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

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.

© 2010 Elsevier B.V. All rights reserved.

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).

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:

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

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:

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

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

$$= \frac{1}{2N_C} \sum_{k=1}^{N_C} r_v^2 + b, \tag{9}$$
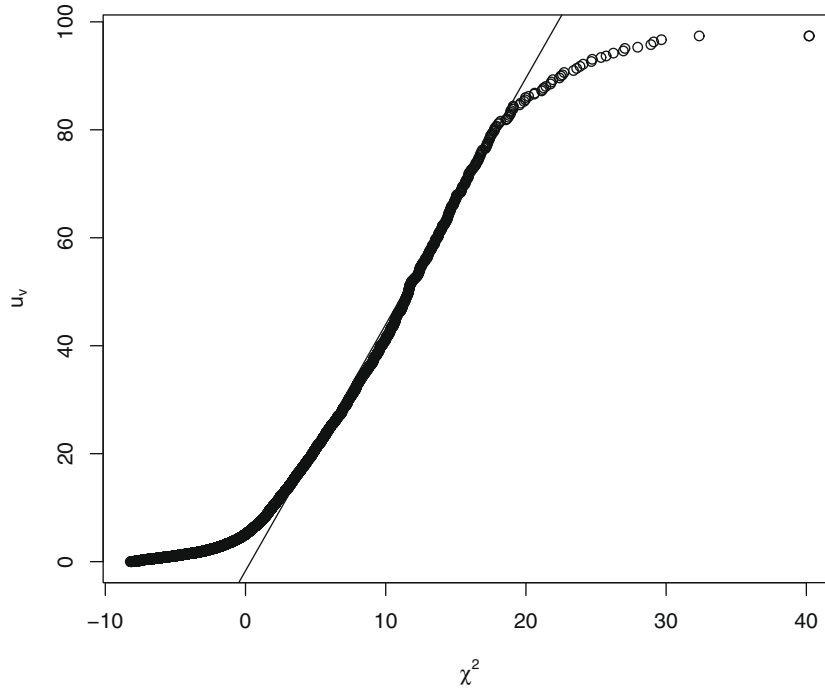
Fig. 1. A QQ-plot of experimental values of u against a scaled and shifted χ^2 distribution verifies the distributional assumption.

where $b = \sum_{k=1}^{N_c} 2b_k$ is a constant and $\sum_{k=1}^{N_c} r_v^2$ is a sum of N_c χ^2-distributed values (for each class), which itself follows a χ^2-distribution with $N_c d$ DOF. In our problem, d equals the number of included imaging modalities only if their information is completely independent. If all modalities contain the same information, then $d = 1$. So the effective DOF ranges in $N_c \leqslant v \leqslant N_c d$. Thus, there is an upper limit of u_v which is χ^2-distributed after proper shifting and scaling. As a consequence of this derivation, followed by experimental verification (see Section 3), we model u_v as a scaled and shifted χ^2-distributed random variable:

$$u_v \sim a\chi^2(v) + b, \tag{10}$$

and determine a, b, v by minimization using Powell's method (Powell, 1964):

$$O = \int_0^\infty \left(F_u(t) - (aF_{\chi^2}(t, v) + b) \right)^2 dt, \tag{11}$$

where $F_u(t)$ is the experimental cumulative density function (CDF) of u and $F_{\chi^2}(t, v)$ is the χ^2 CDF with v DOF, i.e., the regularized gamma function. Computed values u_v are corrected by the optimized parameters $\hat{a}, \hat{b}$ to yield a field of χ^2-distributed values with v DOF. As can be seen from the QQ-plot of an example (Fig. 1), our distributional assumption is sufficiently accurate.

2.3. Finding lesions

Accepting a specific type I error p, we can determine a threshold $u_t = F_{\chi^2}^{-1}(1 - p)$ from the inverse χ^2 CDF to find all lesion voxels. However, this simple approach assumes an (unrealistic) statistical independence of neighboring voxels. In analogy to similar considerations in the statistical analysis of functional imaging data, we perform a second-level analysis using the size of lesion voxel clusters above u_t. This approach is based on random field theory. Rather than repeating basics of this well-developed theory, we refer to the work of Adler (1981) for a thorough introduction, and will focus on the details relevant to our problem.

We consider our 3D spatial map of transformed joint probability values u as a random field of χ^2-distributed values. We want to attach a significance P value to each cluster of voxels above u_t that it represents a "true" lesion. The probability of finding a cluster of at least k voxels by chance is given by:

$$P(s_{max} \geqslant k) = 1 - \frac{\exp[-E(m)P(s > k)] - \exp[-E(m)]}{1 - \exp[-E(m)]}, \tag{12}$$

where $E(m)$ corresponds to the expected number of voxel clusters above u_t. If this probability is small, we consider the cluster a significant detection, i.e., a "true" lesion.

The *expected number of voxel clusters* $E(m)$ above u_t can be approximated by the expectation of the Euler characteristic of the random field (Worsley, 1994). Roughly speaking, the Euler characteristic corresponds to the number of clusters minus the number of holes plus the number of hollows in a 3D random field. For large u_t, holes and hollows tend to disappear (Adler, 1981), and the Euler characteristic approximates the number of clusters. For a χ^2 random field in R^3 with v DOF, the expected number of clusters is approximated by Worsley (1994):

$$E(m) = \sqrt{\frac{|\Lambda| u_t^{v-3}}{(2\pi)^3 2^{v-2}}} \frac{n \exp(-0.5 u_t)}{\Gamma(0.5v)} [u_t^2 - (2v - 1)u_t + (v - 1) \\ \times (v - 2)], \tag{13}$$

where Γ denotes the gamma function, and n corresponds to the number of voxels in the field. The spatial smoothness $|\Lambda|$ of the random field is given as (Worsley, 1994):

$$|\Lambda| = \frac{1}{n - 1} \sum_{v=1}^{n} (d_v \, d_v^T) \quad \text{where } d_v = \frac{(\partial u/\partial t)_v}{(2\sqrt{u_v})}, \tag{14}$$

and $(\partial u/\partial t)$ is approximated by central differences in each dimension.

The *probability for finding a cluster of size k* is, according to Cao (1999):

$$P(s > k) = \exp\left(-0.5\left[\frac{kE(m)Y}{E(n)}\right]^{\frac{2}{v}}\right), \tag{15}$$

where the integral $Y = \int_0^\infty y^{v/2}\exp(-0.5y)dy$ can be computed by discrete integration techniques.

Finally, the *expected total number of voxels* $E(n)$ that have a value greater than u_t is determined by integrating the tail of the χ^2 probability density function (PDF) $f(\cdot)$ beyond u_t:

$$E(n) = n\int_{u_t}^\infty f_{\chi^2}(x)dx = n(1 - F_{\chi^2}(u_t/2, v/2)). \tag{16}$$

Note that the computation of a probability according to Eq. (12) requires the cluster size k only. The threshold u_t, DOF v, the expectation of the total number of voxels n and the smoothness $|A|$ are constant for a given dataset, so $E(m), E(n)$ and Y can be precomputed.

We summarize the lesion detection process as follows: (1) Compute u_v for each voxel. (2) Determine a, b, and v according to Eq. (11), and transform u_v into a χ^2 random field. (3) Determine the smoothness $|A|$ according to Eq. (14) and compute $E(m), E(n)$, and Y for this field. (4) Find connected components for which $u_v > u_t$, where u_t is a given threshold. (5) Use Eq. (12) to compute the probability P that the connected component occurs only by chance based on the χ^2 random field. Here, we consider a supra-threshold cluster in the probability map as a lesion if $P < 0.05$. By introducing the χ^2 random field theory, we take the spatial extent of a lesion into account.

2.4. Remarks

MR image preprocessing includes co-registration, skull stripping, tissue segmentation, and intensity non-uniformity correction. T_1-weighted MR images were transformed into the T_2-weighted MR image space of the same subject by a co-registration procedure based on mutual information (Maes et al., 1997), so both manual and automatic lesion segmentation were performed in the T_2-weighted MR image space. Non-brain tissues (e.g., skull and skin) were removed by a fast and robust algorithm based on a deformable model (Smith, 2002). For estimating the distribution parameters, a prior segmentation into three classes (WM, GM, and CSF) was computed using a segmentation approach based on hidden Markov random fields (Zhang et al., 2001), and intensity non-uniformity was corrected by a non-parametric method (Sled et al., 1998).

In essence, the lesion segmentation method is based on outlier detection, but not all outliers indicate lesions. CSF voxels close to GM often appear as outliers partially due to co-registration errors, partially due to the partial volume effect. We add a constraint to remove some false-positives: The T_2 image intensity of lesion candidate voxels must be higher than the mean T_2 image intensity of WM, which is reasonable because lesions have high intensity in T_2 images.

The algorithm was implemented using the C++ programming language and tested on a Linux server with an AMD Athlon 64 processor (2.21 GHz) and 4 GB of memory. It takes about 6 min to segment subject data including T_1-weighted and T_2-weighted MR images, which is fast enough for studies involving large databases. Note that the only relevant parameters of our procedure are the threshold u_t (which is related to an accepted voxel-wise detection error), and the cluster-wise probability threshold P_t (chosen as $P_t = 0.05$ here).

3. Experimental results

Thirty subjects with mild or moderate dementia (10 males, 20 females, 78.27 ± 5.02 years old) were recruited from the Leipzig Longitudinal Study of Aging (LEILA 75+) (Riedel-Heller et al., 2000). Both T_1-weighted and T_2-weighted MRI brain data sets were acquired on a Siemens Vision 1.5 T scanner. The acquisition parameters for the T_1-weighted images were: TR 11.4 ms, TE 4.4 ms, matrix 256×256, 128 slices, voxel size $0.9 \times 0.9 \times 1.5$ mm. Parameters for the T_2-weighted images were: TR 5016 ms, TE 132 ms, 19 slices, matrix 357×512, voxel size $0.5 \times 0.5 \times 5$ mm, gap 1.5 mm. A neuroradiologist annotated WM lesions in both im-
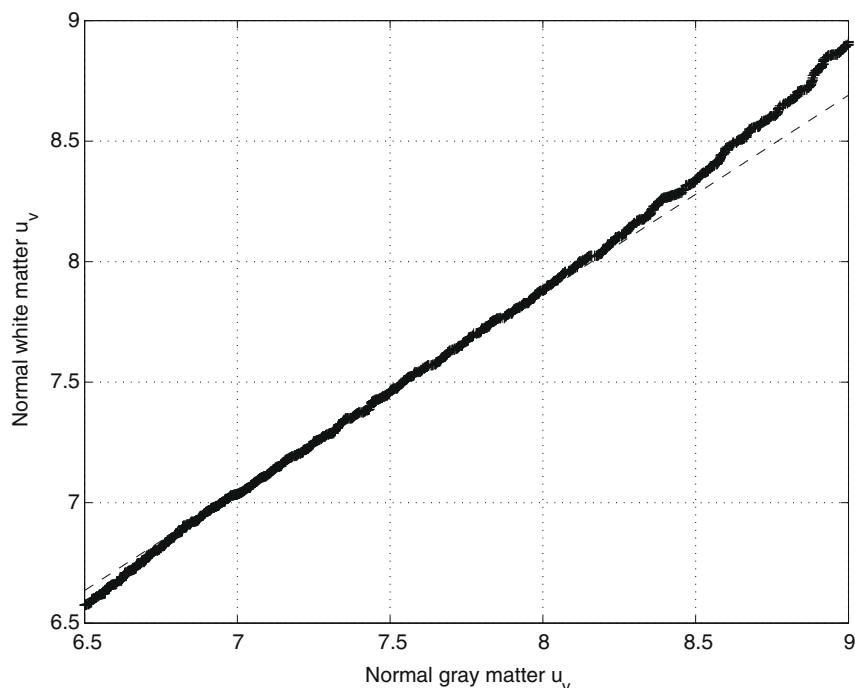

Fig. 2. A QQ-plot of experimental values of u of normal gray vs. white matter verifies the stationarity assumption.
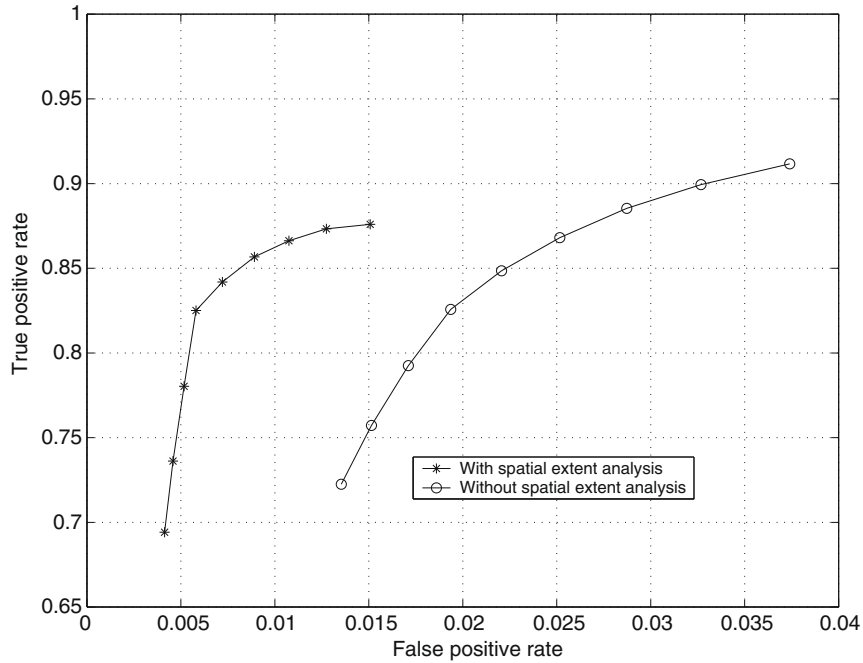
Fig. 3. A statistical rating of the significance of lesion clusters based on their extent remarkably improves detection accuracy. Note that curves are just lines connecting the data points.

age series mainly based on T_2-weighted images with T_1-weighted images available for reference.

We evaluated our unsupervised lesion segmentation algorithm by comparing the computed lesions with manual annotations considered as ground truth. We computed the average true-positive rate (ATPR), average false-positive rate (AFPR), and average similarity index (ASI) of the automatically segmented lesions and the ground truth. Suppose RA and RM are the two sets of voxels classified as lesions for a subject in two segmentations and RM is the

ground truth, then the true-positive rate (TPR), false-positive rate (FPR), and similarity index (SI) are defined as:

$$TPR = \frac{V(RA \cap RM)}{V(RM)}, \tag{17}$$

$$FPR = \frac{V(RA \cap \overline{RM})}{V(\overline{RM})}, \tag{18}$$
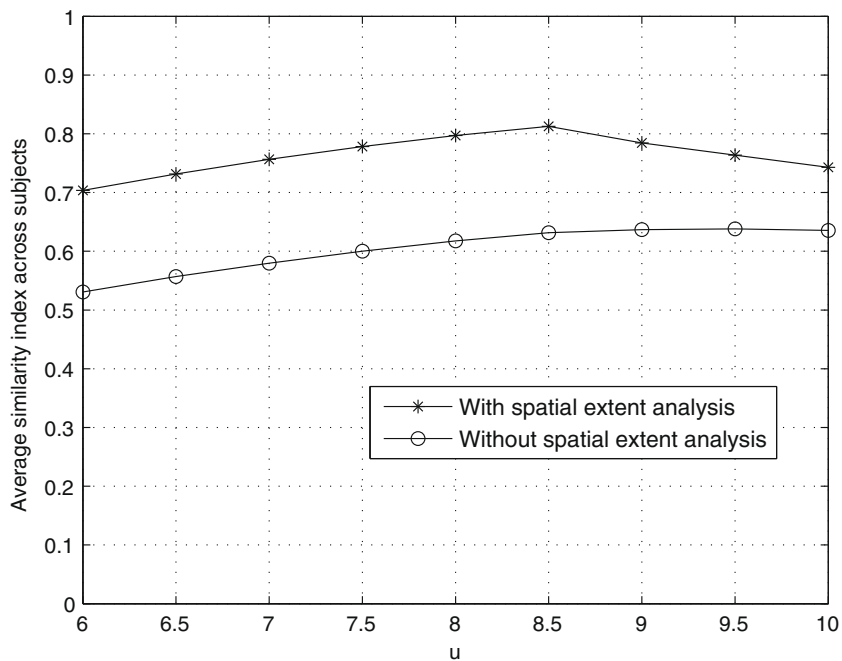
$$SI = \frac{2V(RA \cap RM)}{(V(RA) + V(RM))}, \tag{19}$$

Fig. 4. A statistical rating of the significance of lesion clusters based on their extent also improves the similarity with an expert's segmentation. Note that curves are just lines connecting the data points.
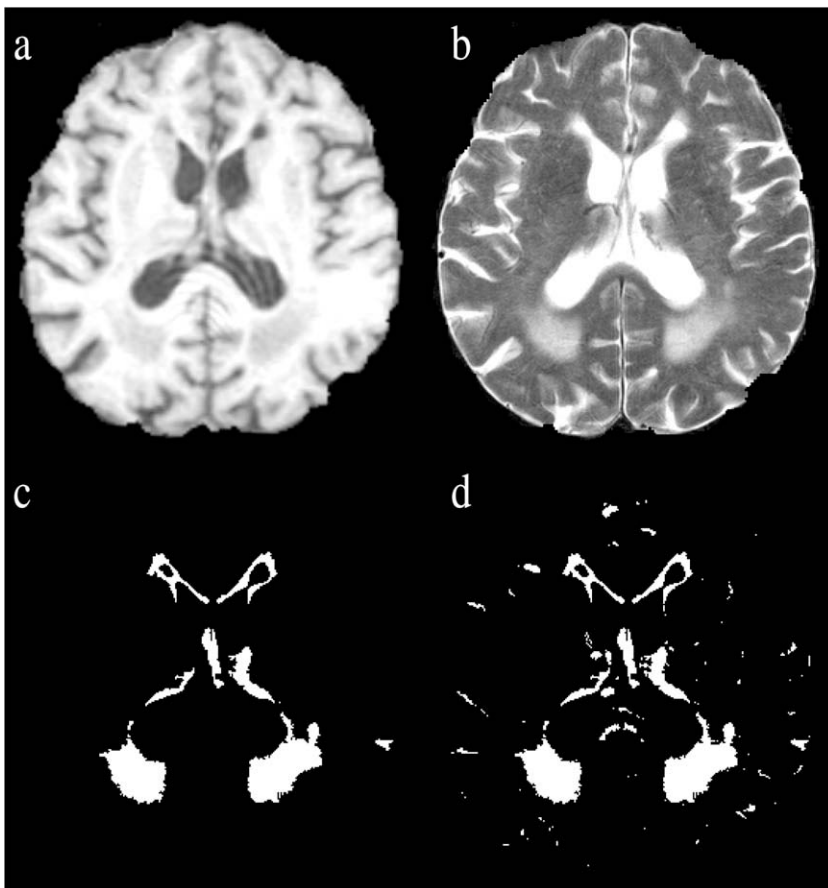
Fig. 5. Effect of rating the cluster extent: T_1-, T_2-weighted image (a, b), segmented lesions with (c) and without the application of cluster rating (d).

where $V(\cdot)$ corresponds to the number of elements in a set, and $\overline{RM}$ the complementary set of RM. The values of TPR, FPR, and SI are in the range of $[0,1]$. A TPR value of 1 indicates that all lesion voxels labeled in the ground truth segmentation were detected by the segmentations RA. A FPR value of 0 means that no voxels were wrongly detected as lesions by the segmentations RA. The SI indicates how
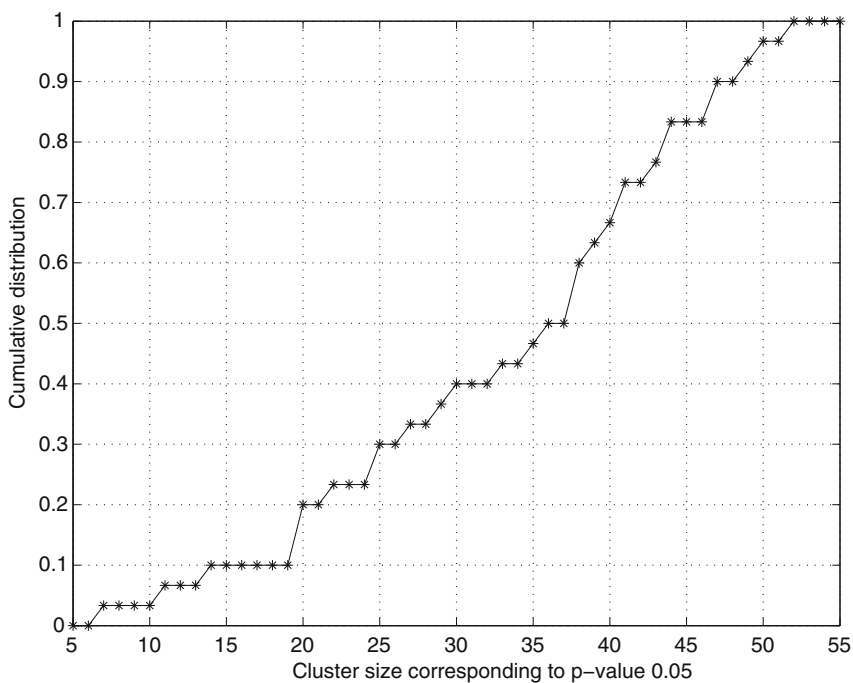

Fig. 6. Experimental CDF of the cluster size threshold for $p = 0.05$ in the data sample of this study.
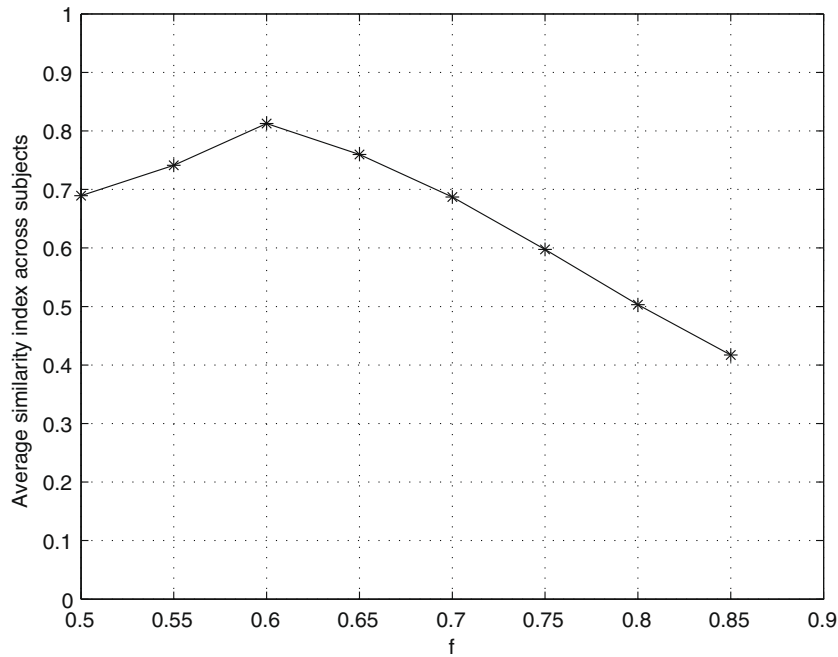
Fig. 7. Influence of the parameter f on the lesion segmentation results. Note that curves are just lines connecting the data points.

much the segmentations overlap. Bartko (1991) pointed out that a $SI > 0.7$ indicates a good agreement between two segmentations; however, the SI values depend greatly on the volume of the seg- mentation viewed as ground truth (Anbeek et al., 2004). In general, higher values of TPR and SI and smaller values of FPR indicate a bet- ter performance of the automatic method.
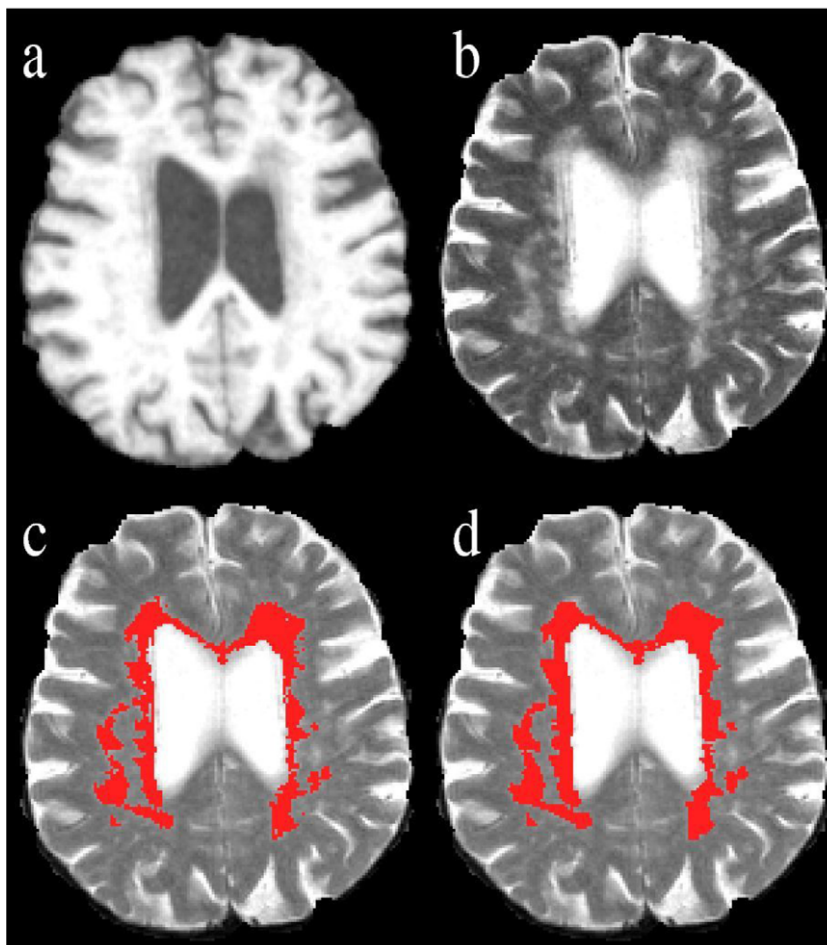

Fig. 8. Lesion segmentation results: T_1-, T_2-weighted image (a, b), segmented lesions (c), and expert annotation (d).

3.1. Validation of the assumed distribution of u_v

In this section, we demonstrate that our simplification of the distribution of u_v as a χ^2 distribution is reasonable. For an example subject, we computed the values of u_v at all locations except those labeled as lesions by the experienced rater. Fig. 1 shows a QQ-plot of the experimental values u_v against a shifted and scaled χ^2 distribution and validates our distributional assumption. Next, we collected experimental values u_v in GM and WM separately. The QQ-plot in Fig. 2 demonstrates that the assumed χ^2 random field is stationary across normal WM and GM.

3.2. Validation of the intra-rater manual segmentation variation

Four subjects were arbitrarily selected from the 30 experimental subjects and manually re-segmented for lesions by the same rater 3 months after the first manual segmentation to investigate the intra-rater manual lesion segmentation variation. The *SI* between the two manual segmentations was 0.71, which verified the reliability of the manual segmentation process.

3.3. Validation of the lesion segmentation accuracy

To demonstrate that the cluster size analysis improves performance, we applied the algorithm with and without this option to all datasets. Fig. 3 shows receiver-operating characteristics (ROC) curves while varying the threshold u_t. Though the maximum *ATPR* with χ^2 random-field-based cluster size analysis is a little smaller than those without the cluster size analysis, the *AFPR*

greatly decreased with the cluster size analysis, which can be seen from Fig. 4. The corresponding *ASI* is consistently and remarkably improved when the size criterion is included. Note that the best threshold is $u_t = 8.5$ at $ASI_{max} = 0.81$. To visualize the effect of the cluster size analysis, slices of the original T_1 and T_2-weighted MR images are shown in Fig. 5, showing the detected lesions with and without cluster size analysis, from which we can see that the statistical rating of a lesion's spatial extent correctly removed several false-positive detections.

We compiled the cluster size corresponding to $p = 0.05$ for all datasets as a CDF in Fig. 6 to demonstrate the range of the size threshold for false-positive detections. Because this threshold varies considerably across datasets, we conclude that the cluster size threshold is dependent on characteristics of a specific dataset and cannot be replaced by a simple (global) threshold. The χ^2 random-field-based analysis provides a rigorous assessment of the lesion cluster size, based on the smoothness of the statistical map.

Experiments were conducted to verify that our cluster size analysis based on random field theory yields better results than applying a simple global threshold on the cluster size. This global threshold was varied from 1 to 101 voxels with a step of 2 voxels. The maximum $ASI = 0.75$ was found at a threshold of 76 voxels, which is smaller than the maximum $ASI = 0.81$ obtained by our cluster size analysis. The *SI* range obtained in the 30 subjects was [0.70, 0.90] for the cluster size analysis, and [0.43, 0.89] if the optimal global threshold was applied. A paired *t*-test of the *SI* values demonstrated the superiority of our cluster size analysis.

Next, we studied the influence of the parameter f in the MCD estimator on the lesion segmentation result. We varied f from
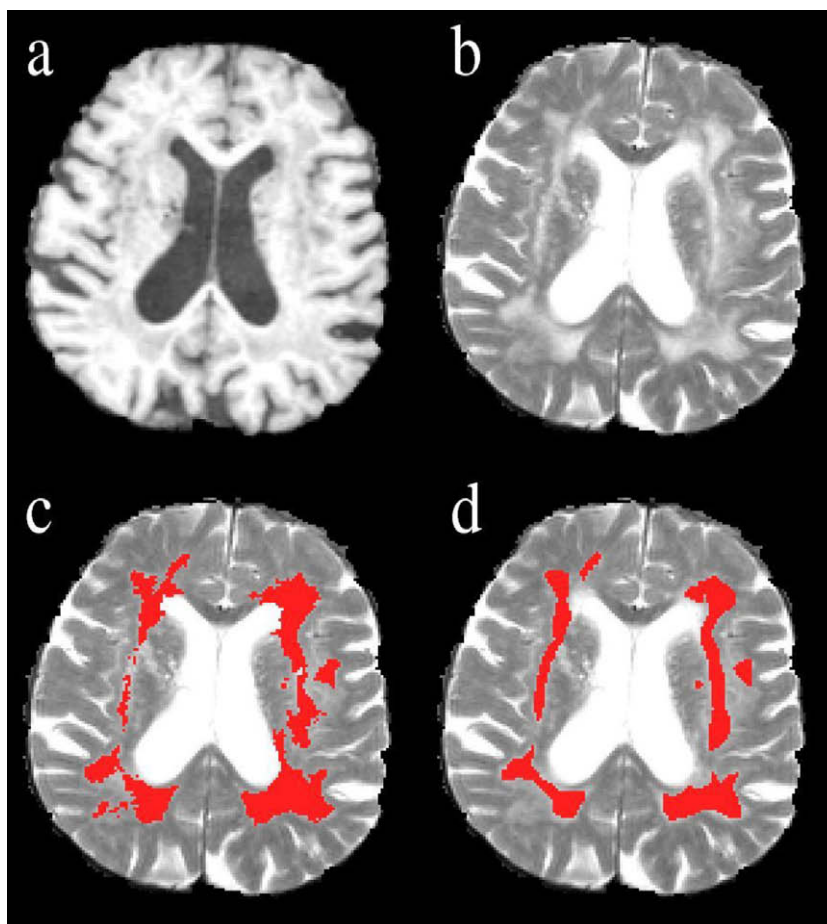

Fig. 9. Lesion segmentation results: T_1-, T_2-weighted image (a, b), segmented lesions (c), and expert annotation (d).
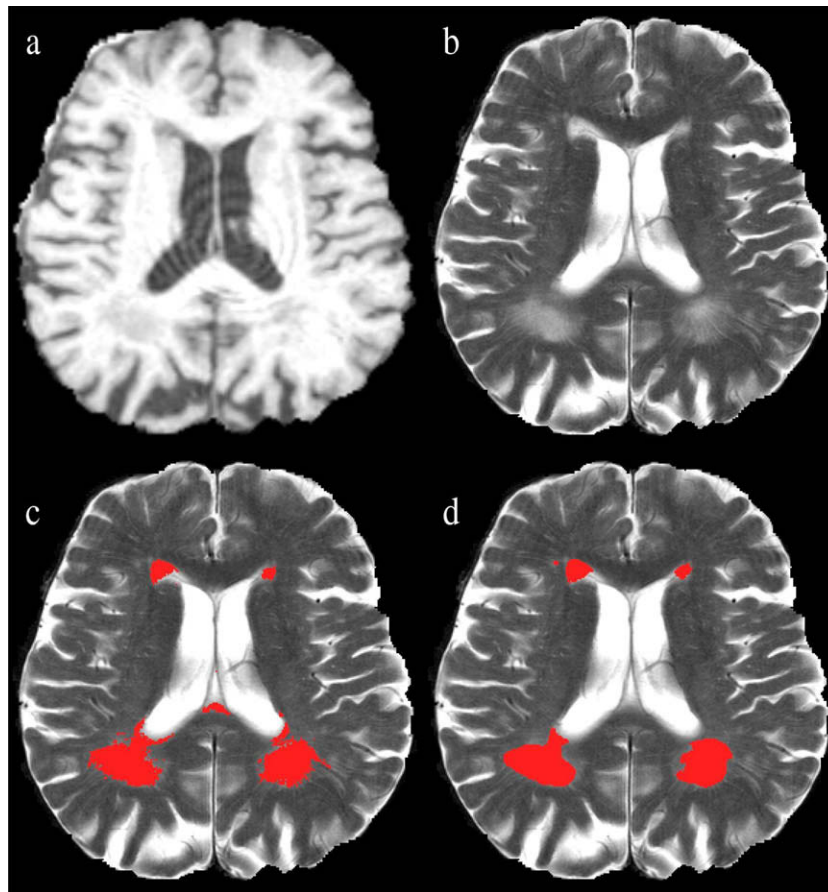
Fig. 10. Lesion segmentation results: T_1-, T_2-weighted image (a, b), segmented lesions (c), and expert annotation (d).

0.5 to 0.85 with a step size of 0.05. The *ASI* across all subjects is shown in Fig. 7. The best fraction value was 0.6, corresponding to the best *ASI* value of 0.81.

We also compared our lesion segmentation method with an approach based on a Bayesian classifier (Herskovits et al., 2008). Although this classifier was effective and accurate when fluid-attenuated inversion-recovery (FLAIR) MR images were available (Herskovits et al., 2008), the *ASI* was only 0.14 when only T_1 and T_2 MR images were used, while the algorithm described here achieved a superior *ASI* value of 0.81.

Finally, manually and automatically segmented lesions are compiled to illustrate the performance of our WM lesion segmentation method with medium (Figs. 8 and 9) and small lesion loads (Fig. 10).

4. Discussion

We proposed an automatic method for segmenting diffuse WM lesions of the human brain. Lesion voxels are detected as outliers of the intensity distribution of normal tissues in multi-sequence MR images. The resulting probability map is modeled as a χ^2 random field, and groups of connected voxels above a threshold are tested for their detection significance using the random field theory.

Rating the significance of lesion detection using the random field theory is the core contribution of this work. Our experiments demonstrate that lesion detection is considerably improved. In addition, it is easier and more meaningful to specify an accepted error for false positives (here, we used $p = 0.05$) than assuming an (arbitrary) global threshold for the cluster size. Note that this approach can easily be integrated with other lesion segmentation

methods that yield probability maps (Kruggel et al., 2008; van Leemput et al., 2001). Our framework can readily incorporate other MR imaging protocols such as fluid attenuation inversion-recovery (FLAIR) or proton density (PD) weighted MRI, which may further improve detection accuracy.

At a first glance, the simplification made when assuming a χ^2 distribution for u_v appears arguable. It is important to fit the long tail well, since we focus on outlier detection here. Our choice of u_t can be compared with a similar threshold used in functional MRI data analysis. Here, Gaussian random fields are used, and a threshold $u_t = 2.0$ is often assumed, which corresponds to an error probability of $p = 0.05$. For a χ^2 distribution with $v = 4$, an error probability of $p = 0.05$ corresponds to a threshold of $u_t = 9.48$, which is close to the optimal value of $u_t = 8.5$ derived from the SI analysis above. This finding underlines that our simplification of the distribution of u_v is reasonable and experimentally corroborated. We also demonstrated that the empirical distribution of u_v is well approximated by a χ^2 distribution.

The automatic estimation of a subject's WM lesion load is a valuable alternative to visual rating. Especially in longitudinal studies, automatically generated results are expected to be more reliable.

References

Adler, R.J., 1981. The Geometry of Random Fields. Wiley, New York.

Admiraal-Behloul, F., van den Heuvel, D.M.J., Olofsen, H., van Osch, M.J.P., van der Grond, J., van Buchem, M.A., Reiber, J.H.C., 2005. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28, 607–617.

Anbeek, P., Vincken, K.L., van Osch, M.J.P., Bisschops, R.H.C., van der Grond, J., 2004. Automatic segmentation of different-sized white matter lesions by voxel probability estimation. Med. Image Anal. 8, 205–215.

Bartko, J.J., 1991. Measurement and reliability: Statistical thinking considerations. Schizophrenia Bull. 17, 483–489.

Cao, J., 1999. The size of the connected components of the excursion sets of χ^2, t and F fields. Adv. Appl. Probab. 31, 579–595.

Chen, S., Zhang, D., 2004. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE Trans. System Man Cybernet. Part B 34, 1907–1916.

de Groot, J.C., de Leeuw, F.E., Oudkerk, M., van Gijn, J., Hofman, A., Jolles, J., Breteler, M.M.B., 2000. Cerebral white matter lesions and cognitive function: The Rotterdam scan study. Ann. Neurol. 47, 145–151.

Deary, I.J., Leaper, S.A., Murray, A.D., Staff, R.T., Whalley, L.J., 2003. Cerebral white matter abnormalites and lifetime cognitive changes: a 67-year follow-up of the scottish mental survey of 1932. Psychol. Aging 18, 140–148.

Dyrby, T.B., Rostrup, E., Baaré, W.F.C., van Straaten, E.C.W., Barkhof, F., Vrenken, H., Ropele, S., Schmidt, R., Erkinjuntti, T., Hansen, L.K., Waldemar, G., 2008. Segmentation of age-related white matter changes in a clinical multi-center study. NeuroImage 41, 335–345.

Gerig, G., Welti, D., Guttmann, C.R., Colchester, A.C., Szekely, G., 2000. Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. Med. Image Anal. 4, 31–42.

Gunning-Dixon, F.M., Raz, N., 2000. The cognitive correlates of white matter abnormalites in normal aging: A quantitative review. Neuropsychology 14, 224–232.

Herskovits, E.H., Bryan, R.N., Yang, F., 2008. Automated bayesian segmentation of microvascular white-matter lesions in the accord-mind study. Adv. Med. Sci. 53, 182–190.

Hirono, N., Kitagaki, H., Kazui, H., Hashimoto, M., Mori, E., 2000. Impact of white matter changes on clinical manifestation of alzheimer's disease: A quantitative study. Stroke 31, 2182–2188.

Hojjatoleslami, S.A., Kruggel, F., 2001. Segmentation of large brain lesions. IEEE Trans. Med. Imaging 20, 666–669.

Kruggel, F., Paul, J.S., Gertz, H.-J., 2008. Texture-based segmentation of diffuse lesions of the brain's white matter. NeuroImage 39, 987–996.

Lao, Z., Shen, D., Jawad, A., Karacali, B., Liu, D., Melhem, E.R., Bryan, R.N., Davatzikos, C., 2006. Automated segmentation of white matter lesions in 3D brain MR images, using multivariate pattern classification. In: Proc. 3rd IEEE Internat. Symp. on Biomedical Imaging, Arlington Virginia, USA, pp. 307–310.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation. Springer, Heidelberg.

Leys, D., Soetaert, G., Petit, H., Fauquette, A., Pruvo, J.P., Steinling, M., 1990. Periventricular and white matter magnetic resonance imaging hyperintensities do not differ between Alzheimer's disease and normal aging. Arch. Neurol. 47, 524–527.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16, 187–198.

Powell, M.J.D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput. J. 7, 155–162.

Riedel-Heller, S.G., Schork, A., Matschinger, H., Angermeyer, M.C., 2000. Recruiment procedures and their impact on the prevalence of dementia. Results from the leipzig longitudinal study of the aged (leila 75+). Neuroepidemiology 19, 130–140.

Rousseeuw, P.J., 1985. Multivariate estimation with high breakdown point. Math. Stat. Appl. B, 283–297.

Rousseeuw, P.J., van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41, 212–223.

Skoog, I., Berg, S., Johansson, B., Palmertz, B., Andreasson, L.-A., 1996. The influnce of white matter lesions on neuropsychological functioning in demented and non-demented 85-year-olds. Acta Neurol. Scandinavica 93, 142–148.

Sled, J.G., Zijdenbos, A.P., Evanse, A.C., 1998. A nonparametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155.

Udupa, J.K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M.A., Grossman, R.I., 1997. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. IEEE Trans. Med. Imaging 16, 598–609.

van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20, 677–688.

Worsley, K.J., 1994. Local maxima and the expected euler characteristic of excursion sets of χ^2, F and t fields. Adv. Appl. Probab. 26, 13–42.

Yang, F., Jiang, T., Zhu, W., Kruggel, F., 2004. White matter lesion segmentation from volumetric MR images. Lect. Note Comput. Sci.: Med. Imaging Augment. Real. 3150, 113–120.

Zhang, D.-Q., Chen, S.-C., 2004. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. Artif. Intell. Med. 32, 37–50.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden markov random field model and the expectation–maximization algorithm. IEEE Trans. Med. Imaging 16, 45–57.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans. Med. Imaging 13, 716–724.