Pattern Recognition Letters 31 (2010) 781–790

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

White matter lesion segmentation based on feature joint occurrence probability and χ^2 random field theory from magnetic resonance (MR) images

Faguo Yang [a,*], Zuyao Y. Shan [b], Frithjof Kruggel [a]

[a] Signal and Image Processing Lab, Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, United States
[b] Division of Translational Imaging Research, Department of Radiological Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

ARTICLE INFO

Article history:
Received 4 September 2009
Received in revised form 20 January 2010
Available online 28 January 2010

Communicated by M. Couprie

Keywords:
White matter lesions
Joint probability
χ^2 Random field theory

ABSTRACT

Lesions of the brain's white matter are common findings in MR examinations of elderly subjects. A fully automatic method for segmenting white matter lesions is proposed here. The joint probability of multi-modality MR image intensities is used as a feature to segment lesions, because lesion intensities usually are outliers of the normal tissue intensities and the lesions' joint intensity probability appears much smaller than those of normal brain tissues. The χ^2 random field theory is used to determine the significance of a detected lesion and provides a strict statistical analysis to exclude small-sized false-positive lesions. Experimental results show that the automatic segmentation of lesions is in high agreement with manual segmentation, and the χ^2 random-field-based statistical analysis greatly improves lesion segmentation results.

1. Introduction

Diffuse white matter (WM) lesions are characterized (mostly) by a loss of myelin and an increase of extracellular space and revealed by magnetic resonance imaging (MRI) techniques due to their higher water content. These lesions are considered as a sign of pathological aging (Deary et al., 2003). The demyelinization of WM fibers may affect their conduction properties and lead to a decrease in cognitive performance, such as a subtle memory loss, a slower processing speed, or an early fatigue (Gunning-Dixon and Raz, 2000; de Groot et al., 2000). Obviously, the presence of WM lesions worsens the cognitive performance of patients suffering from other neurodegenerative processes such as Alzheimer's disease (Skoog et al., 1996; Leys et al., 1990; Hirono et al., 2000).

Brain lesion segmentation approaches can be classified as manual, semi-automatic (Zijdenbos et al., 1994; Udupa et al., 1997; Hojjatoleslami and Kruggel, 2001), and fully automatic (Anbeek et al., 2004; Lao et al., 2006; Kruggel et al., 2008; Dyrby et al., 2008; Herskovits et al., 2008; Admiraal-Behloul et al., 2005; van Leemput et al., 2001; Yang et al., 2004). Due to the large amount of work required for manual segmentation, and the considerable inter- and intra-rater variability of the results, semi- and fully automatic brain lesion segmentation methods are preferred. In semi-automatic methods, the user marks a seed location for a region growing algorithm that segments a lesion (Hojjatoleslami and Kruggel, 2001) or accepts/rejects fuzzy-connected candidate regions as brain lesions (Udupa et al., 1997). Semi-automatic approaches are not adequate for projects involving large databases because of the amount of manual work required. All of the above methods depend mainly on a voxel's intensity to segment lesions. Taking a different approach, Gerig et al. (2000) explored time domain features to classify lesions, but two or more MR scans of the same subject must be available.

Fully automatic algorithms for WM lesion segmentation can be grouped into two classes: supervised approaches (Anbeek et al., 2004; Lao et al., 2006; Kruggel et al., 2008; Dyrby et al., 2008; Herskovits et al., 2008), in which classifiers such as K-nearest neighbors, support vector machines, neural networks, or Bayes classifiers are used to distinguish lesions from normal brain tissues, and unsupervised methods (Admiraal-Behloul et al., 2005; van Leemput et al., 2001; Yang et al., 2004). The first group of approaches requires neuroradiologists to manually segment lesions in datasets of training subjects. Often, intensity normalization across scans is necessary which may have an adverse effect on lesion segmentation. Most approaches of these two types do not take spatial lesion information into account or apply heuristics to remove spurious small lesions (Admiraal-Behloul et al., 2005; Yang et al., 2004). Spatial autocorrelations in the data are ignored, unless the intensity of neighboring voxels is taken into account in the classification process (Lao et al., 2006; Zhang and Chen, 2004; Chen and Zhang, 2004).

In this paper, we propose an unsupervised approach for WM lesion segmentation. As proposed by van Leemput et al. (2001), we model lesions as outliers in the multivariate intensity distribution of healthy tissues. We compute the joint feature occurrence

* Corresponding author. Fax: +1 901 595 3981.
E-mail address: faguo.yang@stjude.org (F. Yang).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.01.025

probability as revealed by multi-sequence MR images. Because WM lesions are small and inhomogeneous, the joint intensity probability of lesion voxels is much smaller than that of healthy brain tissues. The resulting probability map is modeled as a χ^2 random field in a second step, and we consider lesions as "unusual events" in this random field and attach a probability to a cluster of voxels for being a lesion (Cao, 1999). Thus, larger clusters of outliers are more likely to be classified as lesions. This statistically rigorous context for WM lesion detection is the core contribution of this work.

2. Theory

Lesions are considered outliers of a multiple, multivariate intensity distribution that represent the major tissue components in the image, measured by the intensity joint occurrence probability. Clusters of outlier voxels are rated for their probability of being a true lesion using random field theory.

2.1. Intensity joint occurrence probability

Let I_v denote the intensity vector associated with voxel v in co-registered multi-sequence MR images. Suppose a subject image volume is composed of N_C classes denoted by $C_1,\ldots,C_{N_C}$. The joint probability of the occurrence of the intensity vector I_v can be expressed as:

$$P(I_v)=\sum_{k=1}^{N_C}P(v\in C_k)P(I_v|v\in C_k). \tag{1}$$

We assume that intensities of healthy compartments in multi-sequence MR images are multivariate Gaussian distributed, in which the variance is partially due to the structural variability of the compartment itself, partially due to the partial volume effect at boundaries, and partially due to additive white noise:

$$P(I_v|v\in C_k)=((2\pi)^d|\Sigma_k|)^{-1/2}\exp\left(-1/2(I_v-\mu_k)^T\Sigma_k^{-1}(I_v-\mu_k)\right), \tag{2}$$

where μ_k represents the mean, Σ_k is the covariance matrix of (tissue) class k, and d is the dimension of the intensity vector I_v, or here, the number of imaging sequences. Most commonly, model parameters μ_k,Σ_k are determined by maximum likelihood estimation (MLE) (Lehmann and Casella, 1998):

$$\mu_k=\frac{1}{n}\sum_v I_v \quad\text{and}\quad \Sigma_k=\frac{1}{n-1}\sum_v(I_v-\mu)^T(I_v-\mu),\quad v\in C_k. \tag{3}$$

The breakdown point (defined as the proportion of samples tending to infinity also makes the estimation go to infinity) of the approach is 0. Estimating model parameters of healthy tissues in the presence of lesions (with unusual or extreme intensity values) requires a robust method. Therefore, we use the minimum covariance determinant (MCD) estimator (Rousseeuw, 1985). The basic idea of this robust method is to estimate the mean and covariance from a fraction f (usually $0.5\leqslant f<1$) of the whole set of n voxels by minimizing the determinant of the covariance matrix $|\Sigma|$ with respect to the selection of $n\times f$ samples.

Because MCD estimation becomes very time-consuming if n is large, Rousseeuw and van Driessen (1999) proposed a fast variant: First, we randomly select n_s subsets $\Phi_1,\ldots,\Phi_{n_s}$ with a small sample size l (note that $n_s l<n$), and determine their model parameters using Eq. (3). Then, the following process for a set of samples Φ denoted as C-step is performed for each subset Φ_j:

1. Compute the Mahalanobis distance of each sample in the subset using current parameter estimates μ,Σ.
2. Sort the $V(\Phi)$ (the number of samples in set Φ) samples by their Mahalanobis distance in ascending order.
3. Re-estimate the model parameters using the first $V(\Phi)\times f$ samples in the list by Eq. (3).
4. Repeat steps 1–3 until the change in $|\Sigma|$ falls below a pre-set limit.

Thus, we obtain n_s estimates μ_j,Σ_j. Now, we merge all the initial subsets $\Phi_1,\ldots,\Phi_{n_s}$ into a larger one Φ^* (note that Φ^* does not include all the n samples but only $n_s l$ samples), and repeat the C-step n_s times with $\mu_j,\Sigma_j(j=1,\ldots,n_s)$ as the starting estimates of the parameters to find n_s refined model parameter estimates with this larger sample set Φ^* denoted by $\hat{\mu}_j,\widehat{\Sigma_j}(j=1,\ldots,n_s)$. Finally, μ^*,Σ^* are determined for the full set of n samples by performing the C-step n_s times with the n_s refined parameters $\hat{\mu}_j,\widehat{\Sigma_j}(j=1,\ldots,n_s)$ as the starting points and selecting the parameters with the smallest covariant matrix determinant.

We model the brain extracted from T_1-weighted MR images as composed of three classes, roughly, white matter (WM), gray matter (GM), and cerebro-spinal fluid (CSF). Images are first segmented using an algorithm based on hidden Markov random fields (Zhang et al., 2001). Then, the fast MCD method is applied to robustly estimate the mean and covariance matrix of the three classes. The prior probability $P(v\in C_k)$ for each class is assumed to be equal resulting in $P(v\in C_k)=1/N_C$. Finally, the joint occurrence probability $P(I_v)$ is computed for each voxel.

2.2. Modeling the joint probability distribution

For convenience, let us use the logarithm $u_v=-\log(P(I_v))$ of the joint probability (Eq. (1)) in the following. Because lesion voxels have an unusual joint intensity probability compared with normal brain tissues, larger values of u_v indicate a higher probability for being a lesion voxel.

Because u_v corresponds to a logarithm of a multiple, multivariate Gaussian distribution, deriving a closed-form expression for the distribution of u_v is not straightforward. Here we demonstrate that there is an upper limit of u_v, which is χ^2-distributed. Let us consider the case of a single class first. For $N_c=1,u_v$ is related to the squared Mahalanobis distance, denoted by r_v^2:

$$u_v=0.5(I_v-\mu_k)^T\Sigma_k^{-1}(I_v-\mu_k)+b_k=0.5r_v^2+b_k, \tag{4}$$

where b_k corresponds to the sum of all log-transformed class-dependent constants in Eqs. 1 and 2. The squared Mahalanobis distance r_v^2 is χ^2-distributed with degrees of freedom (DOF) d, the number of components in the intensity vector I_v.

For the multi-class case $N_c>1$, we exchange the summation in Eq. (1) and the log-transformation when computing u_v. Using Jensen's inequality and the fact that the negative logarithm $(-\log(x))$ is a convex function and $\sum_{k=1}^{N_C}P(v\in C_k)=1$, we find:

$$u_v=-\log\left(\sum_{k=1}^{N_C}P(v\in C_k)P(I_v|v\in C_k)\right) \tag{5}$$

$$\leqslant-\sum_{k=1}^{N_C}P(v\in C_k)\log(P(I_v|v\in C_k)). \tag{6}$$

Because the class-wise prior probability $P(v\in C_k)$ is often assumed to be $1/N_C$, it follows:

$$u_v\leqslant-\frac{1}{N_C}\sum_{k=1}^{N_C}\log(P(I_v|v\in C_k)) \tag{7}$$

$$=\frac{1}{2N_C}\sum_{k=1}^{N_C}(r_v^2+2b_k) \tag{8}$$

$$=\frac{1}{2N_C}\sum_{k=1}^{N_C}r_v^2+b, \tag{9}$$