

The **next generation** GBCA
from Guerbet is here

Explore new possibilities >

Guerbet | 

© Guerbet 2024 GUOB220151-A

AJNR

Improved Automatic Detection of New T2 Lesions in Multiple Sclerosis Using Deformation Fields

M. Cabezas, J.F. Corral, A. Oliver, Y. Díez, M. Tintoré, C. Auger, X. Montalban, X. Lladó, D. Pareto and À. Rovira

This information is current as
of September 22, 2024.

AJNR Am J Neuroradiol 2016, 37 (10) 1816-1823

doi: <https://doi.org/10.3174/ajnr.A4829>

<http://www.ajnr.org/content/37/10/1816>

Improved Automatic Detection of New T2 Lesions in Multiple Sclerosis Using Deformation Fields

M. Cabezas, J.F. Corral, A. Oliver, Y. Díez, M. Tintoré, C. Auger, X. Montalban, X. Lladó, D. Pareto, and À. Rovira

ABSTRACT

BACKGROUND AND PURPOSE: Detection of disease activity, defined as new/enlarging T2 lesions on brain MR imaging, has been proposed as a biomarker in MS. However, detection of new/enlarging T2 lesions can be hindered by several factors that can be overcome with image subtraction. The purpose of this study was to improve automated detection of new T2 lesions and reduce user interaction to eliminate inter- and intraobserver variability.

MATERIALS AND METHODS: Multiparametric brain MR imaging was performed at 2 time points in 36 patients with new T2 lesions. Images were registered by using an affine transformation and the Demons algorithm to obtain a deformation field. After affine registration, images were subtracted and a threshold was applied to obtain a lesion mask, which was then refined by using the deformation field, intensity, and local information. This pipeline was compared with only applying a threshold, and with a state-of-the-art approach relying only on image intensities. To assess improvements, we compared the results of the different pipelines with the expert visual detection.

RESULTS: The multichannel pipeline based on the deformation field obtained a detection Dice similarity coefficient close to 0.70, with a false-positive detection of 17.8% and a true-positive detection of 70.9%. A statistically significant correlation ($r = 0.81$, P value = $2.2688e-09$) was found between visual detection and automated detection by using our approach.

CONCLUSIONS: The deformation field-based approach proposed in this study for detecting new/enlarging T2 lesions resulted in significantly fewer false-positives while maintaining most true-positives and showed a good correlation with visual detection annotations. This approach could reduce user interaction and inter- and intraobserver variability.

ABBREVIATIONS: BL = baseline; CIS = clinically isolated syndrome; DSC = Dice similarity coefficient; DF = deformation fields; FP = false-positive; FPF = false-positive fraction; FU = follow-up; PD = proton density; TP = true-positive; TPf = true-positive fraction

MR imaging has become a core paraclinical tool for diagnosing and predicting long-term disability and treatment response in patients with multiple sclerosis. Of particular note, several criteria and strategies have been proposed for

prompt identification of suboptimal response in individual patients based on a combination of clinical and MR imaging measures assessed during the first 6–12 months after treatment initiation.^{1–6} These criteria are related to detection of disease activity on follow-up brain MR imaging studies compared with baseline scans, defined as either gadolinium-enhancing lesions or new/enlarging T2 lesions. However, detection of active T2 lesions in patients with MS can be hindered by several factors, such as a high burden of inactive T2 lesions, the presence of small and confluent lesions, inadequate repositioning, and high interobserver variability.⁷ Image subtraction after image registration can overcome these issues by visually cancelling stable disease (lesions that stay the same over time) and providing good visualization and quantification of active T2 lesions (either positively or negatively).^{8,9}

Techniques for automatic detection of active T2 lesions can be classified into 2 categories: intensity-based and deformation-based approaches.¹⁰ In the former, successive scans are analyzed by point-to-point (voxel-to-voxel) comparison, whereas in the

Received October 13, 2015; accepted after revision March 21, 2016.

From the Section of Neuroradiology, Department of Radiology (M.C., J.F.C., C.A., D.P., À.R.) and Centre d'Esclerosi Múltiple de Catalunya, Department of Neurology/Neuroimmunology (M.T., X.M.), Vall d'Hebron University Hospital, Vall d'Hebron Research Institute, Autonomous University of Barcelona, Barcelona, Spain; and Visió per Computador i Robòtica group (M.C., A.O., Y.D., X.L.), University of Girona, Girona, Spain.

Mariano Cabezas held a European Research Committee for Treatment and Research in Multiple Sclerosis/Magnetic Resonance Imaging in MS 2014 fellowship grant. This work has been partially supported by "La Fundació la Marató de TV3", Retos de Investigación grant TIN2014–55710-R, and an MPC UdG 2016/022 grant.

Paper previously presented in part at: Annual Meeting of the American Society of Neuroradiology and Foundation of the ASNR Symposium, April 25–30, 2015; Chicago, Illinois.

Please address correspondence to Mariano Cabezas, PhD, Computer Vision and Robotics Group, Department of Computer Architecture and Technology, Polytechnic School - P-IV Building, University of Girona, 17071 Girona, Spain; e-mail: mariano.cabezas@vhir.org, mcabezas@eia.udg.edu

<http://dx.doi.org/10.3174/ajnr.A4829>

latter, deformation fields obtained by nonrigid registration of the 2 scans are analyzed.

Most of the proposed techniques to detect changes on follow-up images use an image-subtraction process that identifies new T2 lesions¹¹⁻¹³ and include statistical models of intensity changes between scans or other, more complex, supervised strategies. Although segmentation of subtraction images enables quantification of new, enlarging, and resolving MS lesions, automated image analysis that differentiates a true lesion change and noise or artifacts would save considerable time and effort.

Nonrigid registration techniques usually provide a discrete vector field that defines deformations occurring between 2 different images. This vector field can be used to detect evolving processes, including new T2 lesions. Several approaches that use deformation fields (DF) to detect positive changes occurring in longitudinal MR studies have been reported.^{14,15} These approaches focus on detecting and explaining processes undergoing change (ie, lesions shrinking or growing), but not on detecting new lesions, a measure that is now under consideration as a biomarker for monitoring and predicting treatment response.¹⁶

The purpose of this study was to improve automated detection of new T2 lesions on successive brain MR images, by using a novel approach that combines subtraction and DF analysis. This new pipeline will be compared with other approaches, in which a threshold is applied or a postprocessing step is incorporated on the basis of intensity rules.

MATERIALS AND METHODS

Patients

We prospectively analyzed previously acquired data from a cohort of 36 patients with clinically isolated syndrome (CIS) or early relapsing MS (13 women and 23 men; 35.4 ± 7.1 years of age) who underwent brain MR imaging in our center for diagnosis or for monitoring disease evolution or treatment response. All patients with CIS and early relapsing MS demonstrated new T2 lesions on the follow-up scans and were diagnosed according to recent definitions and criteria.^{17,18} Two brain MR imaging acquisitions were obtained in each patient, the first within the first 3 months after the onset of symptoms (baseline [BL]) and the second at 12 months' follow-up after onset (FU). The Vall d'Hebron hospital's ethics committee approved the study, and written informed consent was signed by the participating patients.

MR Image Acquisition

All patients underwent brain MR imaging at BL and FU on the same 3T magnet (Tim Trio; Siemens, Erlangen, Germany) with a 12-channel phased array head coil. The MR imaging protocol included the following sequences: 1) transverse proton density (PD)- and T2-weighted fast spin-echo (TR = 3080 ms/TE = 21–91 ms, voxel size = $0.78 \times 0.78 \times 3.0$ mm³), 2) transverse fast T2-FLAIR (TR = 9000 ms, TE = 87 ms, TI = 2500 ms, flip angle = 120°, voxel size = $0.49 \times 0.49 \times 3.0$ mm³), and 3) sagittal T1-weighted 3D magnetization-prepared rapid acquisition of gradient echo (TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, voxel size = $1.0 \times 1.0 \times 1.2$ mm³).

Expert Analysis

All new and enlarging T2 lesions visually detected on the FU scan were annotated on T2-FLAIR images by using the semiautomated tool included in Jim 5.0 (<http://www.xinapse.com/home.php>). This task was performed by a trained technician who first detected changes visually by using the BL and FU scan and then delineated them semiautomatically by using a subtraction image and both scans. This task was later confirmed by an expert neuroradiologist. The results of this analysis served as the reference standard for comparisons in the study.

Preprocessing

On both BL and FU PD-weighted images, a brain mask was identified and delineated by using the FSL Brain Extraction Tool (bet2 command) (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET>) with the robust center estimation, neck clean-up, and default threshold parameters. The mask was then applied to the other coregistered images (T2-, T2-FLAIR-, and T1-weighted), and the N4 algorithm from the ITK library (<http://www.itk.org/>)¹⁹ was used to correct for bias with the standard parameters for a maximum of 400 iterations. The last preprocessing step was to normalize BL and FU intensity values by using a histogram-matching approach.

Registration and Subtraction

In each patient, T1- and T2-FLAIR-weighted images from the same study were coregistered to the PD-weighted image by using a 3D affine transformation similar to that in previous works.²⁰ The Mattes Mutual Information cost function was minimized by Regular Step Gradient Descent Optimization (https://itk.org/Doxygen320/html/classitk_1_1RegularStepGradientDescentOptimizer.html), and B-spline interpolation was applied. This framework was implemented by using ITK.

The same 3D affine-registration framework was also used before subtraction to warp the BL images to the FU space because patients with CIS and early relapsing MS present with small (or no) overall anatomic changes.²¹ The registration was conducted between both PD-weighted images. After the transformation had been obtained, we applied it to the other images by using B-spline interpolation to subtract the BL PD-, T2-, and T2-FLAIR-weighted images from their corresponding FU images. In the case of BL T2-FLAIR-weighted images, the 2 affine transformations were combined to avoid interpolating more than once.

Affine registration methods are robust to the presence of lesions, and when new lesions appear, deformable models usually show distortions to compensate for the anomalous regions. On the basis of the characteristics of these approaches, we were able to analyze the DF obtained after applying these nonrigid techniques to the registered images. In this study, we applied the multiresolution Demons registration approach²² from ITK initialized with the previous affine transformation. Concretely, we used the DemonsRegistrationFilter (SD = 1) (http://www.itk.org/Doxygen320/html/classitk_1_1DemonsRegistrationFilter.html) with MultiResolutionPDEDformableRegistration (http://www.itk.org/Doxygen320/html/classitk_1_1MultiResolutionPDEDformableRegistration.html) (iterations = 50, levels = 2). This algorithm can produce large localized deformations and has been widely used in brain MR imaging.

Threshold

New and enlarging T2 lesions appear hyperintense in the subtraction image. However, certain regions outside the white matter may also appear hyperintense due to artifacts, noise, inhomogeneity, registration errors, or small anatomic differences. Because our goal was to detect new and enlarging T2 WM lesions, we restricted our search to areas within the WM. To define this region, we applied an automated tissue-segmentation algorithm²³ to the BL and FU scans. This nonparametric algorithm uses an atlas registered to the T1-weighted image in conjunction with the T1-, T2-, and PD-weighted images. This segmentation was applied before the registration step between the 2 image sets. After registration, a final WM mask was obtained as the union of the 2 WM segmentations in the FU space. After defining WM, we smoothed the subtracted images by using the ITK 3D Gaussian filter (DiscreteGaussianImageFilter; http://www.itk.org/Doxygen/html/classitk_1_1DiscreteGaussianImageFilter.html) with a 0.5 SD to reduce the impact of spurious hyperintense regions.²⁰ An automated threshold was then computed for each subtraction image (PD, T2, and T2-FLAIR) and applied separately to obtain 3 initial lesion masks. The thresholds were computed as the mean of the subtraction image within the WM plus 5 SDs to guarantee that only hyperintense regions were detected and to maintain a large number of true-positives (TPs), as proposed previously.²⁰ Lesions of <3 voxels were excluded to reduce the effects of noise.

Lesion Mask Combination

To differentiate between errors and true lesions in each mask, we used the intersection of the 3 masks (PD, T2, and T2-FLAIR). Because differences in the initial masks might still result in false-positive (FP) detections of 1 or 2 voxels, we also applied the lesion size restriction to the combined mask to reduce this effect.

Afterward, the 2 different postprocessing approaches were used independently in order to compare them.

Postprocessing Based on Intensity

While the aforementioned restrictions usually exclude a large number of FPs, they do not completely eliminate this problem. As has been reported,²⁰ some FPs can arise from low intensities in the BL images, caused, for example, by skull-stripping errors. To reduce the effect of these factors and to include local information, we applied a set of suggested intensity-based rules to the BL and FU images²⁰:

- Global rule: To avoid regions with a low intensity, candidate lesions with a mean value under $\mu_{\text{basal}} - 2\sigma_{\text{basal}}$ are discarded, where μ_{basal} and σ_{basal} are the mean and SD of the basal intensities inside the WM ROI.
- Basal neighborhood ratio: New lesions should appear as WM in the basal image. To ensure that, we compute a ratio between the neighboring pixels of the candidate lesions ($\mu_{\text{lesion}}/\mu_{\text{neighbors}}$). If this ratio is <0.9, we discard the candidate lesion. That usually means that there is a dark spot that might appear as a hyperintensity in the subtraction image.
- Follow-up neighborhood ratio: Similarly, new lesions should actually be lesions in the follow-up image. To ensure that, we compute the same ratio. If this ratio is <1, the candidate lesion has a lower intensity profile than its neighboring area, so we discard it.

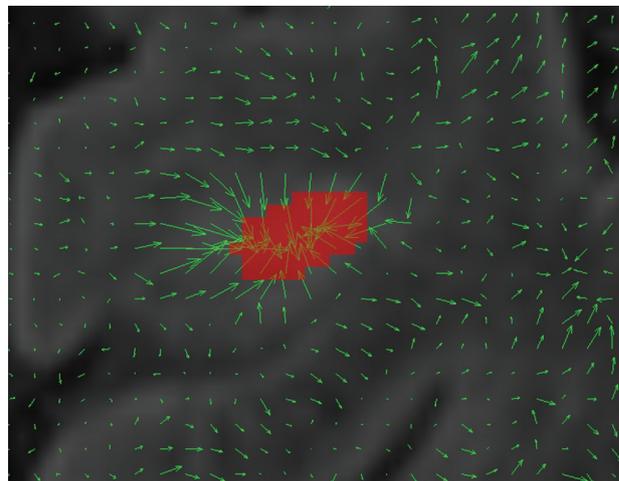


FIG 1. Example of the deformation field inside a new lesion. All arrows point to the lesion center.

Postprocessing Based on Deformation Fields

The Demons algorithm provides DF representing a transformation from the target image (FU scan) to the source image (BL scan). To compensate for hyperintense lesions, the DF goes from outside the lesion to its center (shrinking it), as is illustrated in Figs 1 and 2. Vectors within and in the vicinity of the lesion have a higher modulus than those in other regions of the image. Moreover, no sinking patterns with independent behavior between neighboring vectors are observed far from lesions.

To be able to model and automatically detect this behavior, we defined 3 regional metrics computed from the DF inside each candidate lesion:

- Divergence¹⁵: This vector operator is defined as the volume density of the outward flux of a vector field from an infinitesimal volume around a given point. Given a continuously differentiable vector field \vec{F} , the divergence at a given point is equal to the scalar-valued function:

$$\text{div}\vec{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

In our case,

$$\text{div}DF = G(x)_x + G(y)_y + G(z)_z,$$

where $G(i)_j$ is the j component of the gradient in the i component of the vector field volume.

For new T2 lesions, deformations have an inward flux that is represented by a negative value ($\text{div} DF$ of <0). Therefore, we excluded lesions that had a positive mean value.

- Jacobian¹⁴: We used the Jacobian operator to analyze the DF at each candidate lesion. Values of <1 represent a shrinking process. Regions with a higher value were excluded.
- Concentricity: Due to the inward flux of the vector field within lesions, all vectors point to the center of mass of the lesion. We defined a new operator on the basis of that notion. For each lesion voxel, we computed the vector between the voxel and the center of mass of the lesion. We then com-

puted the scalar product between the DF vector and this concentric vector. Concentric vector fields should have an absolute mean value close to 1; therefore, we excluded all

candidate lesions with an absolute value lower than 0.75. This value indicates that the deformation vector and the concentric vector have a maximum angle of 15°.

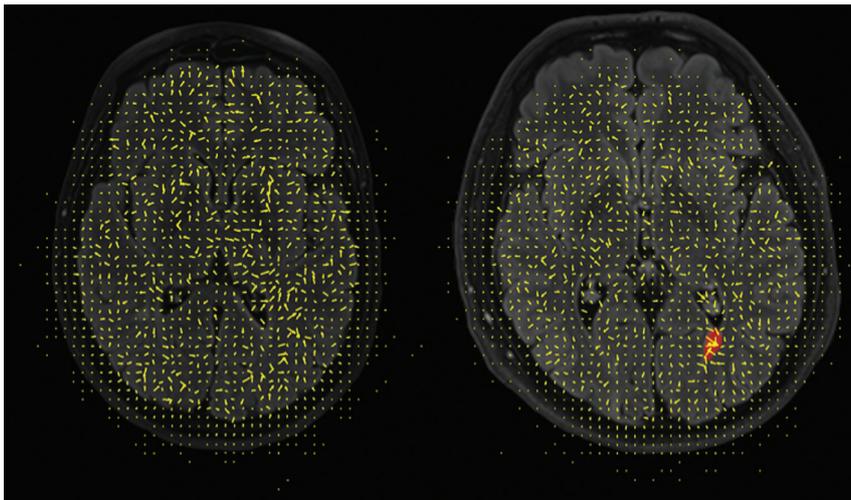


FIG 2. Example of the deformation field for 2 sections. The first image does not contain lesions and presents large deformations with no clear sinking patterns, while in the second image with a lesion, all the arrows inside the lesion point to the center.

Evaluation and Statistical Analysis

To validate use of the DF and the benefits they provide when automatically detecting new T2 lesions, we compared the proposed pipeline to a state-of-the-art approach²⁰ with detection-based measures. In this approach, a lesion is considered TP if it overlaps a ground truth lesion, FP is a detected lesion with no overlap, and FN is a lesion that has not been detected.

The TP fraction (TPf) and FP fraction (FPf) are the ratio measures of TP versus ground truth lesions and FP versus all lesions found, respectively. Therefore, perfect detection would be 100% TPf and 0% FPf. To complement and summarize these measures, we also computed the Dice similarity coefficient (DSC):

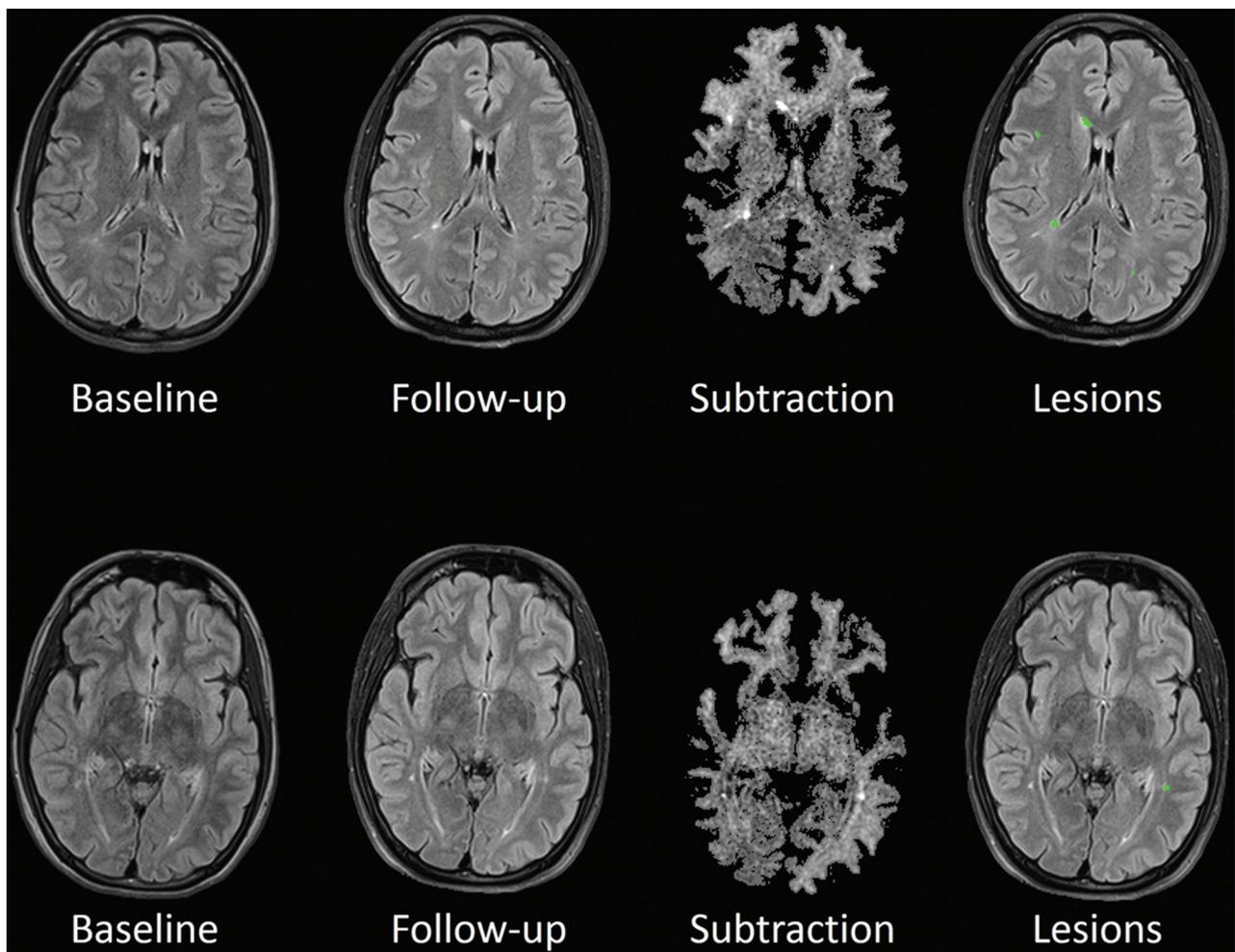


FIG 3. New lesion detection. For each row, the first image is the baseline image, the second is the follow-up image, the third is the subtraction, and the fourth is the lesion analysis over the follow-up image (green = true-positive). The patient has a large number of TPs (100%), with a small number of FPs (0%).

Table 1: Lesion detections obtained for our data base using various approaches

Image	Method	ASD	TPf	FPf	DSC (Lesions)	DSC (Volume)
PD	Threshold	25.80	92.28	93.18	0.11	0.31
	Intensity rules ²⁰	21.90	80.61	83.01	0.24	0.35
	DF	19.91	73.18	77.02	0.30	0.37
T2	Threshold	25.22	93.89	95.88	0.07	0.25
	Intensity rules ²⁰	22.22	64.09	86.35	0.17	0.25
	DF	17.76	81.79	80.84	0.26	0.34
T2-FLAIR	Threshold	27.22	90.24	92.79	0.10	0.26
	Intensity rules ²⁰	21.17	78.34	80.77	0.25	0.31
	DF	21.14	81.22	77.11	0.30	0.33
Combination	Threshold	13.07	91.05	85.61	0.22	0.45
	Intensity rules ²⁰	30.80	51.62	35.87	0.46	0.37
	Proposal	7.89	70.93	17.80	0.68	0.52

Note:—ASD indicates average surface distance.

Table 2: Permutation test ranking of DSC values for the approaches applied on each image separately^a

	Method	Mean P Value
Rank 1 (<1 σ)	T2-FLAIR-DF	.75
	PD-DF	.56
	T2-DF	.53
Rank 2 (<2 σ)	T2-FLAIR ²⁰	.22
	PD ²⁰	.16
Rank 3 (<3 σ)	T2 ²⁰	-.22
	PD-threshold	-.56
	T2-FLAIR-threshold	-.67
	T2-threshold	-.78

^a Methods were ranked relative to the mean and SD of the method with the highest DSC value. Methods in the same rank have similar results, whereas methods in different ranks show significant differences.

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Furthermore, we also performed an evaluation of the actual overlap between lesions by using the volumetric DSC.

Finally, we also included the average surface distance measure from the MICCAI MS Lesion Segmentation Challenge 2008 (<http://www.ia.unc.edu/MSseg/>).²⁴ The border voxels of segmentation and reference are determined. For each voxel along one border, the closest voxel along the other border is determined (by using unsigned Euclidean distance in real-world distances). All these distances are stored, and their average is computed. This value is zero for a perfect segmentation.

A statistical analysis was performed to evaluate the significance of the results obtained. To determine the performance of each key step in our pipeline, we conducted 3 sets of experiments, each focusing on a different aspect. The naïve approach consisted of applying the threshold defined in the “Materials and Methods” section to each subtraction image. We also applied different post-processing approaches to the initial masks separately, and finally, we compared the results of the threshold mask combination to our proposal and a state-of-the-art approach.

First, we performed a Lilliefors test on the measures evaluated and their differences. Due to the number of pipelines evaluated and the statistically proved non-normal distribution of the measures, pair-wise *t* tests were inappropriate. Hence, permutation tests^{20,25} were used to determine significant differences among applying a threshold, using intensity and neighborhood rules, and using DF. Permutation tests yield the mean (μ) and SD (σ) of the fraction of times that the difference in a given measure for a given method is smaller than the remaining methods, with a *P* value \leq

.05. The methods were then ranked by the mean and SD of the method with the highest measured value. Methods in the same rank had similar results, whereas methods in different ranks showed significant differences.

We also performed a Wilcoxon rank sum test among the DSC, TPf, and FPf results for each independent image after the threshold was applied. Finally, the Pearson correlation was used to analyze the manual annotations and the automatic detections obtained with our approach.

RESULTS

The mean results for new T2 lesion detection and segmentation by using each of the approaches are summarized in Table 1. The DSC results with our approach were 0.68 in terms of detection (regions) and 0.52 in terms of segmentation (volume). Moreover, we obtained the lowest average surface difference (7.89 mm) in contrast to the joint threshold (13.07 mm) and with intensity rules (30.80 mm). While the volumetric agreement was lower, it was high enough to validate our detection definition of 1 voxel overlap.

Impact of Postprocessing per Image

Our first set of experiments consisted of applying a threshold to PD-weighted, T2-weighted, and T2-FLAIR-weighted images separately. We compared this naïve approach with a state-of-the-art approach²⁶ based on intensity and spatial rules and the DF rules presented here on each image.

According to Table 1, application of a threshold alone missed some ground truth lesions and resulted in a large number of FPs. Lowering the threshold to include all ground truth detections would be counterproductive because of the number of FPs. In terms of sensitivity alone, both PD-weighted and FLAIR subtractions yielded similar results.

Rankings obtained by statistical permutation testing for the DSC are summarized in Table 2. Negative values indicate lower performance than the method with the highest DSC value. Rank 1 only included approaches that relied on the DF after applying a threshold, whereas rank 2 included approaches that used intensity and neighborhood rules for the PD and T2-FLAIR subtractions. Rank 3 included all methods based on thresholds with a negative *P* value. Because ranking between the approaches differed, we can conclude that there was a significant difference between using DF and intensity/neighborhood rules.

Paired rank sum testing between strategies revealed no significant difference in DSC or TPf among the 3 image subtractions, thus indicating that all 3 images provided similar sensitivity for lesion detection. However, we obtained significant differences for FPf, suggesting that FP detection differed among the images. This difference supports our idea of combining the masks obtained for each subtraction.

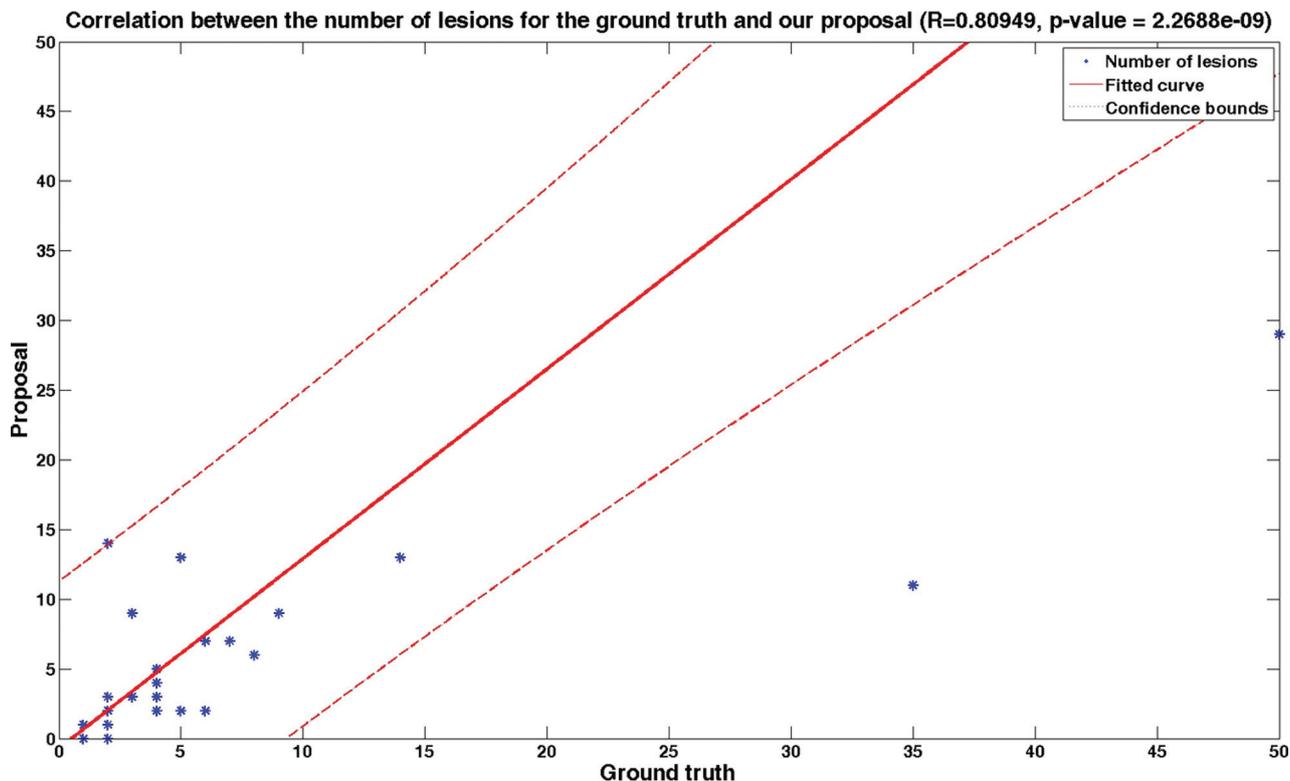


FIG 4. Correlation between the number of ground truth lesions and the number of automatically detected ones (Pearson coefficient = 0.81, $P = 2.2688e-09$).

Impact of the Lesion Mask Combination

When the initial masks for each image were analyzed independently, almost all new T2-WM lesions were detected. However, FP detections were visually different among the images in most cases and, therefore, highly related to the image being visualized.

To validate the assumption that combining the masks significantly improves the results, we performed a second set of experiments and comparisons by using rank sum testing between the lesion mask after applying a threshold to each image independently and the intersection of all 3 masks. Significant differences were found for PPF and DSC ($P < .05$) but not for TPf. Again, this finding suggests that combining all masks reduces the number of FPs without significantly affecting TP detections.

Pipeline Comparison

We also performed an analysis of the last group in Table 1 (mask combination in the 3 different strategies). In this case, we obtained significant differences ($P < .05$) for all 3 measures (DSC, TPf, and FPF) between the intersection mask and the 2 approaches based on postprocessing. This result indicates that the DSC improvement was due to the considerable decrease in FPs detrimental to the number of TPs. This is the usual trade-off encountered when dealing with postprocessing techniques, in which some TPs are excluded (eg, due to image artifacts) to reduce the number of FPs. We also found significant differences ($P < .05$) in all 3 measures between the 2 automatic approaches (our proposal and that of Ganiler et al²⁰), reinforcing the notion that our DF strategy yields better performance. Qualitative examples of the results obtained with our proposal are shown in Fig 3.

Table 3: Analysis of the TPf before and after postprocessing with deformation fields for different sizes^a

Image	Method	3–10	11–50	50+
Combination	Combination (threshold)	71.43	72.38	95.16
	Proposal	42.86	48.57	77.42

^a Lesions between 3 and 10 voxels are considered small; lesions between 11 and 50 voxels, medium; and lesions with >50 voxels, large.

A significant correlation ($r = 0.81$, $P = 2.2688e-09$) was found between annotations based on visual detection and our automated approach for detecting new T2 lesions (Fig 4). We then analyzed the effect of the 3 DF-based measures and found that they all had a similar impact in most cases; however, some FPs were only detected by one of them, with no apparent pattern.

Lesion Analysis per Volume

Finally, we analyzed lesion detection by groups of similar size. Table 3 summarizes the results before and after postprocessing by using the deformation field obtained. As expected, lesions with a small size (between 3 and 10 voxels) have a low detection rate (42.86%). Due to their small size, the deformation field cannot fully capture them and they are discarded during the postprocessing step. As the volume increases, the deformation field presents a clearer pattern that we can detect with the rules presented in this article. Even though for lesions of a medium size (between 11 and 50 voxels) the TPf is still lower than 50% (48.57%), this value increases for large lesions of >50 voxels (77.42%). Moreover, the TPf decreases from 69.23% before postprocessing to 23.08% with lesions of <7 voxels.

DISCUSSION

New/enlarging T2 lesion count is a common measure used to monitor and predict treatment response in patients with MS.¹⁻⁶ Trained radiologists perform this task by visual analysis of 2 successive MR images, a time-consuming task associated with high interobserver variability.⁷ The pipeline proposed in this study may be of value for assisting or even replacing visual analysis for detecting active MS lesions on T2-weighted images.

The method is completely autonomous and automated and does not require user input or a training set. Furthermore, the process is computationally fast because it mainly relies on subtraction and registration. With an optimized Demons algorithm, it takes only minutes to segment all new T2 lesions in a single patient, with a low number of FP detections.

We obtained significant results with a data base of 36 patients, and we also tested our algorithm without any modification with a small clinical trial dataset. This dataset had a reduced number of images ($n = 10$) that were provided by 3 different centers. Even though promising results were obtained with this initial test (DSC for lesions = 0.79, DSC for volume = 0.60, TPf = 74.15, FPf = 9.61), an exhaustive analysis with a larger number of patients should be performed to prove that the method performs similarly with different acquisition setups.

However, currently, it is not possible to detect new black holes (even though a postprocessing step could be included to differentiate between new lesions and new black holes by using the T1-weighted images).

Current studies are working on the definition and implementation of a new “no evidence of disease activity” treatment.^{6,27} This decision model relies on, among others, the detection of new/enlarging T2 lesions as a biomarker and requires a high specificity and sensitivity because the number of FPs could suggest an undesirable change in treatment. Therefore, reducing the number of FPs when using automatic tools is a key factor. However, current subtraction techniques usually rely on intensity information, which can misguide detection due to local inhomogeneities or small changes. While these FPs can be reduced by using spatial information, a registration technique that overfits a free-form deformation incorporates this local information and provides better insight into changes occurring due to development of a new lesion or one that changes in size.

Automated algorithms usually obtain better scores when lesion count or lesion volume is high, but they often have shortcomings when the lesion volume or volume change is small.^{11-13,20} We also compared ours to a current state-of-the-art technique that has been validated with 1.5T imaging. 3T imaging provides better resolution and a higher signal-to-noise ratio, from which registration techniques can benefit. Therefore, to demonstrate that DF provide a better means to differentiate subtraction artifacts and true disease activity (in terms of lesions), we used 3T imaging, in which DF provide a better understanding of evolving processes.

CONCLUSIONS

We have presented a new automated pipeline to detect new brain T2 lesions and positive changes in disease activity in patients with clinically isolated syndrome or early relapsing multiple sclerosis.

This technique relies on DF information and provides more reliable measurement of changes occurring between 2 successive MR images than other currently available approaches. Significant differences in accurate lesion detection were found between this technique and other current approaches, and a strong correlation and higher overlap were seen between our approach and visual lesion detection. These findings indicate that the proposed technique may be of value for application in clinical studies investigating disease activity, monitoring, and treatment effects, providing a decrease in user interaction and likely a reduction in inter- and intraobserver variability.

Disclosures: Mariano Cabezas—RELATED: Grant: Magnetic Resonance Imaging in MS (MAGNIMS), Comments: MAGNIMS/European Research Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) Fellowship 2014. Further information can be found at http://www.ectrims.eu/wp-content/uploads/2013/04/ECTRIMS-MAGNIMS-MRI-fellowship-awardees_for-website_2015.pdf. Mar Tintorè—UNRELATED: Board Membership: Genzyme, Roche, Biogen, Novartis, Teva, Merck, Sanofi, Almirall, Bayer; Consultancy: Biogen; Grants/Grants Pending: Genzyme,* Roche,* Biogen,* Novartis,* Teva,* Merck,* Sanofi,* Almirall,* Bayer*; Payment for Lectures (including service on Speakers Bureaus): Genzyme, Roche, Biogen, Novartis, Teva, Merck, Sanofi, Almirall, Bayer; Payment for Manuscript Preparation: Biogen; Payment for Development of Educational Presentations: Biogen. Cristina Auger—UNRELATED: Payment for Lectures (including service on Speakers Bureaus): Novartis, Stendhal America, Biogen. Xavier Montalban—UNRELATED: Consultancy: Actelion, Almirall, Bayer, Biogen, Genzyme, Merck, Novartis, Octapharma, Receptos, Roche, Sanofi-Genzyme.* Deborah Pareto—UNRELATED: Payment for Lectures (including service on Speakers Bureaus): Novartis, Genzyme. Àlex Rovira—UNRELATED: Consultancy: Biogen Idec, Genzyme, Novartis, Olea Medical; Payment for Lectures (including service on Speakers Bureaus): Biogen Idec, Stendhal America, Genzyme, Novartis, Olea Medical; Payment for Development of Educational Presentations: Novartis, Stendhal America, Genzyme, Biogen.* Money paid to the institution.

REFERENCES

1. Río J, Castelló J, Rovira A, et al. Measures in the first year of therapy predict the response to interferon beta in MS. *Mult Scler* 2009;15:848–53 CrossRef Medline
2. Sormani MP, De Stefano N. Defining and scoring response to IFN- β in multiple sclerosis. *Nat Rev Neurol* 2013;9:504–12 CrossRef Medline
3. Sormani MP, Río J, Tintorè M, et al. Scoring treatment response in patients with relapsing multiple sclerosis. *Mult Scler* 2013;19:605–12 CrossRef Medline
4. Prosperini L, Mancinelli CR, De Giglio L, et al. Interferon beta failure predicted by EMA criteria or isolated MRI activity in multiple sclerosis. *Mult Scler* 2014;20:566–76 CrossRef Medline
5. Freedman MS, Selchen D, Arnold DL, et al; Canadian Multiple Sclerosis Working Group. Treatment optimization in MS: Canadian MS Working Group updated recommendations. *Can J Neurol Sci* 2013;40:307–23 CrossRef Medline
6. Stangel M, Penner IK, Kallmann BA, et al. Towards the implementation of ‘no evidence of disease activity’ in multiple sclerosis treatment: the multiple sclerosis decision model. *Ther Adv Neurol Disord* 2015;8:3–13 CrossRef Medline
7. Altay EE, Fisher E, Jones SE, et al. Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. *JAMA Neurol* 2013;70:338–44 CrossRef Medline
8. Moraal B, Wattjes MP, Geurts JJ, et al. Improved detection of active multiple sclerosis lesions: 3D subtraction imaging. *Radiology* 2010;255:154–63 CrossRef Medline
9. Moraal B, van den Elskamp IJ, Knol DL, et al. Long-interval T2-weighted subtraction magnetic resonance imaging: a powerful new outcome measure in multiple sclerosis trials. *Ann Neurol* 2010;67:667–75 CrossRef Medline
10. Lladó X, Ganiler O, Oliver A, et al. Automated detection of multiple

- sclerosis lesions in serial brain MRI. *Neuroradiology* 2012;54:787–807 CrossRef Medline
11. Battaglini M, Rossi F, Grove RA, et al. **Automated identification of brain new lesions in multiple sclerosis using subtraction images.** *J Magn Reson Imaging* 2014;39:1543–49 Medline
 12. Elliott C, Arnold DL, Collins DL, et al. **Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI.** *IEEE Trans Med Imaging* 2013;32:1490–503 CrossRef Medline
 13. Sweeney EM, Shinohara RT, Shea CD, et al. **Automatic lesion incidence estimation and detection in multiple sclerosis using multi-sequence longitudinal MRI.** *AJNR Am J Neuroradiol* 2013;34:68–73 CrossRef Medline
 14. Rey D, Subsol G, Delingette H, et al. **Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis.** *Med Image Anal* 2002;6:163–79 CrossRef Medline
 15. Thirion JP, Calmon G. **Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences.** *IEEE Trans Med Imaging* 1999;18:429–41 CrossRef Medline
 16. Vrenken H, Jenkinson M, Horsfield MA, et al; MAGNIMS Study Group. **Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis.** *J Neurol* 2013;260:2458–71 CrossRef Medline
 17. Lublin FD, Reingold SC, Cohen JA, et al. **Defining the clinical course of multiple sclerosis: the 2013 revisions.** *Neurology* 2014;83:278–86 CrossRef Medline
 18. Polman CH, Reingold SC, Banwell B, et al. **Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria.** *Ann Neurol* 2011;69:292–302 CrossRef Medline
 19. Tustison NJ, Avants BB, Cook PA, et al. **N4ITK: improved N3 bias correction.** *IEEE Trans Med Imaging* 2010;29:1310–20 CrossRef Medline
 20. Ganiler O, Oliver A, Diez Y, et al. **A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies.** *Neuroradiology* 2014;56:363–74 CrossRef Medline
 21. Diez Y, Oliver A, Cabezas M, et al. **Intensity based methods for brain MRI longitudinal registration: a study in multiple sclerosis patients.** *Neuroinformatics* 2014;12:365–79 CrossRef Medline
 22. Thirion JP. **Image matching as a diffusion process: an analogy with Maxwell's demons.** *Med Image Anal* 1998;2:243–60 CrossRef Medline
 23. Cabezas M, Oliver A, Roura E, et al. **Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding.** *Comput Methods Programs Biomed* 2014;115:147–61 CrossRef Medline
 24. Styner M, Lee J, Chin B, et al. **3D segmentation in the clinic: a grand challenge II: MS lesion segmentation.** *MIDAS J* 2008;1–6
 25. Menke J, Martinez TR. **Using permutations instead of Student's t distribution for p-values in paired-difference algorithm comparisons.** In: *Proceedings of the 2004 IEEE International Joint Conference of Neural Networks*, Budapest, Hungary. July 25–29, 2004
 26. Dancey C, Reidy J. *Statistics without Maths for Psychology: Using SPSS for Windows.* New York: Prentice Hall; 2004
 27. Tintore M, Rovira À, Río J, et al. **Defining high, medium and low impact prognostic factors for developing multiple sclerosis.** *Brain* 2015;138(pt 7):1863–74 CrossRef Medline