Automatic selection of ROIs in functional imaging using Gaussian mixture models

J.M. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C.G. Puntonet, E.W. Lang

Address: Departamento Teoría de la Señal, Telemática y Comunicaciones, Universidad Granada, Spain

Received 22 April 2009, Received in revised form 12 May 2009, Accepted 13 May 2009

Abstract

We present an automatic method for selecting regions of interest (ROIs) of the information contained in three-dimensional functional brain images using Gaussian mixture models (GMMs), where each Gaussian incorporates a contiguous brain region with similar activation. The novelty of the approach is based on approximating the grey-level distribution of a brain image by a sum of Gaussian functions, whose parameters are determined by a maximum likelihood criterion via the expectation maximization (EM) algorithm. Each Gaussian or cluster is represented by a multivariate Gaussian function with a center coordinate and a certain shape. This approach leads to a drastic compression of the information contained in the brain image and serves as a starting point for a variety of possible feature extraction methods for the diagnosis of brain diseases.

Nowadays medical diagnosis benefits from a great variety of available imaging techniques, like for instance single photon emission computed tomography (SPECT), positron emission tomography (PET) or functional magnetic resonance imaging (fMRI). Those methods allow to study functional properties of the human brain and are therefore particularly devised for the diagnosis of neurodegenerative diseases like Alzheimer’s disease (AD) [5]. After the reconstruction of the raw data one obtains a three-dimensional image displaying the local cerebral blood flow, glucose metabolism, etc. encoded in the grey-level intensities I(x) of the brain image. The “fingerprint” of different brain diseases in such images often is a reduction of the intensity in certain regions of the brain image [9], which may be difficult to visually identify in the early stage of the disease [13]. Therefore computer-aided diagnosis tools are widely discussed which may improve the analysis of brain images and open the possibility to detect dementia already in an early stage. Such approaches are, on the one hand, based on studying regions of interest [2], which are defined according to the anatomical regions of the brain. On the other hand, the voxel intensities of the whole brain image are used applying statistical learning methods for the classification [8]. The advantage of the latter approach is that no specific information about the disease is necessary. Due to the large number of voxels, however, this method suffers from the so-called small sample size problem, which occurs if the number of training samples is much smaller than the dimensionality of the feature space [4].

In functional imaging studies, model-based clustering has been employed in fMRI analysis for grouping relevant coordinates in the Talairach space [12]. For this task, Activation Likelihood Estimation (ALE) is firstly employed for reducing the list of activation maxima which have one or more other maxima in their vicinity and then, these coordinates x_i with their membership z_i to each cluster, are subjected to clustering based on finite mixture of probability distributions [6]. In this work we present a different clustering approach based on a Gaussian mixture model (GMM), which allows to automatically represent a functional brain image by a set of Gaussians, each of them denoted by the word “cluster”, approximating the grey-level distribution of the original image. The obtained ROIs may serve as a basis for various possible classification methods, as discussed in the experimental section. The advantage of that approach is that feature vectors with a significantly reduced dimensionality can be extracted, avoiding the small sample size problem. The working principle of the clustering algorithm is demonstrated by applying it to real SPECT images, although other tomographic image modalities could be used as well.

GMMs are among the most statistically mature methods for classical clustering, though they are also used intensively for density estimation see e.g. [7]. In general statistics one is often interested in reconstructing the underlying probability distribution p_0(x) from a given set of samples x_1, ..., x_n drawn from this distribution p_0(x). In this context, Gaussian mixture models are widely used [7], where one fits a sum of multivariate Gaussian distributions to an unknown
probability distribution function (pdf) $p_0(x)$. The unknown pdf is therefore approximated by

$$p_0(x) \approx p(x) = \sum_{n=1}^{k} w_n f_n(x), \quad (1)$$

where the Gauss functions in $d$ dimensions $f_n(x)$ are defined by

$$f_n(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp \left[ -\frac{1}{2}(x - \mu_n)^T \Sigma_n^{-1} (x - \mu_n) \right], \quad (2)$$

and the weight factors $w_n$ are normalized according to $\sum_{n=1}^{k} w_n = 1$. The parameters $\theta = (w_n, \mu_n, \Sigma_n)$ of the Gaussian mixture are determined by maximizing the likelihood $L(\theta|x) = \prod_{i=1}^{n} p(x_i)$, see e.g. [7]. The likelihood corresponds to the probability to observe the given samples $x_1, \ldots, x_N$ subject to the distribution $p(x)$, if independent and identically distributed random variables are assumed.

For the case where the observed data is given in form of a histogram, the maximum likelihood estimation can also be used in a modified way [11]. Let $h_j$ be the height of each histogram bin which corresponds to the number of observations of the variable $x_j$, the likelihood can be expressed as

$$L(\theta|x) = \prod_{j=1}^{B} [p(x_j)]^{h_j}, \quad (3)$$

where $B$ is the number of histogram bars. The total number of observations in that case is the sum of the histogram heights, $N = \sum_{j=1}^{B} h_j$.

Maximizing the likelihood under the constraint $\sum_{n=1}^{k} w_n = 1$ is equivalent to finding the maximum of the quantity

$$J = \sum_{j=1}^{B} h_j \ln p(x_j) - \lambda \left( \sum_{n=1}^{k} w_n - 1 \right), \quad (4)$$

where the first term is the log-likelihood and $\lambda$ is a Lagrange multiplier ensuring the correct normalization. To maximize the likelihood we employ the widely used expectation maximization (EM) algorithm [3], which follows from equation the gradient of $J$ in Eq. (4) with zero. It is a recursive scheme which allows to iteratively compute the parameters $w_n$, $\mu_n$, and $\Sigma_n$ of the Gaussian mixture. Along the same lines as shown in [7] we obtain the following equations to update the desired parameters:

$$w_n = \frac{1}{N} \sum_{j=1}^{B} h_j q_n(x_j), \quad (5)$$

$$\mu_n = \frac{1}{w_n N} \sum_{j=1}^{B} h_j q_n(x_j) x_j, \quad (6)$$

$$\Sigma_n = \frac{1}{w_n N} \sum_{j=1}^{B} h_j q_n(x_j) (x_j - \mu_n) (x_j - \mu_n)^T, \quad (7)$$

where the posterior probability $q_n(x)$ entering the above equations is defined by $q_n(x) = w_n f_n(x)/p(x)$.

Starting with an initial guess for $w_n$, $\mu_n$ and $\Sigma_n$ we can recursively apply Eqs. (5)–(7) until convergence is reached, i.e. the changes in the log-likelihood are smaller than a given threshold.

We now apply the GMMs to brain images, which are three-dimensional intensity distributions discretised to $V$ volume elements or voxels. The voxel positions form a cubic lattice labelled with $x_i$, and the corresponding intensities are $I(x_i)$. In order to adjust the intensity profile of the brain image we associate the histogram heights $h_i$ with the voxel intensities $I(x_i)$, and the number of samples $N$ with the total intensity $I_{tot} = \sum_{j=1}^{V} I(x_j)$ of all $V$ voxels, and then we apply the method of the previous section in a straightforward way. It is worth mentioning that in this case the GMM are used to solve a density estimation problem and then, all voxels (positions) contributes to all Gaussians thus, “membership” depends on the distance of the cluster center to the voxel position $x$. Other clustering methods in functional imaging use an explicit membership, thus the input pattern as a pair: position and membership [12] is required. The key question of them is that the “gray-level” information is not subjected to clustering, but only the list of activation maxima previously defined by the ALE principle.

This leads to a Gaussian mixture $p(x)$ which is normalized to 1. By multiplying with the total intensity we obtain an intensity distribution

$$I_{Gauss}(x) = I_{tot} \sum_{n=1}^{k} w_n f_n(x), \quad (8)$$

which approximates the original brain image $I(x)$ “as good as possible” in terms of the likelihood defined in Eq. (3). This novel approach for obtaining the GMM considers all the voxels (positions) with weight factors proportional to their grey-levels, and each Gaussian (cluster) represents the ROI in the brain image which groups all the voxels inside of it. The clustering algorithm is then used to define the ROIs, where we compare the brain activations in order to classify the image. The result of the proposed algorithm applied to the average SPECT image of the control subjects, defines the ROIs which are used later for the classification task. The resulting clusters are used to extract the cluster activations $I_k$ for each SPECT image, which are obtained by averaging over the intensities within cluster $n$. The $k$-dimensional feature vector for each SPECT image is then defined by $v = (v_1, \ldots, v_k)$ where $k$ is the number of clusters, i.e. we choose $k = 64$. A further analysis of the feature vectors provides an improvement of the performance in the classification rate of the CAD system, i.e. using statistical classifiers such as SVM [1].

The classification is done in two steps: the training and the classification. In the training stage the classifier is defined according to the feature vectors of a given training set, where the class labels are known. Once defined, the classifier can be used to categorize unknown samples. The classifier in general is a function in the feature space assigning positive or negative values to the members of the different classes. For this study we use the SVM methodology which has been successfully applied to many other areas [18]. SVMs have focused recent attention from the pattern recognition community due to a number of theoretical and computational merits derived from the Statistical Learning Theory (SLT) developed by Vladimir Vapnik at AT&T. In particular, we employ SVMs with different types of kernels [18,16]. The basic idea of that approach is to transform the data points, which need to be classified, into a distorted higher dimensional feature space $F$, where a linear hyperplane classifier can be applied. The SVM classifier in general can be written as

$$g_{SVM}(v) = \sum_{i=1}^{N} \alpha_i y_i \Phi(s_i) \cdot \Phi(v) + b \quad (9)$$

with Lagrange multipliers $\alpha_i$, support vectors $s_i$, class labels $y_i$ ($y_i = \pm 1$) and a constant $b$. Here, $\Phi$ denotes the transformation of the feature vectors into the effective feature space $F$ and we see that the classifier is linear in the transformed feature vectors. The parameters of the above equation are the solution of a quadratic optimization problem, which are determined, i.e. by the well-known Sequential Minimal Optimization (SMO) algorithm [14]. The dot product of the transformed feature vectors can be expressed by a suitable kernel function

$$\Phi(s_i) \cdot \Phi(v) = K(s_i, v), \quad (10)$$
so that the transformation $\Phi$ does not enter the computation explicitly. The kernel functions we use for the classification are a linear kernel $K(s, v) = s \cdot v$, polynomial kernels $K(s, v) = (1 + s \cdot v)^p$ of 2nd and 3rd order, and a Gaussian radial basis function (RBF) kernel $K(s, v) = \exp(-|s - v|^2/(2\sigma^2))$ with $\sigma = 9$.

The working principle of the proposed clustering method is demonstrated on the average SPECT image of 23 normal patients (NOR), provided by the “Virgen de las Nieves” hospital (Granada, Spain). The image is symmetrised with respect to the central sagittal plane and the intensities are squared in order to sharpen the structures, which facilitates the definition of the clusters. Fig. 1 displays axial slices of the original image, the clusters and the reconstructed image from the Gaussian mixture according to Eq. (8). We see that the location and shapes of the clusters represent very well the regions of high intensity of the original image; Moreover, the cluster configuration perfectly respects the axial symmetry of the brain image, and the reconstructed image reasonably reproduces the course structure of the SPECT image. The three-dimensional cluster arrangement is shown in the inset of Fig. 2.

To demonstrate the general applicability and stability of the proposed clustering method, we applied the algorithm to 51 real SPECT images of both NOR and Alzheimer’s (AD) patients. In all cases the clustering scheme converged well as shown in Fig. 2. The log-likelihood is normalized to the interval $[-1, 0]$ and for all subjects the algorithm converges similarly and monotonically. We also show the result of the application to the proposed approach to individual subjects in Fig. 3 (no symmetry operation is applied in this case). As clearly shown in this example, normal perfusion patterns provide symmetric clustering configuration (color and shape) with the presence of activation maxima located in the parieto-temporal, posterior cingulate, and medial temporal cortices, unlike AD patients whose cluster configuration shows asymmetries and hypo-perfusion patterns in the previous mentioned areas.

Each cluster is represented by 10 independent parameters: the centre coordinate (3 parameters), the covariance matrix (6 parameters) and the weight factor. Considering the fact that the original image has a size of $39 \times 47 \times 34 = 62322$ voxels and the reconstructed image is defined by 640 parameters, we reach a compression rate of about 100. Therefore, this clustering approach is also applicable for the compression of grey-scale images, if one is interested in the local grey-level distribution of the images.

The classification performance of our approach is also tested using SVM-based supervised learning [10,1] and the leave-one-out method, see [15], that is, the classifier is trained with all but one images of the database. The remaining image, which is not used to define the classifier, is then categorized. In particular, The non-linear RBF-SVM classifier [18] in combination with the proposed features extraction scheme actually yields the highest accuracy rate of 94%, corresponding to 3 (AD) out of 51 misclassifications while
the linear voxel as features (VAF) approach only provides a rate of 84.3% [17]. The reason for the good performance of this feature extraction scheme is that the relevant information necessary for the classification is compressed in the parameters of the clusters, and the irrelevant information is ignored as in [1]. No-linear classifiers outperform linear ones since we drastically reduce the dimensionality of the features vectors which is comparable with the number of samples, i.e. scans. With this feature extraction method one partially avoids the so-called small sample size problem and non-linear classifiers may be efficiently used.

Finally, we also test our approach on fMRI data. The database used in this part was obtained from the experiment conducted by the FIL methods group at The Wellcome Department of Cognitive Neurology[1]. The goal in this preliminary experiment is to model the acquisitions using the previous model-based clustering and to check the reliability of the proposed model in detection of functional cortical areas that are relevant for the investigated cognitive function (including artifacts). The location and the shapes of the clusters coincide very well with the regions of high intensity in the original images (see Fig. 4). Also the reconstructed image according to Eq. (1) reproduces the coarse structure of the original image. The obtained ROIs indicate the locations of the high intensity regions for a single subject and could also serve as a mask for the feature extraction scheme.

We have presented a method to subdivide a three-dimensional brain image into clusters according to the regions of high and low activation. The clusters are defined by means of a Gaussian mixture which approximates the intensity distribution of the brain image following a maximum likelihood criterion. The resulting ROIs are useful for the computer-aided diagnosis of brain diseases like Alzheimer’s disease, since the classification methods based on statistical learning improve their performance, if the feature vectors are low dimensional. This can be achieved, for instance, by using the average cluster intensities as features so that the dimensionality of the feature vectors corresponds to the number of clusters. The clusters may also be used to define regions of interest or to extract feature vectors using the parameters of the Gaussian mixtures directly. Moreover, we could apply so-called component-based SVM, where the clusters may be used as components [10].

**Acknowledgments**

This work was partly supported by the MICINN under the PETRI DENCLES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) and HD2008-0029 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Project (TIC-02566). Furthermore it was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD). We are grateful to M. Gómez Río and coworkers from the “Virgen de las Nieves” hospital in Granada (Spain) for providing and classifying the SPECT images used in this work.

**References**


---

1 see at http://www.fil.ion.ucl.ac.uk/spm/ for a description.