corresponds to the largest eigen values of the covariance matrix. In other words, the directions with the most variation. The second principle component is the direction in the data with the most second variation. After founding all principle components, the observed data is projected onto these components. The number of components selection is generally performed by trial and error. By selecting only the first *d* rows of Y, we have projected the data from *n* down to *d* dimensions. For large matrices, Singular Value Decomposition (SVD) is used to find eigenvectors. In [6], more information about SVD and also the relationship between SVD and PCA can be read, which is one of the nice papers about this topics. In the next section, the PCA model is described.

PCA APPROACH FOR MS LESION DETECTION [7]

Kroon et al. identify that PCA based MS lesion detection in MR images can be thought as two group problem. One of them is the pixels displays MS properties

and the other is the group of pixels that does not display any properties of MS. Every pixel can be identified with a feature vector, which contains features such as gray levels of neighbor pixels. A training set obtained with known MS pixels and non-MS pixels can be used for construction of the two groups. A kind of distance measure between the feature vector of a certain pixel and the class feature means and variances can be used to categorize a pixel as MS or non-MS.

This PCA based method consists of the following steps: First step is construction of a L_x matrix by using the feature vectors *x* of MS pixels from training dataset, and also construction of a matrix G_x by using features vectors x' of non-MS pixels in the training dataset.

$$L_x = [x_1, x_2, ..., x_n] \qquad G_x = [x_1', x_2', ..., x_n'] \tag{5}$$

Generally, MR neighborhood intensities are used as a feature vector since because they describe the texture of the lesion regions. In addition to the intensity values, the authors add other information to their feature vector, such as histogram information and MS probability atlas to obtain the lesion probability values for each voxel.

Second step is the subtraction of the mean « value to center the feature vector matrices.

$$«_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{6}$$

$$L_x = [x_1 - «_x, x_2 - «_x, ..., x_n - «_x], G_x = [x_1 - «_x, x_2 - «_x, ..., x_n - «_x] \tag{7}$$

Third step is to calculate the eigenvectors U_G and eigen values a in the mean centered non-MS feature data G_m

$$G_m = U_G \Sigma_G V_G^T \sqrt{n} \tag{8}$$

$$\mathsf{a}_i = \Sigma^2_{(i,i)}, \quad \mathsf{a}_i = [\mathsf{a}_1, \mathsf{a}_2, ... \mathsf{a}_n] \tag{9}$$

The next step is to kept the biggest eigen values (99%), which describe the big variance in the dataset U_G', and to eliminate the smallest eigen values (1%).

The last step is projection, which is performed by multiplying the eigen vectors with data sets and the orthogonal non-MS data is obtained

$$L_{pca1} = U_{G^T}' L_m, \quad G_{pca1} = U_{G^T}' G_m \tag{10}$$

Similar steps are repeated for the MS feature data, the 99% largest eigen values which describe the main variance in the data are kept U_L' and the 1% smallest are discarded.

$$L_{pca1} = U_L' \Sigma_L V_L^T \sqrt{n} \tag{11}$$

Then the normalization of the variance in the eigenvectors is performed, it is normalized to one. After the multiplication of eigenvectors with the data sets, the MS data is obtained as orthogonal.

$$U_{LN} = \frac{U_L'}{diag(\Sigma_L)} \tag{12}$$

$$L_{pca2} = U_{LN^T} L_{pca1}, \quad G_{pca2} = U_{LN^T} G_{pca1} \tag{13}$$

The eigenvectors and values in the non-MS feature data are calculated again using the same steps. Both the datasets can have orthogonal features again by multiplying eigenvectors with datasets.

$$G_{pca2} = U_{G2} \Sigma_{G2} V_{G2^T} \sqrt{n} \tag{14}$$

$$L_{pca3} = U_{G2^T} L_{pca2}, \quad G_{pca3} = U_{G2^T} G_{pca2} \tag{15}$$

One matrix rotation is obtained using the previous PCA rotations as U_{tot} and the covariance matrix $C_{G_{pca3}}$ is calculated by using the G_{pca3} data.

$$U_{tot} = U_G U_{LN} U_{G2} \tag{16}$$

Mahalanobis distance can be used as the distance measure from the feature vectors of a test pixel to the MS and non-MS pixels. Classification of pixels is performed by applying a threshold the log-likelihood ratio of the distances $\tilde{\wedge}(u,v)$

$$u = U_{tot}{}^T (x - \ll_G), \quad v = U_{tot}{}^T (x - \ll_L) \tag{16}$$

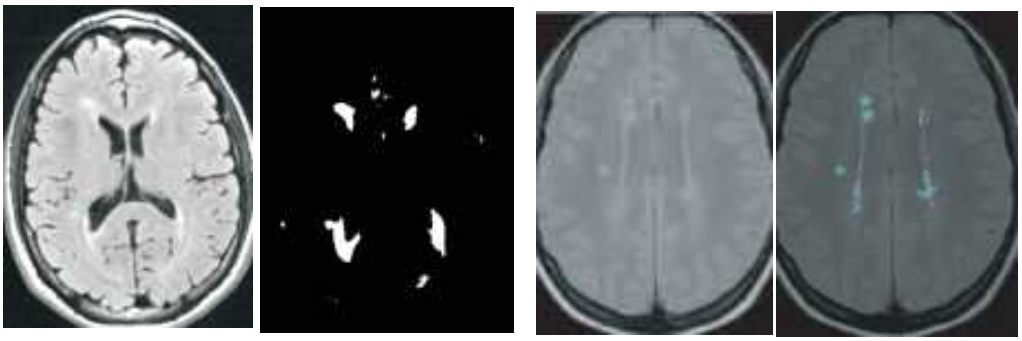
$$\wedge(x,y,z) = \frac{1}{2}(u^T C_{G_{pca3}}{}^{-1} u - \|v\|^2) + \ln(\left|C_{G_{pca3}}{}^{-1}\right|) \tag{17}$$

$$MS(x,y,z) = \tilde{\wedge}(x,y,z) \geq t \tag{18}$$

CONCLUSIONS AND RESULTS

Application results for these two approaches are given in Figure 3. The first example result is shown in Figure 3.(a) [7] uses a FLAIR image as input and PCA

based segmentation. The second example result is shown in Figure 3.(b) [4] uses a PD-MRI as input and Bayesian approach.



(a) (b)

Figure 3. MS Lesion Detection with PCA and Bayesian Approaches

PCA classifier provides better understanding of data. However, PCA projects data onto a set of orthogonal vectors, this restricts the new input components to be the linear combination of old ones. Although Bayesian approach gives effective results, it is not easy to use in high dimensional spaces. Therefore, PCA based Bayesian detection will give much more accurate results for automatic MS lesion detection.

As a future work, an implementation on PCA based Bayesian lesion detection will be performed.

REFERENCES

[1] Dugas-Phocion, G., Gonzalez, M. A., Lebrun, C., Chanalet, S., Bensa, C., Malandain, G., Ayache, N. (2004) Hierarchical segmentation of multiple sclerosis in multi-sequence MRI, *Proc. of the IEEE Intl. Symposium on Biomedical Imaging (ISBI)*, 157-160

[2] Kamber, M., Collins, D. L., Narayanan, S., Pioro, E. , Evans, A. C. and Arnold, D. L. (1993) Tissue probability models in standardized brain space and the classification of MS plaques in MRI. *Proc. Society of Magnetic Resonance in Medicine*, **3**(1),1600

[3] Marroquin, J. L., Vemuri, B. C., Botello, S., Calderon, E., Fernandez, B. A. (2002) An accurate and efficient Bayesian method for automatic segmentation of brain MRI. *IEEE Trans Medical Imaging*, **21**(8),934-45.

[4] Harmouche, R., Arnold, D., Francis, S., Collins, D. L., Arbel, T. (2006) Bayesian MS lesion classification modeling regional and local spatial information. *Pattern Recognition 18th International Conference (ICPR)*, Ottawa, Canada.

[5] Li, S. Z. (1995) On discontinuity-adaptive smoothness priors in computer vision. *IEEE Trans. on Pattern and Machine Intelligence,* **17**(6),576–586

[6] Wall, M. E., Rechtsteiner, A., Rocha, L. M. (2003) Singular value decomposition and principal component analysis. *A Practical Approach to Microarray Data Analysis. D. P. Berrar, W. Dubitzky, M.Granzow, eds., Norwell, MA*, 91-109.

[7] Kroon, D. J. and Van Oort, E. S. B. and Slump, C. H. (2008) Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis. *MIDAS Journal (MICCAI2008 Conf. USA),*604-617