IMaGe: Iterative Multilevel Probabilistic Graphical Model for Detection and Segmentation of Multiple Sclerosis Lesions in Brain MRI

Nagesh Subbanna(✉), Doina Precup, Douglas Arnold, and Tal Arbel

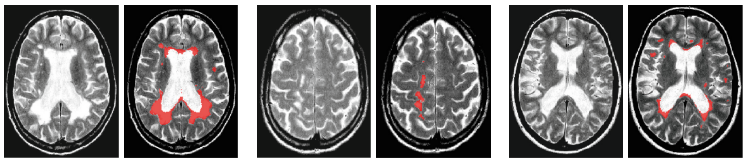
McGill University, Montreal, QC, Canada
nagesh@cim.mcgill.ca

Abstract. In this paper, we present IMaGe, a new, iterative two-stage probabilistic graphical model for detection and segmentation of Multiple Sclerosis (MS) lesions. Our model includes two levels of Markov Random Fields (MRFs). At the bottom level, a regular grid voxel-based MRF identifies potential lesion voxels, as well as other tissue classes, using local and neighbourhood intensities and class priors. Contiguous voxels of a particular tissue type are grouped into regions. A higher, non-lattice MRF is then constructed, in which each node corresponds to a region, and edges are defined based on neighbourhood relationships between regions. The goal of this MRF is to evaluate the probability of candidate lesions, based on group intensity, texture and neighbouring regions. The inferred information is then propagated to the voxel-level MRF. This process of iterative inference between the two levels repeats as long as desired. The iterations suppress false positives and refine lesion boundaries. The framework is trained on 660 MRI volumes of MS patients enrolled in clinical trials from 174 different centres, and tested on a separate multi-centre clinical trial data set with 535 MRI volumes. All data consists of T1, T2, PD and FLAIR contrasts. In comparison to other MRF methods, such as [5,9], and a traditional MRF, IMaGe is much more sensitive (with slightly better PPV). It outperforms its nearest competitor by around 20 % when detecting very small lesions (3–10 voxels). This is a significant result, as such lesions constitute around 40 % of the total number of lesions.

1 Introduction

Multiple Sclerosis (MS) is an inflammatory, demyelinating disease of the Central Nervous System (CNS)[1]. Establishing patient lesion load, through counts and volume measurements, provides important measures of disease progression and drug efficacy [1]. Current clinical practice involves manual labelling of MS lesions by experts, a task that is time-consuming, expensive and subject to inter- and intra-rater variability. Automatic MS lesion detection and segmentation are particularly challenging tasks, as lesions display wide variability in terms of shape, size, appearance, and location throughout the white matter (WM) (and possibly grey matter (GM)) of the brain. Normalization of different field strengths,

S. Ourselin et al. (Eds.): IPMI 2015, LNCS 9123, pp. 514–526, 2015.
DOI: 10.1007/978-3-319-19992-4_40



(a) (b) (c)

Fig. 1. (left) Brain images of MS patients with (right) lesions segmented by experts. We show (a) a volume with heavy peri-ventricular lesion load, (b) volume with supra-ventricular lesions, and (c) volume with juxta-cortical lesions. The different shapes, sizes, positions and intensities make lesion segmentation challenging.

acquisition sequences and different scanner types yields subtle differences in the appearances and absolute intensities of the lesions. These variabilities are further dependent on disease stage [2]. Figure 1 shows some lesion examples.

Although various automatic approaches have been developed for the segmentation of MS lesions, their adoption into real clinical practice has been limited. This is primarily due to strong simplifying assumptions, which restrict the robustness to variability displayed in large clinical datasets [2]. For example, techniques that define lesions as outliers of healthy class distributions [3,7] risk missing a large set of subtle lesions located throughout the brain, for which these assumptions are violated. Techniques based on topological features and template matching (e.g. [10]) presume that lesion shape can easily be modelled by a set of candidate lesion shapes, which is difficult to achieve in practice. Finally, most techniques are tested on synthetic lesion data, or on the limited MICCAI Challenge dataset [4,7], neither of which reflects the variability present in large clinical datasets.

Various MRF techniques have been proposed for pathology segmentation [3,11] but typically only class relationships are reflected in the priors (e.g. Ising models). However, traditional MRFs tend to smooth class labels excessively, and hence may lead to the removal of smaller MS lesions. To alleviate this problem, a modified MRF [9] has been proposed, which models both voxel intensities and intensity differences in the cliques. However, a local model of a voxel and its neighbourhood is insufficient for this task, and a more global view of context is necessary for accurate classification. Recently, several different, multi-level graphical models have been introduced for segmentation of pathologies [12,13]. However, none have been designed for the unique context of MS lesion segmentation.

In this work, we present IMaGe, a new, iterative, two-stage, probabilistic, graphical model for the detection and segmentation of MS lesions in multi-channel MRI. The goal of this architecture is to take advantage of both local information, as well as global context, in order to remove false assertions and refine boundaries. The lower level consists of nodes associated with each voxel, and models the intensities of healthy tissues and lesions, as in [9]. It produces a set of lesion candidates, passed to the higher level, whose goal is to determine which of these candidates are indeed lesions, by looking at a larger context

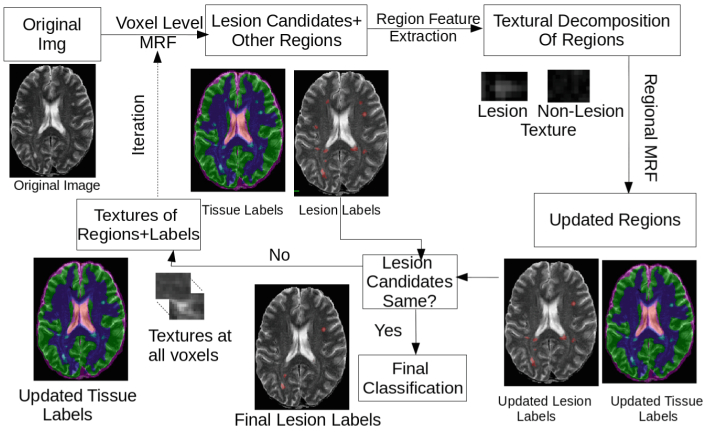


Fig. 2. Flowchart of the proposed approach. The multimodal MRI are processed by a 2 stage iterative MRF. At the bottom level, a voxel based MRF identifies potential lesion voxels, as well as other tissue classes. Contiguous voxels of a particular tissue type are grouped into regions. A higher, non-lattice based MRF is then constructed, in which each node corresponds to a region, and edges are defined based on neighbourhood relationships between regions. Regional features such as textures are extracted from each region. The goal of this MRF is to evaluate the probability of candidate lesions, based on group intensity, texture, and neighbouring regions. If the resulting labels have changed at this stage, the labels and textures are then propogated back down to the voxel-level MRF. This process of iterative inference continues until convergence.

around each region. The higher level consists of a non-lattice based MRF (whose structure is not a regular, uniform grid). Each node corresponds to a contiguous region of voxels currently labelled with the same class, and nodes are connected if the corresponding regions have at least one voxel adjacent to the other region. Each region is modelled using both intensity and textural features, and the relations between different regions are also considered. The class label computed based on this information, as well as the texture information, is then passed to the lower level. Inference alternates between the two levels, until a consensus is reached about the lesion candidates, or a given number of iterations has elapsed. Figure 2 illustrates this process.

IMaGe is trained on 660 MS patient MRI volumes from a clinical trial involving 174 centres. In order to test the robustness of the framework to different trial data, testing was performed on a different clinical trial data set of 535 MRI MS patient volumes from 128 centres. Both datasets include T1, T2, PD and FLAIR contrasts. We compared IMaGe to a recent MRF technique based on FLAIR [5], a traditional MRF, and the result of the voxel-based MRF [9]. IMaGe outperformed the competitors based on sensitivity and Positive Predictive Value, particularly in the context of smaller lesion detection (i.e. in the 3–10 voxel range), comprising around 40 % of the clinical trial dataset. These small lesions often constitute a very small portion of the total volume of the lesions, but it is important to detect them in clinical settings since the activity of the disease is measured by the number of lesions, both large and small.

2 Methodology

Given a multichannel MRI volume, the task is to identify the voxels corresponding to MS lesions. The proposed hierarchical MRF technique identifies potential lesion candidates at the lower level, and the validity of the candidates is inferred at the higher level, where the lesion boundaries are also refined. We describe in detail the framework in Fig. 2.

2.1 Voxel-Level MRF

Traditional voxel level MRFs [8,11] are typically lattice-based and infer the class of a voxel using observations local to that voxel (such as intensity) and the classes of the neighbouring voxels. However, this process tends to smooth over small patches of voxels whose class differs from their neighbours. This is problematic for MS lesions, which tend to be very small (3–10 voxels). To alleviate this problem, we employ a modified MRF at the voxel level, as in [9], which uses at each voxel the contrast of the local observation with those of the neighbours. Specifically, at each node we consider the observation to include the intensity difference between the corresponding voxel and its neighbours.

The MRF is structured as a regular lattice, with a node for each voxel and connectivity as depicted in Fig. 3. Let Q be the number of voxels in the volume, and $\mathbf{I}_i$ denote the vector of multimodal intensities at voxel i and N_i denote i's neighbourhood (Fig. 3). The goal is to infer the class label C_i, corresponding to voxel i's tissue type (lesion, white matter (WM), grey matter (GM), cerebro-spinal fluid (CSF) or partial volume[1] (PV)). Let $\mathbf{I}_{N_i}$ denote the vector of intensities of the voxels in N_i and $\mathbf{C}_{N_i}$ denote the vector of class labels for these voxels. Let $\mathbf{C} = (C_0, C_1, \ldots, C_{Q-1})$ be the collection of all class labels for the corresponding voxels in the volume, and $\mathbf{I} = (\mathbf{I}_0, \mathbf{I}_1, \ldots, \mathbf{I}_{Q-1})$ be the corresponding intensities. The goal is to compute $P(\mathbf{C} \mid \mathbf{I})$, which in an MRF is given by:

$$P(\mathbf{C} \mid \mathbf{I}) \propto e^{-U(\mathbf{C}|\mathbf{I})}, \tag{1}$$

where $U(\mathbf{C} \mid \mathbf{I})$ is the energy of the configuration. Following [9], the posterior probability of class C_i is given by:

$$\begin{aligned} P(C_i \mid \mathbf{I}_i, \mathbf{I}_{N_i}) &= \sum_{\mathbf{C}_{N_i}} P(C_i, \mathbf{C}_{N_i} \mid \mathbf{I}_i, \mathbf{I}_{N_i}) = \sum_{\mathbf{C}_{N_i}} \frac{P(\mathbf{I}_{N_i}, \mathbf{I}_i \mid C_i, \mathbf{C}_{N_i}) P(C_i, \mathbf{C}_{N_i})}{P(\mathbf{I}_i, \mathbf{I}_{N_i})} \\ &\propto \sum_{\mathbf{C}_{N_i}} P(\mathbf{I}_{N_i} \mid \mathbf{I}_i, C_i, \mathbf{C}_{N_i}) P(\mathbf{I}_i \mid \mathbf{C}_{N_i}, C_i) P(\mathbf{C}_{N_i} \mid C_i) P(C_i). \end{aligned} \tag{2}$$

where the sums marginalize over all combinations of class labels that can be assigned to the neighbourhood voxels. In Eq. (2), $P(C_i)$ is the prior probability

[1] Partial volume denotes the class ascribed to voxels which are a mix of GM and CSF. This class is created in order to reduce the number of false negatives at the edges of the ventricles.

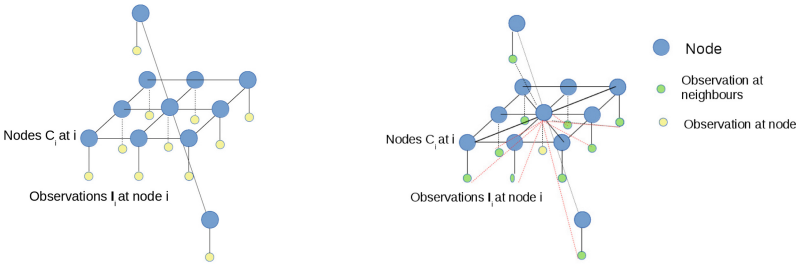


Fig. 3. (a) The regular MRF model and (b) the MRF model used in [9]. Note that the neighbouring intensities are considered in the computation of the label at every node i in this model.

at voxel i, $P(\mathbf{C}_{N_i} \mid C_i)$ represents the probability of class transitions in the neighbourhood, $P(\mathbf{I}_i|C_i)$ is the likelihood of $\mathbf{I}_i$ given class C_i and $P(\mathbf{I}_{N_i}|\mathbf{C}_{N_i}, C_i, \mathbf{I}_i)$ models the likelihood of the intensity of the neighbouring voxels, given their classes and the information at voxel i.

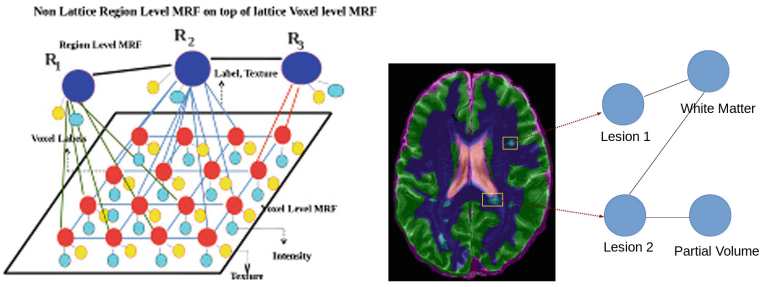
Let $\mathbf{\Delta I}_{N_i}$ be defined as the intensity difference between $\mathbf{I}_i$ and $\mathbf{I}_{N_i}$ in the clique (see Sect. 3.1 for details on computing intensity differences). Since $\mathbf{I}_{N_i}$ can be computed deterministically from $\mathbf{\Delta I}_{N_i}$ and $\mathbf{I}_i$, in Eq. 2, we replace $P(\mathbf{I}_{N_i} \mid \mathbf{I}_i, C_i, \mathbf{C}_{N_i})$ with $P(\Delta\mathbf{I}_{\mathrm{N_i}} \mid \mathbf{I}_{\mathrm{i}}, \mathrm{C_i}, \mathbf{C}_{\mathrm{N_i}})$. We can also assume that the difference in intensity between a voxel and its neighbours depends on the classes in the neighbourhood, but not on the absolute intensity of the voxel itself, and hence we can replace $P(\mathbf{\Delta I}_{N_i} \mid \mathbf{I}_i, \mathbf{C}_{N_i}, C_i)$ with $P(\mathbf{\Delta I}_{N_i} \mid \mathbf{C}_{N_i}, C_i)$. Furthermore, we assume that the intensity of a voxel is conditionally independent of the neighbourhood classes, given the class label at the voxel. Hence, we can replace $P(\mathbf{I}_i \mid C_i, \mathbf{C}_{N_i})$ with $P(\mathbf{I}_i \mid C_i)$. With these assumptions, Eq. (2) simplifies to:

$$P(C_i \mid \mathbf{I}_i, \mathbf{I}_{N_i}) \approx P(\mathbf{I}_i \mid C_i)P(C_i) \sum_{\mathbf{C}_{N_i}} P(\Delta\mathbf{I}_{N_i} \mid C_i, \mathbf{C}_{N_i})P(\mathbf{C}_{N_i} \mid C_i). \quad (3)$$

Using Eqs. 1 and 3, the Gibbs energy function is given by: $P(\mathbf{C} \mid \mathbf{I}) \propto \prod_{i=0}^{Q-1} e^{-U(C_i|\mathbf{I}_i, \mathbf{I}_{N_i})}$. To infer the class labels, the total energy must be minimised. The energy at each voxel i can be expressed as:

$$\begin{aligned} U(C_i|\mathbf{I}_i, \mathbf{I}_{N_i}) = &- \log P(\mathbf{I}_i|C_i) - \log P(C_i) \\ &- \sum_{k \in \mathrm{Cliques}(N_i)} \sum_{j \in k} (\log P(\Delta\mathbf{I}_{i,j}|C_i, \mathbf{C}_j) - \alpha m(\mathbf{C}_j, C_i)), \end{aligned} \quad (4)$$

where k indexes over all the possible cliques in N_i and j indexes over all the possible elements of clique k, $m(\mathbf{C}_j, C_i)$ is the potential associated with the relationship between C_i and the vector of classes in the jth clique, $\mathbf{C}_j$, and α is a weighting parameter used to handle the differences between inter-slice and intra-slice distances.



(a) (b)

Fig. 4. In (a), the schematic diagram of the hierarchical MRF and the features used at each level are shown. Red nodes represent the classes at each voxel. Light blue nodes are the derived textures, and yellow are the intensities. The voxels with the same labels are merged into regions and moved into the higher level MRF. In (b), all the different classes of a volume in one slice are marked. All the links of the two lesions (highlighted with yellow rectangles) are shown. Lesion 1 has only one neighbour, white matter (seen in blue in the MRI). Lesion 2 has two neighbours, partial volume (seen in magenta on the MRI) and white matter (Color figure online).

In order to estimate the lesion candidates, we run inference in the MRF until convergence. The parameters of the voxel-based MRF have been trained so as to obtain high sensitivity in lesion detection, at the expense of a higher false positive rate, as pruning of the false positives will occur at the region level. All voxels for which the lesion class probability is higher than a certain threshold are labelled as lesions. For all the other voxels, the label is the MAP class. The value of the lesion threshold is also chosen to ensure high sensitivity.

2.2 Region-Level MRFs

The goal of the region level MRF is to adjust the class labels by taking larger scale contextual information into account. The region level MRF is non-lattice-based and is constructed adaptively based on the labels assigned by the voxel-level MRF (depicted in Fig. 4). A region is defined as a set of voxels that have been assigned the same class by the voxel-based MRF and that are connected in 3D. Each region is represented by a node in the MRF. The nodes corresponding to two regions r and q have an edge between them if there are at least two voxels $i \in r, j \in q$ which are adjacent. Due to this process, regions correspond to different numbers of voxels and have different shapes. The region based MRF aims to assign classes to the voxels in each region based on the intensities of the voxels in the region and in the neighbouring regions, as well as based on texture information obtained from the voxels in each region. A region's class will be inferred as the class of its corresponding MRF voxel.

Let $\mathbf{T}_r$ denote the vector of texture features at node r. Let the intensity of the region be $\mathbf{I}_r$ and intensity difference between the region r and its neighbouring

regions be $\mathbf{\Delta I}_{N_r}$. The probability of C_r at node r is given by

$$P(C_r \mid \mathbf{T}_r, \mathbf{I}_r, \mathbf{\Delta I}_{N_r}) = \sum_{\mathbf{C}_{N_r}} P(C_r, \mathbf{C}_{N_r} \mid \mathbf{I}_r, \mathbf{T}_r, \mathbf{\Delta I}_{N_r})$$
$$\propto \sum_{\mathbf{C}_{N_r}} P(\mathbf{\Delta I}_{N_r} \mid \mathbf{T}_r, \mathbf{I}_r, C_r, \mathbf{C}_{N_r}) P(\mathbf{T}_r \mid \mathbf{I}_r, C_r, \mathbf{C}_{N_r})$$
$$P(\mathbf{I}_r \mid C_r, \mathbf{C}_{N_r}) P(\mathbf{C}_{N_r} \mid C_r) P(C_r) \tag{5}$$

Following Bayesian reasoning, and assuming the texture features are also conditionally independent of all other information given the class label for any node r, we have:

$$P(C_r \mid \mathbf{I}_r, \mathbf{T}_r, \mathbf{I}_{N_r}) \approx P(\mathbf{T}_r|C_r) P(\mathbf{I}_r|C_r) P(C_r) \sum_{\mathbf{C}_{N_r}} P(\mathbf{\Delta I}_{N_r} \mid C_r, \mathbf{C}_{N_r}) P(\mathbf{C}_{N_r}| C_r) \tag{6}$$

The MRF energy equation, corresponding to Eq. (6) is given by $P(\mathbf{C} \mid \mathbf{I}, \mathbf{T}) \propto \prod_{r=0}^{R-1} e^{-U(\mathbf{C_r}|\mathbf{I}_r, \mathbf{I}_{N_r} \mathbf{T}_r)}$, where R is the total number of regions in the volume. By reasoning similar to Eq. (3), and assuming that we consider cliques of size 2 at maximum, we obtain:

$$U(C_r \mid \mathbf{I}_r, \mathbf{I}_{N_i} \mathbf{T}_r) = -\log P(\mathbf{I}_r \mid C_r) - \log P(\mathbf{T}_r \mid C_r) - \log(C_r)$$
$$- \sum_{s \in \text{Cliques}(N_r)} (\log P(\Delta \mathbf{I}_{r,s} \mid C_r, C_s) - \alpha m(C_s, C_r)), \tag{7}$$

where s indexes over all possible cliques of size 2. The inference process in the MRF minimizes the energy in this model, and re-labels all the regions. The result of this process is typically an elimination of false positives from the voxel-based MRF.

2.3 Iterations

After the region-level inference, the result is compared to the low-level MRF. If there are discrepancies in the lesion voxel labelling, further iterations are needed. In this case, the information computed by the region-level MRF has to be communicated to the voxel-level. Two types of information are transmitted: the class labels that are inferred for each region, and the texture information that was computed. Each voxel $i \in r$ will initially have $C_i = C_r$ meaning that all the labels in the region are moved back down to their corresponding voxels to initiate the process. Similarly, there is a voxel- associated texture $\mathbf{T}_i$ (whose computation is described in Sect. 2.2). At this stage, further refinement of boundaries is performed.

The voxel-level MRF is exactly as described in Sect. 2.1, except it also uses the texture information $\mathbf{T}_i$, Using the same assumptions as in Eq. (3) and assuming that $\mathbf{T}_i$ is conditionally dependent only on C_i, we get

$$P(C_i \mid \mathbf{I}_i, \mathbf{T}_i, \mathbf{I}_{N_i}) = \sum_{\mathbf{C}_{N_i}} P(\mathbf{\Delta I}_{N_i} \mid C_i, \mathbf{C}_{N_i}) P(\mathbf{T}_i \mid C_i) P(\mathbf{I}_i \mid C_i) P(\mathbf{C}_{N_i} \mid C_i) P(C_i) \tag{8}$$

and the corresponding energy at voxel i given a particular configuration of classes is given by:

$$U(C_i|\mathbf{I}_i, \mathbf{I}_{N_i}) = -\log P(C_i) - \log P(\mathbf{I}_i|C_i) - \log P(\mathbf{T}_i|C_i) - \sum_{k \in \text{Cliques}_{N_i}} \sum_{j \in k} (\log P(\Delta\mathbf{I}_{i,j}|C_i, \mathbf{C}_j) - \alpha m(\mathbf{C}_j, C_i)), \quad (9)$$

where, again, k indexes over all possible cliques in N_i and j indexes over all the possible elements of clique k. Once the MRF inference has been run to convergence at the voxel level, inferring new class labels C_i, and the process is iterated.

3 Experiments and Results

We used two proprietary clinical trial data sets of patients with Relapsing Remitting MS for training and testing. The training data set consisted of 660 MS patients from 174 centres. The test data set contained 535 MS patients' MRI volumes from 128 centres. The test data and training data come from entirely different clinical trials. Both data sets included semi-manual lesion segmentations by trained experts following a strict protocol. All volumes consisted of four contrasts: T1, T2, FLAIR (T2w fluid attenuation inversion recovery), and PD (proton density). All volumes underwent bias-field inhomogeneity correction using N3 [14], intra-subject registration of multispectral volumes [15], extraction of non-brain regions from the MRI (brain parenchyma) [16], and intensity range normalization [17]. Intra-subject registration of volumes involves registering contrasts to stereotactic space before any further processing is done.

In our experiments, we compare our results against 3 different methods. The first [5] is an MRF which categorizes lesions based on outliers in $T1$ and labels them in $FLAIR$. The second is a traditional MRF, which encapsulates class relations using priors and similar classes in the neighbourhood. The last is the modified MRF from [9], where classification is based on modelling both intensities and intensity differences, as well as class similarities in the neighbourhood.

3.1 Implementation of IMaGe

Computation of Intensity Difference $\Delta\mathbf{I}_{N_i}$: For a 2 voxel clique, we have the node and one of its neighbours. Assume that the intensity of the node is $\mathbf{I}_1$ and that of the neighbour is $\mathbf{I}_2$. The $\Delta\mathbf{I}_{1,2}$ is computed simply by $[\mathbf{I}_1 - \mathbf{I}_2]$. For a 3 voxel clique, we have the node and two of its neighbours. Assuming that the intensity of the node is $\mathbf{I}_1$ and that of the neighbours are $\mathbf{I}_2$ and $\mathbf{I}_3$, we have $\mathbf{I}_{1,2,3}$ given by

$$\Delta\mathbf{I}_{1,2,3} = \begin{bmatrix} \mathbf{I}_1 - \mathbf{I}_2 \\ \mathbf{I}_1 - \mathbf{I}_3 \end{bmatrix}. \quad (10)$$

The computation of intensity differences for 4 and 5 node cliques follows similarly.

Voxel Level MRF: There are two stages in the technique. In the training stage, all the models are learned from the expert labelled volumes.

Training: The intensities and intensity differences of both the healthy tissues and the lesions are modelled using multivariate Gaussian Mixture Models, with two components per contrast since all tissue classes are either unimodal or bimodal [8]. The neighbourhood contains the 8 in plane neighbours and the two corresponding voxels in the slices above and below. We generate the intensity difference models for all the cliques possible ($2, 3, 4$ and 5 voxel cliques for the neighbourhood chosen). The method models combinations of labels in cliques, not the order (position) in which they occur. The spatial transition probabilities between the different classes are learned from the frequency of co-occurrence in the labelled training volumes.

Classification: In the classification stage, the prior probabilities required for healthy tissues $P(C_i)$ are taken from an ICBM152 healthy brain atlas registered non-linearly to the volumes. The lesion probabilities are taken from a proprietary lesion atlas, built manually from over 3000 MS patients from different clinical trials (separate from the training and testing data sets), non-linearly registered to the volume being classified. If the lesion posterior probability $P(C_i \mid \mathbf{I}_i, \mathbf{I}_{N_i})$ of a voxel i is above a particular threshold, it is considered a potential lesion candidate. Iterated Conditional Modes are used to minimze the total energy at this stage.

Regional MRF: Here, $\mathbf{I}_r$ is computed as the mean intensity of all voxels associated with region r. $\mathbf{\Delta I}_{N_r}$ is computed based on the 'border' voxels of the two regions, as:

$$\mathbf{\Delta I}_{N_r} = \frac{1}{S_r} \sum_{i \in r} \sum_{j \in s, j \text{adj} i} (\mathbf{I}_j - \mathbf{I}_i) \tag{11}$$

where S_r is the number of edges connecting voxels i in r with voxels j in s, and denotes adjacency of i and j.

Texture can be modelled in many ways, and the approach proposed can accommodate any choice of texture features. Here, we use multi-window Gabor textures. Multi-window Gabor coefficients are very flexible, and can be customized to problems since the analysis and synthesis windows are de-coupled [6]. To compute $\mathbf{T}_r$, at every voxel i in the region r, we choose the first Gabor window at a particular orientation centred at i that completely crosses the boundary of the region r before the Gabor coefficient value falls to a thousandth of the peak. This gives the value of $\mathbf{T}_r$ at the voxel i. The mean texture $\mathbf{T}_r$ is obtained for each Gabor window for the class using the textures at all the individual voxels in the region. Gabor texture computation for every region has been described in detail in [18] for the case of brain tumours.

Training and classification: Gabor textures $\mathbf{T}_r$, the regional intensities $\mathbf{I}_r$ and the intensity differences $\mathbf{\Delta I}_r$ are all modelled as Gaussian mixtures. Regional class transition probabilities are computed from the frequency of co-occurrence of the classes. During classification, we consider only cliques of size 2. We start with labels obtained by the voxel-based MRF and apply ICM for the inference.

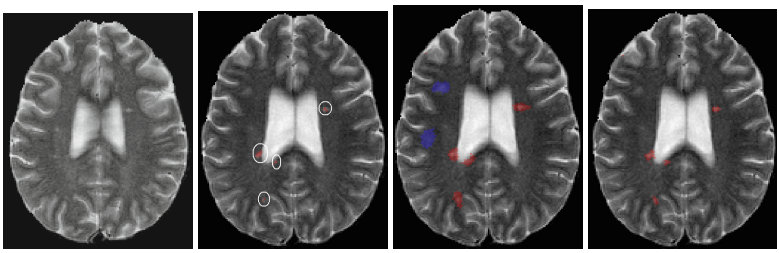
Iterations. We compare the lesions estimated from the voxel level MRF and the region level MRF. If the difference between the two is less than a threshold (in number of voxels), then iterations are stopped. Otherwise, the iterations proceed in the same fashion, using the currently estimated parameters. The textures for each voxel obtained at the region level MRF and the labels of each region are passed down to the voxel level MRF and the process iterates as described in Eq. (9). In some cases, it is possible that the iterations will oscillate between the solution of voxel-level MRF and the regional level MRF. To account for such cases, we set a maximum number of iterations. The solution obtained at the end of the number of maximum iterations is taken to be the final answer. The maximum time taken by the technique on a Dell Optiplex 980 I7 machine is 40 minutes per volume.

3.2 Qualitative Results

Juxta-cortical lesions, often resembling grey matter in intensity, are hard to detect with just voxel level information. When lesion candidates are chosen (Fig. 5c), not only are all lesions detected by experts in Fig. 5b chosen as lesion candidates, but some additional false positive lesions are also chosen (Fig. 5c). Final results by IMaGe (Fig. 5d) show false positive regions removed iteratively, and true positive boundaries refined.

3.3 Quantitative Results

The clinical goal of this work is to detect all lesions accurately, rather than attain high precision in lesion boundaries, particularly given that there is little agreement regarding boundaries among expert raters. In clinical trials, detection accuracy is crucial, as lesion counts are used to assess if a treatment is working, and even one new lesion can indicate that the disease is active. Hence,



(a) (b) (c) (d)

Fig. 5. Qualitative results on T2 slices of patient (a) Original unlabelled volume (b) Expert classification (lesions circled), (c) Lesion Candidates from first level of IMaGe, and (d) Final classification by IMaGe. In all images, red=true positive, blue=false positive lesions (Color figure online).

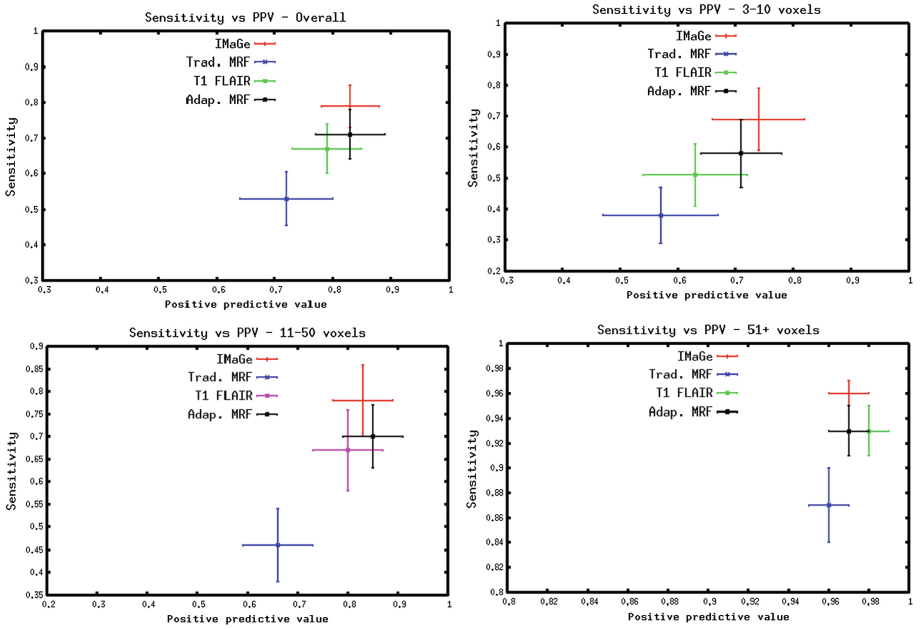


Fig. 6. Sensitivity against PPV for the overall data set, and broken down according to lesion size ranges in voxels. Lesion breakdown: Small lesions (3–10 voxels) 39 % (12737 lesions), moderate lesions (11–50 voxels) 36 % (11767 lesions), and large lesions (51+ voxels) 25 % (8124 lesions). IMaGe outperforms competing approaches in terms of sensitivity, particularly for small lesions of size 3–10 voxels, and shows some sensitivity and PPV improvements for all lesion sizes. In the figures, Trad. MRF refers to a standard Potts model MRF, Adap. MRF refers to the technique used in [9], and T1-FLAIR refers to the method used in [5]

metrics that consider false positives and false negatives with respect to the number of lesions in the volume are the best way to measure the algorithm's efficacy. We present quantitative results which include sensitivity and positive predictive value (PPV) at the lesion level. We use positive predictive value, since the number of true negative regions is very small and consequently, the relative number of false positives is better reflected by using PPV instead of specificity. According to the clinical protocol, a lesion is a set of 3 or more voxels, contiguous in 3D, that are labelled as lesion by the expert. An expert lesion is counted as a true positive (TP) when 3 or more of its voxels overlap with the algorithm-detected lesion. Otherwise, it is a false negative (FN). A false positive (FP) lesion has fewer that than 3 voxels overlapping with the expert lesion. Sensitivity is defined as TP/(TP + FN) and positive predictive value (PPV) is defined as TP/(TP+FP).

In Fig. 6, there is an overall improvement in sensitivity by 11.26 % compared to the next closest competitor, with the largest gain obtained for smaller lesion detection. The sensitivity for small lesions (3–10 voxels) is 18.96 % higher than for the closest competitor, and for moderate lesions (11–50) voxels, it is 11.42 %

higher. For large lesions, (51+ voxels), all methods have sensitivities greater than 0.9.

In our test data set, small lesions (between 3–10 voxels) accounted for 39 % of total lesions. Hence, overlap metrics such as Dice statistics, which do not properly penalize missing smaller lesions [2], are not as appropriate as lesion-based metrics. However, since most results in MS lesion segmentation literature are in terms of Dice similarity values, we also provide the equivalent Dice similarity coefficient values for all 4 techniques. The Dice similarity for Standard MRFs is 0.48, for Adaptive MRFs is 0.67, for T1-FLAIR is 0.63 and for IMaGe is 0.69. IMaGe results in a slightly higher Dice value, but its real advantage is in its ability to provide accurate detection of smaller lesions.

4 Conclusion and Future Work

In this paper, we introduced IMaGe, a new, iterative, multi-level probabilistic graphical model for the detection and segmentation of MS lesions. The method leverages both local and global, contextual information. Training and experimental validation are based on two, large, clinical trial datasets of relapsing remitting MS patients, in which multi-channel MRI data was recorded. The proposed method outperforms other state-of-the-art graphical models, providing much higher sensitivity for the detection of small lesions, with a slight improvement in PPV. In the future, we plan to explore more expressive models for the different classes, as well as alternative optimization techniques, in order to improve the computational speed and further improve the method's accuracy.

References

1. MacDonald, I.W., et al.: Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. Ann. Neurol. **50**(1), 121–127 (2001)
2. Garcia-Lorenzo, D., et al.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med. Image Anal. **17**(1), 1–18 (2013)
3. von Leemput, K., et al.: Automated segmentationo of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imag. **20**(8), 677–688 (2001)
4. Souplet, J., et al.: An automatic segmentation of T2-FLAIR Multiple Sclerosis lesions. In: Midas Jounal (2008)
5. Schmidt, P., et al.: An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. NeuroImage **59**, 3774–3783 (2012)
6. Subbanna, N., et al.: Existence conditions for non canonical multiwindow gabor functions. Trans. Signal Process. **55**(11), 5112–5117 (2007)
7. Weiss, N., Rueckert, D., Rao, A.: Multiple sclerosis lesion segmentation using dictionary learning and sparse coding. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 735–742. Springer, Heidelberg (2013)
8. Harmouche, R., et al.: Bayesian MS Lesion classification modelling regional and local spatial information. In: Proceedings of ICPR 2006, pp. 984–987 (2006)

9. Subbanna, N., et al.: Adapted MRF Segmentation of MS Lesions uisng Local Contextual Information. In: Proceedings of MIUA 2011, pp. 445–450 (2011)
10. Wu, Y., et al.: Automated segmentation of multiple sclerosis subtypes with multichannel MRI. NeuroImage **32**, 1205–1215 (2006)
11. Khayati, R., et al.: Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov Random field model. Comput. Bio. Med. **38**, 379–390 (2008)
12. Karimaghaloo, Z., Rivaz, H., Arnold, D.L., Collins, D.L., Arbel, T.: Adaptive voxel, texture and temporal conditional random field for detection of gad-enhancing multiple sclerosis lesions in brain MRI. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) Proceedings MICCAI 2013, Part I. LNCS, vol. 8149, pp. 543–550. Springer, Heidelberg (2013)
13. Subbanna, N.K., Precup, D., Collins, D.L., Arbel, T.: Hierarchical probabilistic gabor and MRF segmentation of brain tumours in MRI volumes. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, part i. LNCS, vol. 8149, pp. 751–758. Springer, Heidelberg (2013)
14. Sled, J.G., Pike, G.B.: Correction for b(1) and b(0) variations in quantitative T(2) measurements using MRI. Magn. Reson. Med. **43**(4), 589–593 (2000)
15. Collins, D.L., et al.: Automatic 3D model based neuro-anatomical segmentation. Hum. Brain Mapp. **3**, 190–208 (1995)
16. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3), 143–155 (2002)
17. Nyl, L.G., et al.: New variants of a method of MRI scale standardization. IEEE Trans. Med. Imag. **19**(2), 143–150 (2000)
18. Subbanna, N., et al.: Iterative multilevel MRF leveraging context and voxel information for brain tumour segmentation in MRI. In: Proceedings of Computer Vision and Pattern Recognition 2014, Columbus, June 2014