Automatic histogram-based segmentation of white matter hyperintensities using 3D FLAIR images

Rita Simões[a], Cornelis Slump[a], Christoph Mönninghoff[b],
Isabel Wanke[b], Martha Dlugaj[c], Christian Weimar[c]

[a]Signals and Systems Group; MIRA Institute for Biomedical Technology and Technical Medicine; University of Twente, Enschede, The Netherlands
[b]Department of Diagnostic and Interventional Radiology and Neuroradiology; University Hospital Essen, Germany
[c]Department of Neurology; University Hospital of Essen, Germany

ABSTRACT

White matter hyperintensities are known to play a role in the cognitive decline experienced by patients suffering from neurological diseases. Therefore, accurately detecting and monitoring these lesions is of importance. Automatic methods for segmenting white matter lesions typically use multimodal MRI data. Furthermore, many methods use a training set to perform a classification task or to determine necessary parameters. In this work, we describe and evaluate an unsupervised segmentation method that is based solely on the histogram of FLAIR images. It approximates the histogram by a mixture of three Gaussians in order to find an appropriate threshold for white matter hyperintensities. We use a context-sensitive Expectation-Maximization method to determine the Gaussian mixture parameters. The segmentation is subsequently corrected for false positives using the knowledge of the location of typical FLAIR artifacts. A preliminary validation with the ground truth on 6 patients revealed a Similarity Index of 0.73 ± 0.10, indicating that the method is comparable to others in the literature which require multimodal MRI and/or a preliminary training step.

Keywords: Brain, white matter hyperintensities, MRI, FLAIR, Gaussian Mixture Model

1. INTRODUCTION

1.1 Background

White matter hyperintensities (WMHs) are diffuse white matter abnormalities that typically present high intensities in T2-weighted images. They occur often in the elderly and have been shown to predict an increased risk of stroke and cognitive decline.[1,2] The monitoring of WMHs can therefore provide useful information on the early onset of these diseases.[3]

Several methods have been proposed to segment WMHs from MRI images,[4–6] most of them using various types of MRI modalities. The use of multimodal data presents several disadvantages. First, the acquired datasets must be coregistered, which is computationally intensive and prone to errors. In particular, motion artifacts are seen frequently in the MRI data from elderly patients, who are often not able to stay still during the whole acquisition period. This imposes a serious limitation on the registration algorithms and their outcomes.[7] Furthermore, it happens often that the segmentation of, for instance, T1 images is also affected by the presence of lesions.[5]

Fast fluid-attenuated inversion-recovery (FLAIR) is a T2-weighted MRI modality in which the cerebrospinal fluid (CSF) signal is attenuated. In FLAIR images, WMHs are characterized by high intensities within a range that only partially overlaps with that of normal brain regions, making this MRI modality well suited for lesion segmentation purposes.

A few methods have been proposed that perform histogram-based thresholding of FLAIR images. Hirono *et al.*[8] suggest an empirical threshold of 3.5 standard deviations from the white matter intensities. Their

Further author information: (Send correspondence to R.S.)
R.S.: E-mail: A.R.Lopessimoes@ewi.utwente.nl, Telephone: +31 53 489 2897

Medical Imaging 2012: Computer-Aided Diagnosis, edited by Bram van Ginneken, Carol L. Novak,
Proc. of SPIE Vol. 8315, 83153K · © 2012 SPIE · CCC code: 1605-7422/12/$18 · doi: 10.1117/12.911327

approach is semi-automatic, since it requires an initial manual rough estimate of the WMHs locations. Jack *et al.*[7] propose an automatic method that selects a threshold using data from 10 elderly patient datasets. Similarly, de Boer *et al.*[5] use a training set to find the most appropriate parameters for the determination of the WMH threshold. Subsequently, the false positives are corrected using information provided by other MRI modalities. More recently, Ong *et al.*[9] employ an outlier detection analysis to select the WMH threshold.

All of the histogram-based methods described above require a training set or at least some *a priori* information to determine an appropriate threshold for the white matter hyperintensities. In this work, we aim at an unsupervised method that can be applied to any FLAIR dataset without the need for a training phase.

Gaussian Mixture Models (GMM), estimated by the Expectation-Maximization (EM) algorithm, have been widely used in unsupervised image classification. They provide a statistical description of the voxels' intensities and allow for fuzzy classification.[10] In the traditional GMM-EM approach, only spectral information is considered. The FLAIR histogram-based WMH segmentation methods mentioned above take also only spectral information into account. However, it has been recognized that using only intensity information the methods become highly sensitive to noise. In particular, boundary detection becomes problematic in noisy images. Furthermore, the assumption that the voxel intensities are independent does not hold in practice. In reality, and intuitively, we can expect a certain voxel's value to be affected by those in its neighborhood.[11] Therefore, we use a modified GMM-EM approach that considers contextual information.

1.2 Contribution

In this work, we propose an unsupervised and fully automatic segmentation method that uses only three-dimensional FLAIR images. We model the data using a mixture of three Gaussians. The Gaussian parameters are determined using a modified Expectation-Maximization (EM) approach, which incorporates anisotropic contextual information in the E-step at each iteration. Finally, we perform a post-processing step to correct for false positives, using prior information about the typical location of FLAIR artifacts.[12]

We apply the method to 40 datasets of patients suffering from Mild Cognitive Impairment (MCI) and presenting WMHs of various loads.

2. METHODS

Figure 1 shows the general overview of our method.

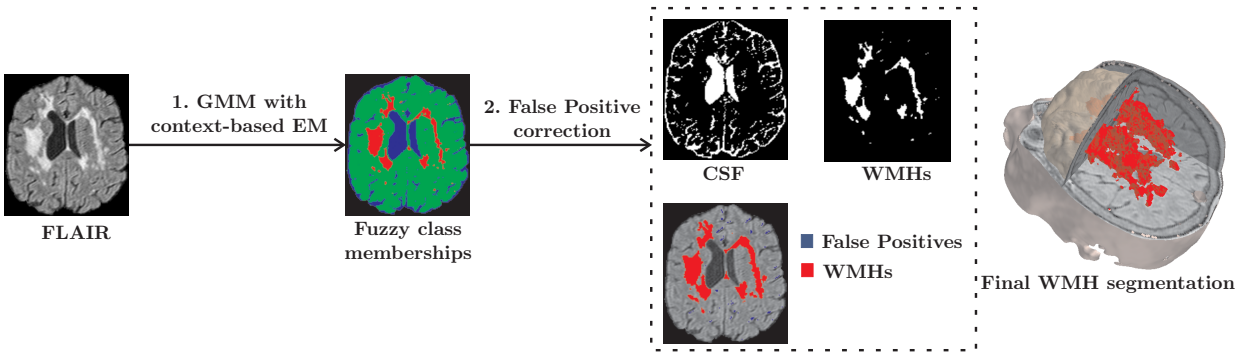


Figure 1. General overview of the segmentation method.

We start by computing the brain image histogram and fitting a Gaussian Mixture Model to it, using a modified Expectation-Maximization (EM) approach initialized with the normal EM method. Subsequently, we remove the false positives that are located at the interface between the cerebrospinal fluid and the gray matter, in the ventricles and in structures within the sagittal midplanes, such as the septum pellucidum. In the following sections we will describe these steps in detail.

2.1 Gaussian Mixture Model

Figure 2 shows the histograms of the FLAIR images of two patients. Two peaks can be easily distinguished: the lowest one, at lower intensities, corresponds to cerebrospinal fluid voxels; the highest peak refers to white and gray matter voxels. Additionally, in Figure 2b) a low and broad peak is present at the right-end tail of the histogram. This peak is especially prominent in patients with a large lesion load and corresponds to WMH intensities.

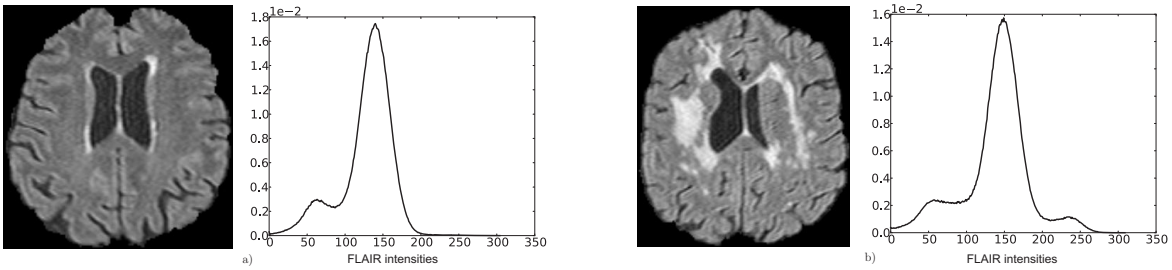


Figure 2. FLAIR image and respective histogram from a patient: a) with a low WMH load; b) with a high WMH load.

We assume that the data are generated by a Gaussian Mixture Model (GMM) and that each voxel belongs to one of three distinct classes—cerebrospinal fluid (CSF), white and gray matter (WM/GM), or white matter hyperintensity (WMH)—. The probability density function (pdf) of a gray-level x can then be described by:

$$p(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{k=1}^{3} \pi_k \mathcal{N}(x|\mu_k, \sigma_k) \tag{1}$$

with $k = 1, 2, 3$ respectively corresponding to the CSF, WM/GM and WMH classes. Each Gaussian component $\mathcal{N}$ is characterized by a mixing weight π_k, a mean value μ_k and a standard deviation σ_k. We use the Expectation-Maximization (EM) algorithm to find these parameters.

2.1.1 Traditional Expectation-Maximization

The EM algorithm is an iterative procedure that maximizes the log-likelihood of the parameters.[13,14] It alternates between two consecutive steps: the E-step and the M-step. In the E-step, the parameters at the current iteration are used to compute the log-likelihood. In the M-step, the computed log-likelihood is maximized to determine the new parameters.

Assuming that the data, $\mathbf{X} = (x_1, ..., x_N)$, are independent and identically distributed, the log-likelihood of the parameters given the data is defined as:

$$\ell(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{X}) = \log \prod_{n=1}^{N} p(x_n|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \log p(x_n|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \tag{2}$$

The M-step parameter estimates are derived by maximizing Eq. 2:

$$\begin{aligned} \mu_k^{(i+1)} &= \frac{1}{N} \sum_{n=1}^{N} x_n T_{k,n}^{(i)} \\ \sigma_k^{(i+1)} &= \sqrt{\frac{\sum_{n=1}^{N} \left(x_n - \mu_k^{(i+1)}\right)^2 T_{k,n}^{(i)}}{\sum_{n=1}^{N} T_{k,n}^{(i)}}} \\ \pi_k^{(i+1)} &= \frac{1}{N} \sum_{n=1}^{N} T_{k,n}^{(i)} \end{aligned} \tag{3}$$

where $T_{k,n}^{(i)}$ is determined at the E-step by:

$$T_{k,n}^{(i)} = \frac{\pi_k^{(i)} \mathcal{N}(x_n | \mu_k^{(i)}, \sigma_k^{(i)})}{p(x_n | \boldsymbol{\pi}^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{(i)})} \tag{4}$$

The initial parameters are computed from the histogram as follows: $\mu_{WM/GM}^{(0)}$ and $\mu_{CSF}^{(0)}$ correspond to the first and second highest peaks in the histogram, respectively; $\mu_{WMH}^{(0)}$ is taken as the average value of the maximum gray level present in the image and $\mu_{WM/GM}^{(0)}$; all standard deviations are initialized with the average of half the distance between the peaks; finally, the CSF and WM/GM weights are selected based on the ratio between $\mu_{WM/GM}^{(0)}$ and $\mu_{CSF}^{(0)}$; the initial WMH weight is set to 1%.

The algorithm has converged when the absolute normalized difference between the log-likelihood values at two consecutive iterations is lower than 0.1%.

Although it may be sufficient to obtain a first rough approximation of the voxels' statistical distributions, the traditional EM algorithm has the disadvantage of taking only intensity information into account. However, in an image, the location of the voxels can provide further information. In particular, we can expect adjacent voxels to belong to the same class.[15] Because the normal EM method described above does not consider such contextual information, we use a previously proposed[11] adaptation to the E-step. The difference between the normal and the adapted EM approaches is particularly significant in cases with low WMH loads, as we will show in Section 3.

2.1.2 Context-Sensitive Expectation-Maximization

In,[11] the authors introduced contextual information into the traditional EM method as follows. At each iteration, the posterior probability (Equation 4) is substituted by:

$$T_{k,n}^{(i)CC} = \frac{\pi_k^{(i)} C_{k,n}^{(i)} \mathcal{N}(x_n | \mu_k^{(i)}, \sigma_k^{(i)})}{p(x_n | \boldsymbol{\pi}^{(i)}, \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{(i)})}, \tag{5}$$

which incorporates a context-sensitive penalty term $C_{k,n}^{(i)}$. This term imposes that, at each iteration, the probability that a voxel belongs to class k depends not only on the voxel's intensity, but also on its neighbors' current class probabilities. We define the penalty term as follows:

$$C_{k,n}^{(i)} = NF\{I_k^{(i)}\}(x_n) \tag{6}$$

with $I_k^{(i)}$ being the membership image which, at each brain voxel x_n, represents the probability that the voxel belongs to class k. $NF\{\cdot\}$ represents the filter used to take the voxel's neighborhood into account. In,[11] the authors used, as a filter, the mean value within a local window, respectively. In this work, we use the $3 \times 3 \times 3$ neighborhood of each voxel.

We initialize the context-sensitive (CS-) EM method with the parameters resulting from applying the traditional EM method to the dataset. After convergence, we apply a threshold to the resulting WMH and CSF membership images. Instead of using Bayes' Maximum A Posteriori (MAP) rule, we loosen the constraint and define the threshold as 10^{-5}.

2.2 False Positive correction

After applying the threshold, we still obtain some false positives — voxels that are initially considered to be lesions but are in reality FLAIR artifacts. We apply a postprocessing step that consists of eliminating these voxels from the segmentation, based on the knowledge on the location of common FLAIR hyperintensities.

2.2.1 GM/CSF interface and Flow artifacts

A common location of false positives is in the interface between the CSF and the cortical gray matter. To eliminate these voxels from our initial segmentation, we use the CSF mask obtained after thresholding the CSF class membership image that results from the anisotropic-EM method described above. We perform binary dilation of this mask with a three-dimensional $3 \times 3 \times 3$ cubic structure with connectivity 1, and we stop after 3 iterations. We mask our first WMH segmentation obtained after applying the EM method with the dilated CSF mask.

Other hyperintense voxels, resulting from flow artifacts (located mainly in the ventricular system)[12] are also eliminated in this step by morphologically "closing the holes" in the dilated CSF mask.

Finally, and because the lesion voxels adjacent to the ventricles are also eliminated after this step, we perform binary propagation to the initial WMH segmentation in order to recover these wrongly eliminated voxels.

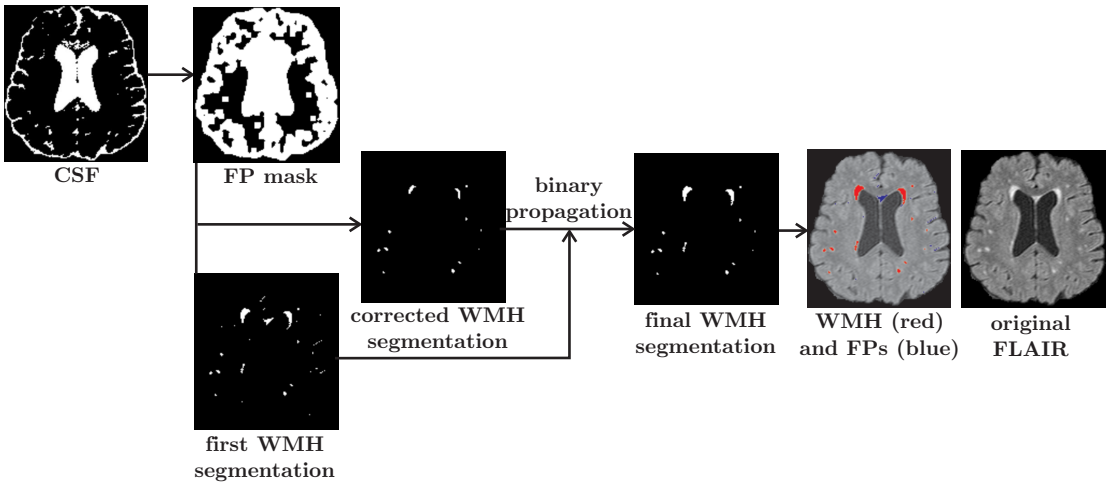


Figure 3. False Positive correction: the WMH segmentation obtained in Section 2.1.2 is first masked with the dilated CSF mask. The result is then propagated to the original segmentation to recover true lesion voxels. The result is shown as an overlay, with the lesions shown in red and the eliminated false positives in blue.

2.2.2 Septum Pellucidum

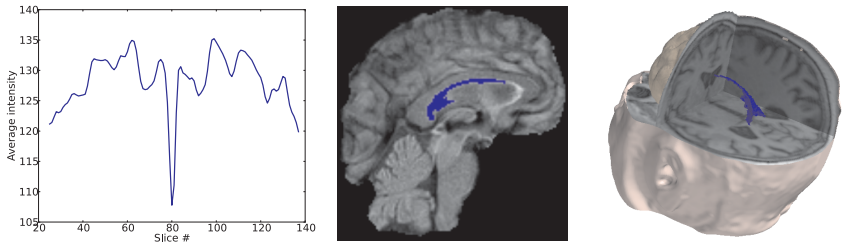
The septum pellucidum is a structure located between the left and the right ventricles that also shows high intensities in FLAIR.[?,12] To eliminate it from the segmentation, we first search for the sagittal midslice. This brain slice is characterized by having lower intensities than its immediately adjacent neighbors (Figure 4a)).

We then select the 16 middle slices and look for the lesions that were segmented within this volume. The septum pellucidum is the largest connected region. In some cases, other structures such as the falx cerebri and the corpus pineale show also as hyperintensities. We make then the decision of eliminating all lesions present within the midsagittal slices.

3. EXPERIMENTS AND RESULTS

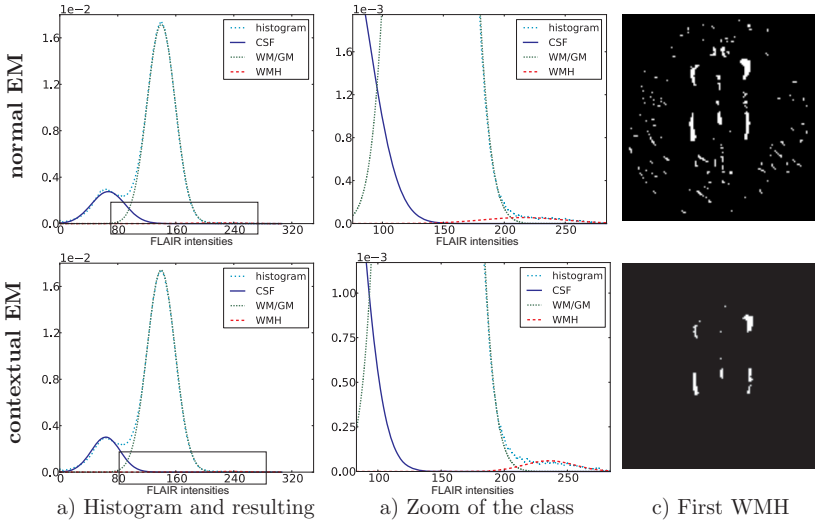
3.1 Data and preprocessing

Forty 3D isotropic FLAIR images are utilized in this study (Siemens Avanto, 1.5 T; TR=40ms; TE=5ms; voxel size = 1mm^3). The datasets were retrieved from a large database of a cognition study with MCI patients carried out at the University Hospital of Essen, Germany. We apply the following preprocessing steps:
- brain extraction using BET (Brain Extraction Tool, http://fsl.fmrib.ox.ac.uk/fsl/bet2/);[16]
- bias field correction using FAST (FMRIB's Automated Segmentation Tool, http://fsl.fmrib.ox.ac.uk/fsl/fast4/)[17]



a) b) c)

Figure 4. Elimination of the septum pellucidum voxels. a) average intensity of the sagittal slices, with the midslice presenting a global minimum; b) sagittal midslice with the septum pellucidum voxels (previously segmented as WMH) outlined in blue; c) 3D view of the eliminated septum pellucidum voxels.



a) Histogram and resulting GMM components. a) Zoom of the class overlapping regions. c) First WMH segmentation.

Figure 5. Example of the result of fitting a GMM to the FLAIR histogram: a) Histogram and the three GMM components; b) detail of the gray-level region where there is class overlap; c) first WMH segmentation.

3.2 Expectation-Maximization approaches

An example of applying the two EM methods (traditional and context-sensitive) to a dataset from a patient with a low WMH load is shown in Figure 5.

We observe that the WMH segmentation results (after applying a low threshold to the respective class probability image) are more accurate when the modified EM approach is used. In particular, in the first case the resulting WMH segmentation is corrupted with more false positives (voxels that are clearly part of the gray matter rather than the WMH class) than when the voxels' context is taken into account. Therefore, we choose this method to determine the GMM parameters and consequently the first WMH segmentation, which is afterwards subject to False Positive correction.

Furthermore, the class overlap is smaller in the CS-EM approach. The difference between the normal and modified approaches is more significant at the WM/GM-WMH overlap.

We determine the class overlapping (CO) area of the resulting GMMs for the 40 patient datasets (Figure 6):

$$\mathrm{CO} = \int \min\left(p_{\mathrm{CSF}}(x), p_{\mathrm{WM/GM}}(x)\right) \mathrm{d}x + \int \min\left(p_{\mathrm{WM/GM}}(x), p_{\mathrm{WMH}}(x)\right) \mathrm{d}x \tag{7}$$

The overlap is greater for the normal EM method, suggesting that taking neighborhood information into account will lead to a smaller classification error.

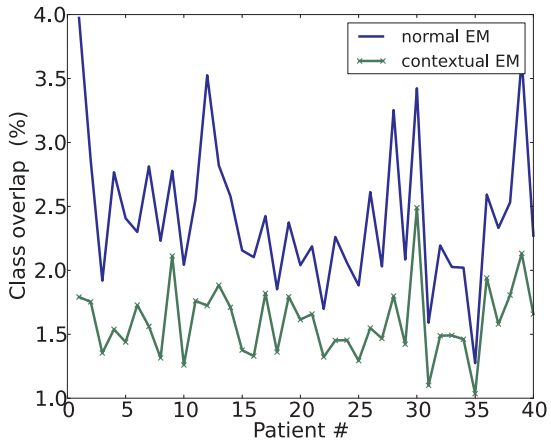


Figure 6. Class overlapping area for the 40 patient datasets.

3.3 False positive correction

As described in Section 1, most methods that use FLAIR images to determine a first segmentation of the WMHs comprise a subsequent False Positive correction step using the segmentation of an additional MRI modality, such as T1. As a comparison, we use FAST (FMRIB's Automated Segmentation Tool, http://fsl.fmrib.ox.ac.uk/fsl/-fast4/)[17] to segment the corresponding and previously co-registered T1 image. We apply the T1-based mask instead of our FP mask (cf. Figure 3) to remove the voxels that are located outside the white matter. We show in Figure 7 that such T1-based mask can be inadequate to detect the voxels at the GM/CSF interface (white arrows). In addition, we can also observe that the T1 segmentation result is significantly affected by the presence of large WMHs (blue arrows in the T1-based mask). This problem is, in this case, partially compensated by the binary propagation step we apply to the corrected WMH segmentation (see Figure 3).

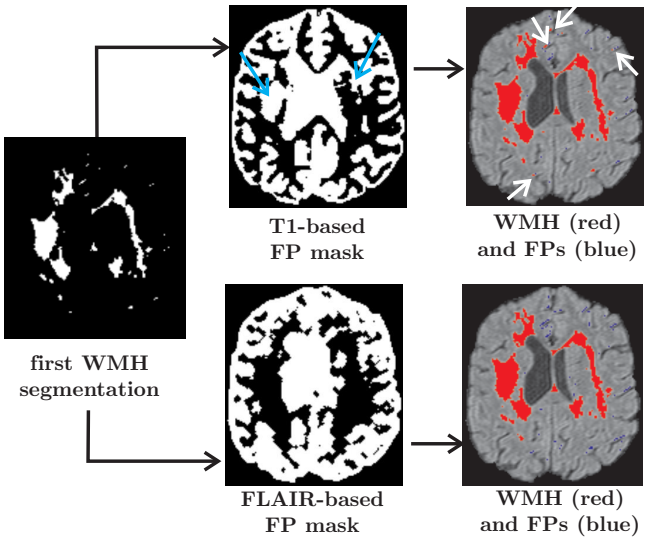


Figure 7. Comparison of the two masks used to remove the false positives.

3.4 Quantitative evaluation

To evaluate the method, we compare our results with the ground truth provided by a trained neuroradiologist. We use the following metrics for comparison: similarity index (SI), overlap fraction (OF) and extra fraction (EF):[4]

$$\mathrm{SI} = \frac{2 \times \#\mathrm{TP}}{\#\mathrm{AS} + \#\mathrm{GT}} \tag{8}$$

$$\mathrm{OF} = \frac{\#\mathrm{TP}}{\#\mathrm{GT}} \tag{9}$$

$$\mathrm{EF} = \frac{\#\mathrm{FP}}{\#\mathrm{GT}} \tag{10}$$

with TP and FP being the true and false positives, respectively, AS the automatic segmentation and GT the ground truth provided by the expert.

Our results (mean ± standard deviation) on six patient datasets are the following: SI= 0.73 ± 0.10, OF= 0.67 ± 0.18 and EF= 0.13 ± 0.07. All values are shown per patient in Figure 8.

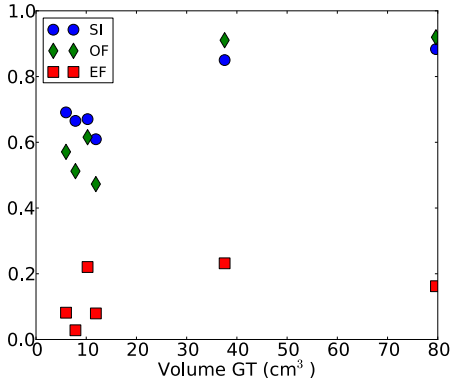


Figure 8. Performance measures (Similarity Index, Overlap Fraction and Extra Fraction) for six patients with various WMH loads. The ground truth corresponds to the manual segmentation performed by an experienced neuroradiologist.

According to Bartko *et al.*,[18] an SI value of 0.7 or higher indicates a very good agreement. The SI values obtained by previous WMH segmentation methods are shown in Table 1. Other studies have reported high correlation values between automatically and manually determined WMH volumes, but have not evaluated the quality of their segmentation.[19,20]

Table 1. Similarity Index (SI) values of automatic WMH segmentation methods available in the literature, for three lesion load categories.

method	**SI**			
	small (< 10 cm^3)	moderate ($10 - 30$ cm^3)	large (> 30 cm^3)	All
Admiraal-Behloul[6]	0.7	0.82	0.82	0.75
Anbeek[4]	0.5	0.75	0.85	0.80
Dyrby[21]	0.45	0.62	0.65	0.56
Samaille[22]	0.55	0.68	0.80	0.66
de Boer[5]	-	-	-	0.72

As is clear from Table 1, all methods perform worst for small lesion loads. In our evaluation, four out of the six patients for which a ground truth is provided have a lesion volume lower than 12cm^3. The average SI is, for these four patients, 0.66. Anbeek[4] and de Boer[5] reported an average Extra Fraction of 0.20 and 0.50, respectively. The overlap fractions were 0.81 and 0.79. It is worth noting that these values vary highly with the population. In particular, our preliminary evaluation considered a small sample of patients with low lesion loads, for which it is recognized that the existing segmentation methods are less accurate.

4. CONCLUSIONS

In this work, we present a method for automatically segmenting WMHs using only FLAIR images. It uses a Gaussian Mixture Model to approximate the image's gray-level histogram and initially segment the WMHs. Unlike existing approaches (to the best of our knowledge), our method requires no additional MRI modalities nor atlases, thereby avoiding the need for co-registrations and allowing for real-time analysis. Finally, our approach is totally unsupervised, meaning that we do not need any training set to find the method's parameters and that it can be applied as is to any FLAIR dataset.

We use a modified GMM-EM method as means of obtaining an initial unsupervised classification of the brain into three tissues. We observe that this approach improves the first WMH segmentation and reduces the class overlap, when compared to the traditional GMM-EM method.

We also show the limitations of considering a T1 segmentation for correcting the false positives. The T1-based false positive mask is not only incorrect due to the presence of a diffuse lesion but it also does not consider some of the false positives located at the CSF/GM interface.

A preliminary evaluation on six patient datasets shows that the method is at least comparable to other existing methods which require extra MRI modalities and, in some cases, a training set. The following step is to perform a more extensive validation with a larger number of patients. In addition, a reproducibility study would also be of interest. This consists of evaluating how stable the results are when the method is applied to FLAIR acquisitions over time.

ACKNOWLEDGMENTS

This work is part of the VIP-BrainNetworks project, which is funded by the department of Economic Affairs of the Netherlands and the provinces of Gelderland and Overijssel.

REFERENCES

[1] E. E. Smith, S. Egorova, D. Blacker, R. J. Killiany, A. Muzikansky, B. C. Dickerson, R. E. Tanzi, M. S. Albert, S. M. Greenberg, and C. R. G. Guttmann, "Magnetic Resonance Imaging white matter hyperintensities and brain volume in the prediction of mild cognitive impairment and dementia.," *Arch Neurol* **65**, pp. 94–100, Jan 2008.

[2] S. Debette and H. S. Markus, "The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis.," *BMJ* **341**, p. c3666, 2010.

[3] M. Yoshita, E. Fletcher, and C. DeCarli, "Current concepts of analysis of cerebral white matter hyperintensities on magnetic resonance imaging.," *Top Magn Reson Imaging* **16**, pp. 399–407, Dec 2005.

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

[5] R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen, "White matter lesion extension to automatic brain tissue segmentation on MRI.," *Neuroimage* **45**, pp. 1151–1161, May 2009.

[6] F. Admiraal-Behloul, D. M. J. van den Heuvel, H. Olofsen, M. J. P. van Osch, J. van der Grond, M. A. van Buchem, and J. H. C. Reiber, "Fully automatic segmentation of white matter hyperintensities in MR images of the elderly.," *Neuroimage* **28**, pp. 607–617, Nov 2005.

[7] C. R. Jack, P. C. O'Brien, D. W. Rettman, M. M. Shiung, Y. Xu, R. Muthupillai, A. Manduca, R. Avula, and B. J. Erickson, "FLAIR histogram segmentation for measurement of leukoaraiosis volume.," *J Magn Reson Imaging* **14**, pp. 668–676, Dec 2001.

[8] N. Hirono, H. Kitagaki, H. Kazui, M. Hashimoto, and E. Mori, "Impact of white matter changes on clinical manifestation of Alzheimer's disease: A quantitative study.," *Stroke* **31**, pp. 2182–2188, Sep 2000.

[9] K. H. Ong, D. Ramachandram, and R. Mandava, "Automated white matter lesion segmentation in MRI using Box-Whisker plot outlier detection," in *Medical Image Analysis and Understanding Conference (MIUA 2010). Univ. of Warwick, U.K.: British Machine Vision Association*, 2010.

[10] H. Caillol, W. Pieczynski, and A. Hillion, "Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation," *IEEE Transactions on Image Processing* **6**(3), pp. 425–440, 1997.

[11] H. Tang, J.-L. Dillenseger, X. D. Bao, and L. M. Luo, "A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model.," *Comput Med Imaging Graph* **33**, pp. 644–650, Dec 2009.

[12] M. Neema, Z. D. Guss, J. M. Stankiewicz, A. Arora, B. C. Healy, and R. Bakshi, "Normal findings on brain fluid-attenuated inversion recovery MR images at 3T.," *AJNR Am J Neuroradiol* **30**, pp. 911–916, May 2009.

[13] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)* **39, No. 1**, pp. 1–38, 1977.

[14] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Technical Report, Center for Biological and Computational Learning and the Artificial Intelligence, Laboratory of the Massachusetts Institute of Technology* , 1994.

[15] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," **16**(2), pp. 494–498, 2005.

[16] S. M. Smith, "Fast robust automated brain extraction.," *Human Brain Mapping,* **17**, pp. 143–155, November 2002.

[17] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.," *IEEE Trans Med Imaging* **20**, pp. 45–57, Jan 2001.

[18] J. J. Bartko, "Measurement and reliability: statistical thinking considerations.," *Schizophr Bull* **17**(3), pp. 483–489, 1991.

[19] M. Wu, C. Rosano, M. Butters, E. Whyte, M. Nable, R. Crooks, C. C. Meltzer, C. F. Reynolds, and H. J. Aizenstein, "A fully automated method for quantifying and localizing white matter hyperintensities on mr images.," *Psychiatry Res* **148**, pp. 133–142, Dec 2006.

[20] J. F. A. Jansen, M. C. G. Vlooswijk, H. M. Majoie, M. C. T. F. M. de Krom, A. P. Aldenkamp, P. A. M. Hofman, and W. H. Backes, "White matter lesions in patients with localization-related epilepsy.," *Invest Radiol* **43**, pp. 552–558, Aug 2008.

[21] T. B. Dyrby, E. Rostrup, W. F. Baar, E. C. van Straaten, F. Barkhof, H. Vrenken, S. Ropele, R. Schmidt, T. Erkinjuntti, L. Wahlund, L. Pantonih, D. Inzitari, O. B. Paulson, L. K. Hansen, and G. Waldemar, "Segmentation of age-related white matter changes in a clinical multi-center study," *Neuroimage* **41, Issue 2**, pp. 335–345, 2008.

[22] T. Samaille, O. Colliot, D. Dormont, and M. Chupin, "Automatic segmentation of age-related white matter changes on flair images: Method and multicentre validation," in *Proc. IEEE Int Biomedical Imaging: From Nano to Macro Symp*, pp. 2014–2017, 2011.