# Automated Segmentation of MS Lesions from Multi-channel MR Images

Koen Van Leemput[1], Frederik Maes[1], Fernando Bello[2], Dirk Vandermeulen[1], Alan Colchester[2], and Paul Suetens[1]

[1] Medical Image Computing, Radiology-ESAT
Katholieke Universiteit Leuven
UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
[2] Neurosciences Medical Image Analysis Group, Electronic Engineering Laboratory
University of Kent of Canterbury, Canterbury Kent CT2 7NT, UK
Koen.VanLeemput@uz.kuleuven.ac.be

**Abstract.** Quantitative analysis of MR images is becoming increasingly important as a surrogate marker in clinical trials in multiple sclerosis (MS). This paper describes a fully automated model-based method for segmentation of MS lesions from multi-channel MR images. The method simultaneously corrects for MR field inhomogeneities, estimates tissue class distribution parameters and classifies the image voxels. MS lesions are detected as voxels that are not well explained by the model. The results of the automated method are compared with the lesions delineated by human experts, showing a significant total lesion load correlation and an average overall spatial correspondence similar to that between the experts.

## 1   Introduction

The role of magnetic resonance (MR) imaging in assessing the progression of multiple sclerosis (MS) and in monitoring the effect of a drug therapy is of increasing importance. This is caused by the higher sensitivity and objectivity of an MR-based surrogate index compared to traditional clinical disability scales [1], in combination with the widespread availability of MR imaging.

In clinical trials, manual analysis of the MR images by human experts is too time-consuming because of the large amounts of data involved. Furthermore, the inter- and intra-observer variability associated with manual delineations complicates the analysis of the results. Finally, it is not clear how a human rater combines information obtained from the different channels when multi-spectral MR data are examined.

Therefore, considerable efforts have been made by the medical imaging community to come up with fast automated methods that produce more objective and reproducible results [2,3,4]. However, most of these techniques still require some human interaction and/or ad hoc processing steps, which can make the results not fully objective. Zijdenbos *et. al* [5,6] proposed and validated a fully automated pipeline for MS lesion segmentation from T1-, T2- and PD-weighted

images. However, they used a fixed classifier that must be retrained in cases where different scanner types or pulse sequences produce contrast variations.

We present a fully automated technique for segmenting MS lesions from T1-, T2-, and PD-weighted MR images that automatically retrains the classifier. More specifically, a model-based iterative algorithm is used that simultaneously corrects for MR field inhomogeneities, estimates tissue class distribution parameters, and classifies the image voxels. The MS lesions are detected as voxels that are not well explained by the model.

The paper is organized as follows. The method is explained in section 2. Section 3 presents a validation of the automatic lesion segmentation by comparing it with the delineations of human experts. We discuss the results in section 4 and briefly formulate our conclusions in section 5.

## 2   Method

### 2.1   Model-Based Segmentation of Normal MR Images of the Brain

Recently, we described a model-based method for fully automated classification of MR images of normal brains [7,8]. Since we will build on this method in the rest of this paper, we briefly describe it here.

Suppose that there are $J$ tissue types or so-called classes present in an MR image of the brain. Let the intensity of voxel $i$ be denoted as $y_i$, then $y = \{y_1, \ldots, y_i, \ldots, y_N\}$ describes the observed intensities where $N$ is the total number of voxels. An often-used simple model for the intensity distribution of a voxel $i$ that belongs to class $j$ is a normal distribution with parameters mean $\mu_j$ and variance $\sigma_j^2$, grouped in $\theta_j = \{\mu_j, \sigma_j^2\}$. MR images often suffer from an imaging artifact that introduces a spatially smoothly varying intensity inhomogeneity or so-called bias field in the images. We model the bias field in image $y$ as a linear combination $\sum_k c_k \phi_k(x)$ of $K$ smoothly varying basis functions $\phi_k(x)$, where $x$ denotes the spatial position. With this model, the intensity distribution of a voxel $i$ that belongs to class $j$ is given by $p(y_i \mid \Gamma_i=j, \theta_j, C) = G_{\sigma_j}(y_i - \mu_j - \sum_k c_k \phi_k(x_i))$ where $G_\sigma()$ denotes a zero-mean normal distribution with variance $\sigma^2$, $\Gamma_i \in \{1, \ldots, j, \ldots, J\}$ denotes the class to which voxel $i$ belongs, and $C = \{c_1, \ldots, c_K\}$ contains the bias field parameters. Assuming that the tissue types of the voxels are independently sampled from the $J$ classes with some known probability $p(\Gamma_i=j)$, the overall model becomes $p(y \mid \theta, C) = \prod_i p(y_i \mid \theta, C)$ where

$$p(y_i \mid \theta, C) = \sum_j p(y_i \mid \Gamma_i=j, \theta_j, C)p(\Gamma_i=j) \tag{1}$$

and $\theta = \{\theta_1, \ldots, \theta_J\}$ denotes all the normal distribution parameters.

As shown in [7], assessing the maximum likelihood (ML) model parameters $\{\theta, C\}$ given the observed intensities $y$ results in an iterative so-called Generalized Expectation-Maximization (GEM) algorithm that interleaves the fol-

lowing equations:

$$p(\Gamma_i{=}j \mid y_i, \theta, C) = \frac{p(y_i \mid \Gamma_i{=}j, \theta_j, C)p(\Gamma_i{=}j)}{\sum_j p(y_i \mid \Gamma_i{=}j, \theta_j, C)p(\Gamma_i{=}j)} \qquad (2)$$

$$\mu_j = \frac{\sum_i p(\Gamma_i{=}j \mid y_i, \theta, C)(y_i - \sum_k c_k \phi_k(x_i))}{\sum_i p(\Gamma_i{=}j \mid y_i, \theta, C)} \qquad (3)$$

$$\sigma_j^2 = \frac{\sum_i p(\Gamma_i{=}j \mid y_i, \theta, C)(y_i - \mu_j - \sum_k c_k \phi_k(x_i))^2}{\sum_i p(\Gamma_i{=}j \mid y_i, \theta, C)} \qquad (4)$$

$$C = (A^T W A)^{-1} A^T W R \ , \ A_{ik} = \phi_k(x_i) \ , \ R_i = y_i - \tilde{y}_i \ , \ W = \text{diag}(w_i) \quad (5)$$

$$\text{where} \quad \tilde{y}_i = \frac{\sum_j w_{ij} \mu_j}{\sum_j w_{ij}} \ , \ w_i = \sum_j w_{ij} \ , \ w_{ij} = \frac{p(\Gamma_i{=}j \mid y_i, \theta, C)}{\sigma_j^2}$$

This algorithm interleaves the following 3 steps until convergence: classification of the voxels (equation 2), estimation of the normal distribution parameters (equations 3 and 4), and estimation of the bias field (equation 5).

As described in [7], the method can be fully automated by introducing a digital brain atlas that contains spatially varying prior probabilities for gray matter, white matter and csf. After affine registration of the study image with a T1 template associated with the atlas, these prior probability maps can be used to initialize the algorithm, which makes the method fully automated. Additionally, the atlas spatially constrains the classification since it contains a spatially varying prior $p(\Gamma_i = j)$ that is used in equation 2.

Previously, we described the algorithm and its practical implementation in more detail. The interested reader is referred to [7,8]; suffice it here to say that the method is easily extended to multi-spectral MR images by substituting the normal distributions with multi-variate normal distributions with mean $\boldsymbol{\mu}_j$ and covariance matrix $\Sigma_j$.

## 2.2  Adaptation for Automated MS Lesion Segmentation

The simple mixture model described above works satisfactorily for MR images of normal brains. However, it does not include a model for MS lesions. Therefore, those lesions could be detected as voxels that are not well explained by the mixture model. We adopt the approach that was proposed by Guillemaud and Brady [9] for modeling non-brain tissues in MR images by adding a uniform intensity distribution to the mixture model, i.e. equation 1 is substituted by

$$p(y_i \mid \theta, C) = \sum_j p(y_i \mid \Gamma_i{=}j, \theta_j, C)p(\Gamma_i{=}j) + \lambda p(\Gamma_i{=}reject)$$

where $\lambda$ is a small constant that is defined by the condition that the integral over all the intensities is unity.

Similar to the approach of Guillemaud and Brady, it can be shown that equations 2, 3 and 4 remain valid, provided that the new rejection class is added

to equation 2. Voxels that are not well explained by the normal distributions, such as MS lesions, are pushed into the uniform rejection class. Also equation 5 for the bias estimation remains unchanged except that the weights $w$ are now only calculated with respect to the normal distributions. That is, voxels that are rejected from the normal distributions have a zero weight for the estimation of the bias field.

Almost 95 % of the MS lesions are located inside white matter. This information can be added to the model by assigning the atlas prior probability map of white matter to $p(\Gamma_i=reject)$. Besides this spatial constraint, additional intensity constraints can be added. We use multi-spectral MR images that consist of T1-, T2- and PD-weighted images. MS lesions have an intensity between that of white matter and csf in T1, and appear hyper-intense in T2 and PD. We therefore exclude voxels with an intensity darker than the mean of csf in T1 or darker than the mean of gray matter in T2 and PD from the rejection class.

Upon convergence of the adapted GEM algorithm, a classification of the voxels is obtained along with an estimation of the parameters $\theta$ of the normal distributions and the bias field parameters $C$. The MS lesions can be expected to be found in the rejection class.

## 2.3   Post-Processing

In practice, the rejection class does not contain the more subtle MS lesions, while on the other hand, non-lesion voxels that are not well explained by the normal distributions also end up in the rejection class. We therefore add a post-processing step in order to decrease the number of false positives and false negatives.

**Final Classification Rule** Given a single normal distribution with mean $\mu_j$ and variance $\sigma_j^2$, an intensity $y_i$ can be said to be abnormal with respect to this distribution if its so-called mahalanobis-distance $d_j^i = (y_i - \mu_j)/\sigma_j$ exceeds a predefined threshold. In the case of a mixture of normal distributions, however, assessing the abnormality of an intensity $y_i$ is more involved. Intuitively, an intensity is abnormal if the probability that it is generated by the mixture model is small, i.e. when the condition

$$\sum_j G_{\sigma_j}(y_i - \mu_j - \sum_k c_k \phi_k(x_i))p(\Gamma_i=j) \quad \leq \quad \kappa p(\Gamma_i=lesion)$$

holds with $\kappa$ a small threshold. However, this would give bad results since a voxel that belongs to a class with a small variance is only detected as abnormal if its mahalanobis-distance is very large compared to a voxel belonging to a class with a large variance. The same rationale also explains the false positives and false negatives in the rejection class of the adapted GEM algorithm. Furthermore, it is not clear how $\kappa$ should be chosen.

Instead, we weight each normal distribution with its variance, i.e. a voxel is labeled as MS lesion if the condition

$$\sum_j \sigma_j G_{\sigma_j}(y_i - \mu_j - \sum_k c_k \phi_k(x_i)) p(\Gamma_i{=}j) \quad \leq \quad \kappa p(\Gamma_i{=}lesion) \qquad (6)$$

holds. The meaning of $\kappa$ is now clear: it indirectly defines a mahalanobis-distance threshold above which a voxel is detected as abnormal, independent of the variance of the classes. This can be written more explicitly by $\kappa = 1/\sqrt{2\pi}exp(-0.5T^2)$ with $T$ a mahalanobis threshold that we experimentally set to $\sqrt{3}$.

Given the parameters $\theta$ and $C$ as calculated by the GEM algorithm, voxels where equation 6 holds, are classified as MS lesions. The other voxels are classified following equation 2.

**Markov Random Fields** Besides MS lesions, other voxels exist that are not well explained by the mixture model and, as a result, are misclassified as MS lesion. This is typically true for partial volume voxels along the sulci. Such misclassifications could be discarded by only withholding lesions at locations where the white matter a priori map exceeds a certain threshold (cf Zijdenbos *et al.* [5]), but this method is too crude in our experience.

Instead, we incorporate contextual information in the final segmentation process by making use of Markov-Random-Fields (MRF's). The segmentation $\Gamma$ is assumed to be the realization of a random process where the probability that voxel $i$ belongs to class $j$ depends on the classification of its neighbors. The Hammersley-Clifford theorem states that the configurations of such a random field obey the distribution $p(\Gamma) = Z^{-1}exp(-U(\Gamma))$ where $U(\Gamma)$ is an energy function and $Z$ is a normalization constant. We use the Potts model: $U(\Gamma) = \sum_i \sum_j \beta_{\Gamma_i,j} u_{i,j}$ where $u_{i,j}$ counts the number of neighbors of voxel $i$ that belong to class $j$ and $\beta_{l,j}, 1 \leq l, j \leq J$ are MRF parameters. We estimated these MRF parameters from an image that was manually labeled into grey matter, white matter, MS lesions, csf and non-brain tissues, using a histogramming technique [10]. Since the slice thickness in MRI can vary widely, we only use the 8 in-plane neighbors, although a full 3D neighborhood could also be used if a manual segmentation exists to estimate the appropriate 3D MRF parameters.

We incorporate contextual information in the final classification process by using the so-called Iterated-Conditional-Modes algorithm (ICM) [11]. More specifically, the prior probability that voxel $i$ belongs to class $j$ depends on the classification of its neighbors: $p(\Gamma_i{=}j) \sim exp(-\sum_j \beta_{\Gamma_i,j} u_{i,j})$ This prior replaces the atlas in the post-processing step, except for $p(\Gamma_i{=}lesion)$ where it is multiplied with the atlas prior probability for white matter. Starting from the segmentation obtained with the final classification rule as described above, we calculate $p(\Gamma_i{=}j)$ and re-apply the same rule with the updated prior. This is repeated until the classification stabilizes, for which 8 iterations are sufficient in our experience.

# 3   Results on BIOMORPH MS Data

As part of the BIOMORPH project [12], we analyzed 12 serial scans from each of 20 MS patients, where each scan contained low-resolution T1-, T2- and PD-weighted images (24 axial 256x256 slices, voxel dimensions 0.9x0.9x5.5 mm$^3$). In addition to these time series, for each patient there was also at least one higher-resolution scan with the same modalities (52 axial 256x256 slices, voxel dimensions 0.9x0.9x2.5 mm$^3$). We processed these images after registering and resampling the T1-weighted images to the corresponding T2-weighted images using the affine multi-modality registration algorithm based on maximization of mutual information of Maes *et al.* [13]. The PD images were assumed to be perfectly aligned with the T2 images since they were acquired simultaneously. The images were then spatially normalized with the atlas by registering the T2-weighted images with the T1 template associated with the atlas using the same registration method.

## 3.1   Validation on Low-Resolution Images

From 10 of the patients, 2 consecutive time points were manually analyzed by tracing MS lesions using only the T2-weighted images. We compared the delineation of the automatic algorithm with these expert segmentations by comparing the so-called total lesion load (TLL), measured as the number of voxels that were classified as MS lesion, on these 20 scans. Figure 1 shows the TLL of the expert segmentation along with the TLL of the automatic method. Also shown is the linear regression: the slope is unity, and the intercept is close to zero. The correlation coefficient between the automatically and manually detected lesion volumes is high: 0.96, $p < 10^{-10}$. However, a paired t-test reveals a significant difference between the TLL estimated by the expert and the TLL of the automated method (p = 0.029).
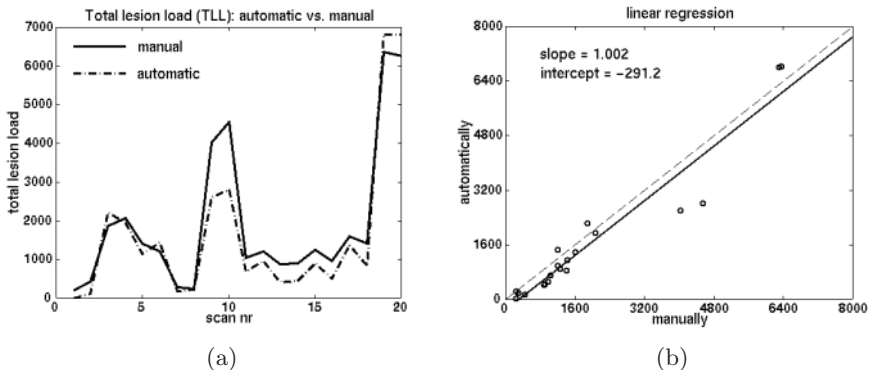


(a)                                        (b)

**Fig. 1.** Comparison between manual and automatic tracings of MS lesions on 20 low-resolution scans: total lesion load for each scan (a), and linear regression (b)
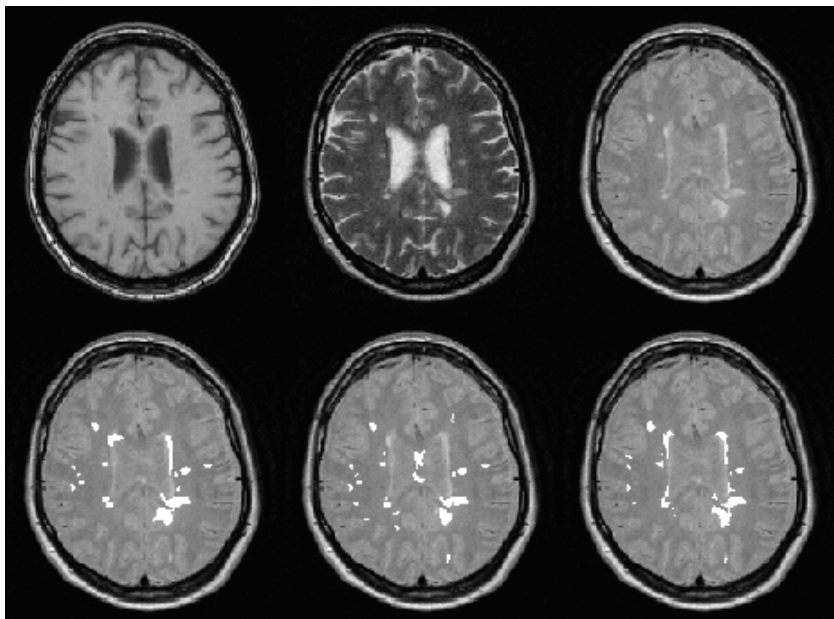
**Fig. 2.** Visual comparison of the MS lesion labeling by two experts and by the algorithm. Top: T1-, T2- and PD-weighted image. Bottom: MS lesions overlayed in bright color on the PD-weighted image. Left to right: expert 1, expert 2, automatic algorithm

### 3.2   Results on High-Resolution Images

In addition to the manual expert segmentations on 20 low-resolution scans, 3 high-resolution (HR) scans from different patients were analyzed by 2 different human experts. Figure 2 shows a representative slice of the data with the lesion segmentation of the two experts and the automatic algorithm overlayed in bright color. The TLL of each of the experts and the TLL of the automatic algorithm are depicted in table 1. Expert 1 consistently found a larger lesion volume than expert 2, while the volume obtained by the automatic algorithm did not show a systematic relationship with that segmented by the human observers .

   Comparing the TLL of two raters does not take into account any spatial correspondence of the segmented lesions [14]. We therefore calculated indices, the definition of which is given in table 2, which take account of the degree of correspondence between two segmentations: the similarity index which was previously used by Zijdenbos *et al.* [2], the overlap index, and the global spatial correspondence indices recently proposed by Bello and Colchester [14].

   The correspondence indices between expert 1, expert 2 and the automatic algorithm are depicted in table 3. We fused the segmentation maps of all 3 scans for each rater in order to have a single average index between each pair of raters. The non-symmetric nature of the global spatial correspondence indices

is apparent in table 3 where the columns have been considered as the reference segmentation and the rows as the current observation. Thus, expert 1 has an average overall correspondence with expert 2 of 50.9 and expert 2 has an average overall correspondence of 43.8 with expert 1. This means that, on average, roughly half of the voxels labeled as lesion by expert 1 were also identified as lesion by expert 2, and that the common lesion voxels have a greater correspondence with the segmentations by expert 2 than with those by expert 1.

Regarding the automatic segmentation, the average similarity and overlap indices between the automatic method and the experts are smaller than between the experts, in particular for expert 1. The global spatial correspondence measures show the results of the automatic method to have an average overall correspondence of 33.4 with expert 1 and 39.8 with expert 2. At the same time, the segmentations done by expert 1 have an average overall correspondence of 39.9 with the automatic method, whereas those by expert 2 have an average overall correspondence of 40.8. These values indicate that the voxels identified as lesion by either of the experts and the automatic method in both cases had a slightly better correspondence with the automatic segmentation than with the manual segmentation. As expected from the definitions in table 2, the values of S are higher than those of O and C. However, all three measures show that the agreement between the experts is, on average, slightly better than between the automatic method and either of the experts, and that the automatic method has a better agreement with expert 2 than with expert 1.

**Table 1.** Total lesion load by two human experts and the automatic algorithm for 3 high-resolution data sets

| TLL | scan 1 | scan 2 | scan 3 |
|---|---|---|---|
| expert 1 | 5303 | 939 | 8172 |
| expert 2 | 3772 | 598 | 6745 |
| automatic | 4609 | 228 | 8789 |

**Table 2.** Definition of spatial correspondence indices. $V_1$ and $V_2$ denote the total lesion volume of rater 1 and 2, and $V_{12}$ denotes the volume of the voxels that both indicated as lesion. $I(\boldsymbol{X}; \boldsymbol{Y})$ is the average total mutual information between rater 1 ($\boldsymbol{X}$) and rater 2 ($\boldsymbol{Y}$). $H(\boldsymbol{X})$ and $H(\boldsymbol{Y})$ denote the entropy of rater 1 and rater 2, respectively

| symbol | name | definition |
|---|---|---|
| S | similarity index | $2V_{12}/(V_1 + V_2)$ |
| O | overlap index | $V_{12}/(V_1 + V_2 - V_{12})$ |
| C | global spatial correspondence | $I(\boldsymbol{X}; \boldsymbol{Y}) / H(\boldsymbol{Y})$ and $I(\boldsymbol{X}; \boldsymbol{Y}) / H(\boldsymbol{X})$ |

**Table 3.** Spatial correspondence indices between each pair of raters on the HR scans

|         |           | expert 1 | expert 2 | automatic |
|---------|-----------|----------|----------|-----------|
|         | expert 1  | 100      | 59.2     | 51.5      |
| S (%)   | expert 2  | 59.2     | 100      | 55        |
|         | automatic | 51.5     | 55       | 100       |
|         | expert 1  | 100      | 42       | 34.7      |
| O (%)   | expert 2  | 42       | 100      | 38        |
|         | automatic | 34.7     | 38       | 100       |
|         | expert 1  | 100      | 50.9     | 39.9      |
| C (%)   | expert 2  | 43.8     | 100      | 40.8      |
|         | automatic | 33.4     | 39.8     | 100       |

## 4   Discussion

We validated the method by comparing the total lesion load (TLL) of automatically detected lesions on 20 scans with those detected by a human expert. A paired t-test revealed a significant difference between the TLL's of the automatic and the human rater. However, as pointed out in [1], the most important requirement for an automated method is that its measurements change in response to a treatment in a manner proportionate to manual measurements. We therefore performed a linear regression analysis, rather than concentrating on the absolute difference in TLL between the automated method and that of the manual segmentations. The regression analysis shows that the automatic segmentations indeed change proportionately to the manual segmentations.

Comparing the TLL of lesions detected by two different raters does not take into account the spatial correspondence between the two segmentations. If the automated method is used to study time correlations of lesion groups and lesion patterns in MS time series, it is important that the lesions are also spatially correctly detected. For 3 patients, we therefore calculated measures which assess this degree of correspondence between two human experts and between each of the experts with the automatic algorithm. Although the inter-observer variability for the experts was very large, the average overall agreement between the two experts was still better than the agreement between any of the experts and the automated algorithm. However, the average overall agreement between expert 2 and the automatic method was comparable to that between the experts.

The segmentation results were also analysed qualitatively. It was notable that the different segmentation methods (expert 1, expert 2 and automatic) showed certain differences which were consistent across patients. Expert 1 segmented many more brainstem and cerebellar lesions. Expert 2 generally segmented a larger number of lesions in the hemispheres. The automatic method tended to place the lesion boundary inside that chosen by the experts. Overall, very careful scrutiny of the segmentations did not reveal consistent failings in any of the methods. MS lesion segmentation is challenging for human observers and auto-

mated methods alike. Disagreement over *boundary placement* of specific lesions contributed only a small fraction of the total error. More important was disagreement over whether or not a lesion was present in a certain locality (*object identification*).

In this paper, the automatic method was validated by comparison of the segmentations with those of human experts. The automated algorithm uses multi-spectral data, while the manual segmentations used for validation were only based on T2-weighted images, which might have introduced errors in the manual tracings. Although the results presented in this paper look promising, a more thorough validation and assessment of intra- and inter-observer variability associated with manual delineation will require data of repeated manual tracings based on multi-spectral data by several human experts, as well as a study of the spatial correspondence between individual lesions and between group of lesions.

## 5    Summary and Conclusions

This paper presented a fully automated model-based method for segmenting MS lesions from multi-channel MR images. The method simultaneously corrects for MR field inhomogeneities, estimates tissue class distribution parameters and classifies the image voxels. MS lesions are detected as voxels that are not well explained by the model. The results of the automated method were compared with the lesions delineated by human experts, showing a significant total lesion load correlation. When the degree of spatial correspondence between two segmentations was taken into account, considerable disagreement was revealed, both between the expert manual segmentations, and between expert and automatic methods. Qualitative evaluation of the results showed that the major source of disagreement lies in identification of lesions as opposed to boundary placement. There is no true gold standard available for evaluating methods, and correct identification of MS lesions remains a major challenge for human observers and automated methods alike.

## Acknowledgments

# References

1. A.C. Evans, J.A. Frank, J. Antel, and D.H. Miller. The role of MRI in clinical trials of multiple sclerosis: Comparison of image processing techniques. *Annals of Neurology*, 41(1):125–132, january 1997. 11, 19

2. A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724, december 1994. 11, 17

3. B. Johnston, M.S. Atkins, B. Mackiewich, and M. Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, 15(2):154–169, april 1996. 11

4. J.K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M.A. van Buchem, and R.I. Grossman. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging*, 16(5):598–609, october 1997. 11

5. A. Zijdenbos, A. Evans, F. Riahi, J. Sled, J. Chui, and V. Kollokian. Automatic quantification of multiple sclerosis lesion volume using stereotaxic space. In *Proceedings of Visualization in Biomedical Computing – VBC'96*, Lecture Notes in Computer Science, pages 439–448, 1996. 11, 15

6. A. Zijdenbos, R. Forghani, and A. Evans. Automatic quantification of MS lesions in 3d MRI brain data sets: Validation of INSECT. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 439–448. Springer, 1998. 11

7. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automatic segmentation of brain tissues and MR bias field correction using a digital brain atlas. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 1222–1229. Springer, 1998. 12, 13

8. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated bias field correction and tissue classification of MR images of the brain using a digital atlas. *IEEE Transactions on Medical Imaging*, october 1999. 12, 13

9. R. Guillemaud and M. Brady. Estimating the bias field of MR images. *IEEE Transactions on Medical Imaging*, 16(3):238–251, june 1997. 13

10. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, october 1999. 15

11. S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Computer Science Workbench. Springer, 1995. 15

12. European project on brain morphometry (BIOMORPH, EU-BIOMED2 project nr. BMH4-CT96-0845, 1996–1998). 16

13. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997. 16

14. F. Bello and A.C.F. Colchester. Measuring global and local spatial correspondence using information theory. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 964–973. Springer, 1998. 17