Medical Image Analysis 16 (2012) 114–126

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Differential MRI analysis for quantification of low grade glioma growth

Elsa D. Angelini [a,*], Julie Delon [b], Alpha Boubacar Bah [c], Laurent Capelle [c], Emmanuel Mandonnet [d]

[a] Institut Telecom, Telecom ParisTech, CNRS LTCI, Department of Signal and Image Processing, 46 rue Barrault, 75013 Paris, France
[b] CNRS LTCI, Telecom ParisTech, Department of Signal and Image Processing, 46 rue Barrault, 75013 Paris, France
[c] Neurosurgery Department, Hôpital Pitié Salpêtrière, 47-83 boulevard de l'Hôpital, 75013 Paris, France
[d] Neurosurgery Department, Hôpital Lariboisière, 2, Rue Ambroise Paré, 75010 Paris, France

ARTICLE INFO

Article history:
Received 19 May 2010
Received in revised form 17 May 2011
Accepted 20 May 2011
Available online 6 June 2011

Keywords:
MRI
Brain tumor
Longitudinal studies
Low-grade glioma

ABSTRACT

A differential analysis framework of longitudinal FLAIR MRI volumes is proposed, based on non-linear gray value mapping, to quantify low-grade glioma growth. First, MRI volumes were mapped to a common range of gray levels via a midway-based histogram mapping. This mapping enabled direct comparison of MRI data and computation of difference maps. A statistical analysis framework of intensity distributions in midway-mapped MRI volumes as well as in their difference maps was designed to identify significant difference values, enabling quantification of low-grade glioma growth, around the borders of an initial segmentation. Two sets of parameters, corresponding to *optimistic* and *pessimistic* growth estimations, were proposed. The influence and modeling of MRI inhomogeneity field on a novel midway-mapping framework using image models with multiplicative contrast changes was studied. Clinical evaluation was performed on 32 longitudinal clinical cases from 13 patients. Several growth indices were measured and evaluated in terms of accuracy, compared to manual tracing. Results from the clinical evaluation showed that millimetric precision on a specific volumetric radius growth index measurement can be obtained automatically with the proposed differential analysis. The automated *optimistic* and *pessimistic* growth estimates behaved as expected, providing upper and lower bounds around the manual growth estimations.

1. Introduction

Magnetic resonance imaging plays a prominent role in the follow-up of patients harboring a grade II gliomas. Together with an increase in seizures frequency, radiological growth is the most reliable mean to assess disease progression. A recent study for example in (Caseiras et al., 2009) has shown that low-grade gliomas tumor volume growth over the course of 6 months was the best predictor of time to transformation into higher grades, independent of relative cerebral blood volume, DTI information, age, sex and histological findings. As an alternative to tumor volume, it has been shown in the past few years that grade II glioma kinetics can be estimated by monitoring the mean tumor diameter, deduced from total tumor volume (Mandonnet et al., 2008). The mean tumor diameter was studied in various clinical settings such as before any treatment (Mandonnet et al., 2003), after surgery (Mandonnet et al., 2010), after chemotherapy (Ricard et al., 2007), or after radiotherapy (Pallud et al., 2009). The clinical importance of individual assessment of the radiological kinetics of tumor growth before any treatment cannot be over-emphasized. Indeed, the prognosis of patients with a velocity of diametric expansion (VDE) greater than 8 mm/year is similar to patients harboring a high grade glioma (Pallud et al., 2006). This finding underlines the need of an automated and reliable tool to accurately measure glioma growth. Such tool could enable to discriminate fast progressors (VDE > 8 mm/year) from slow progressors (VDE < 8 mm/year) from two MRI exams acquired over a 3 months period. The aim of this paper is to provide a new computational framework to address this challenge. A simple difference map of two co-registered MRI volumes will likely fail to detect any tumor changes, due to contrast differences between the two images, generated by field inhomogeneity and differences in scanner sensitivity. Longitudinal comparison of MRI exams needs to be insensitive to typical scanning differences, including head position, scanner noise, dynamic range and voxel size, while being able to detect biological evolutions of the tumor (and eventually other cerebral brain structures) depicted as morphological and signal changes. Before presenting our method, based on the statistical analysis of difference maps after non-linear normalization of the MRI intensity data, we first provide an overview of the literature on the topic of change detection in longitudinal image acquisitions, focusing on monitoring of brain pathologies with MRI.

Monitoring change detection on series of image acquisitions has been getting a lot of attention recently for applications such as remote sensing and video processing, as reviewed in (Radke et al.,

* Corresponding author. Tel.: +33 1 45 81 78 90; fax: +33 1 45 81 37 94.
E-mail address: elsa.angelini@telecom-paristech.fr (E.D. Angelini).

1361-8415/$ - see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.media.2011.05.014

2005). For brain imaging, this problem has been addressed in several contexts: mapping longitudinal changes of brain structures for Alzheimer disease (Thompson et al., 2004), diameter measurements to monitor responses to tumor treatment (Therasse et al., 2000), formulation of the evolution of tumor delineation observed on longitudinal MRI to adapt a generic dynamic tumor model, to a patient-specific case (Hogea et al., 2008; Konukoglu et al., 2010a,b), or image comparison to quantify the evolution of white matter lesions for multiple sclerosis (MS) disease (Nyul et al., 2000; Bosc et al., 2003; Prima et al., 2002).

Regarding the underlying methodology enabling comparison of image information, we can distinguish three types of approaches: direct image comparison, segmentation-based comparison and registration-based comparison.

For *direct image comparison*, previous work from Nyul et al. (2000) proposed an MRI standardization protocol based on histogram mapping on a generic mean histogram determined through training. In addition, landmark gray values were identified on the histogram to perform separate histogram matching for landmark intervals. Bosc et al. (2003) proposed a histogram mapping of longitudinal MRI exams maximizing their joint histogram with a median value estimator. They made the observation that optimizing joint histogram values was very sensitive to image noise. Non-parametric image difference analysis using scattergrams (i.e. series of conditional histograms pairs of intensity values) was studied in Bromiley et al. (2002), relying on the probabilistic detection of abnormal pairs of intensity values, which is computationally costly and might not be robust to small spatial clusters of slow intensity changes. Prima et al. (2002), used polynomial fit of longitudinal MRI scans onto a target average scan, via least trimmed square approximation, which required complex numerical implementation. In Jager and Hornegger (2010), a non-rigid joint histogram mapping was proposed to normalize intensities of multi-modal MRI exams onto a reference multimodal exam. Several applications were illustrated, demonstrating the robustness of the mapping to the presence of a brain pathology with increasing areas. While this approach provides an elegant solution to global MRI mapping, robust to the presence of localized pathologies, it was not evaluated in the context of pathological growth quantification. In addition, the use of non-rigid histogram registration involves some risks in terms of contrast modification, non-convergence of the intensity normalization process and attenuation of the contrast of pathological structures.

In the specific context of brain tumor change detection, an interesting approach was proposed by Liu et al. (2003) to select only significant changes on SPGR, T2-weighted and FLAIR longitudinal brain MRI exams using an elaborated processing framework on direct difference maps. This processing involved noise level map estimations from repeated exams on a given subject, based on previous work by Lemieux et al. (1998), thresholding of the difference maps above noise level and filtering of the remaining values with a learned anatomical map of artifacts, computed via spatial normalization of difference maps from 40 subjects. In comparison, our proposed method avoids the needs of repeated MRI exams and offers a framework likely to be more easily reproducible and consistent across different clinical studies.

For *segmentation-based* longitudinal brain MRI studies, previous works include Meier and Guttmann (2003) for MS disease and Prastawa et al. (2003) for tumor growth. For brain volume change detection, a consistent longitudinal alignment approach was proposed by Xue et al. (2006) which incorporated adaptive image clustering, spatio-temporal smoothness constraints, and image warping to jointly segment longitudinal MRI brain data sets of individual subject and characterize temporal changes due to development, aging, or disease. In Clark et al. (1998), an automated tumor segmentation was tested on longitudinal series of four exams for patients with glioblastomas multiforme. They reported very high true-positive values when comparing tumor volumes to manual tracing. They also observed errors in growth or shrinkage predictions based on tumor volume measurements from the segmented data.

For *registration-based* longitudinal brain MRI studies, displacement fields obtained via non-rigid registration of longitudinal MRI were used to detect regions corresponding to growth and atrophy (Rey et al., 2002). These methods provided accurate localization of MS lesion changes, but inaccurate segmentations of the lesion morphology. More recently, smooth diffeomorphisms were proposed to quantify shape changes, as reviewed in Miller (2004). These methods offer the great advantage of being invariant to small rigid registration errors, but have a high computational complexity, when performed in 3D, and typically require a spatial resolution superior to the one available with FLAIR data. In addition, these methods are limited to global growth indices and are not suited for precise growth area boundaries. A study in Konukoglu et al. (2008) compared direct image comparison to a registration-based method, for global growth quantification of brain meningiomas.

We finally point out an original approach, based on time series analysis taking into account the characteristic time function of variable lesions, that was proposed for MS disease in Gerig et al. (2000), involving a large set of longitudinal MRI scans, and a training phase to learn the different types of lesion evolution.

Brain tumor segmentation remains a challenging task, requiring to include necrosis, edema, infiltrations and eventually post-surgery cavities. Since inter-observer variability for manual tracing of brain tumor can range up to 15% (Kaus et al., 2000), our goal was to propose a direct differential MRI analysis framework, relying on longitudinal gray-values normalization. Grade II glioma are better delineated on FLAIR MRI sequences (Bynevelt et al., 2001), and we focused our study on this protocol, although a 5–6 mm slice thickness coupled with the fact that axial planes orientations may change between two successive MRI exams will obviously limit the accuracy of tumor growth measurement (both for the manual and automated methods).

In this paper we present a novel method for detection and quantification of low-grade tumor evolution in longitudinal single-protocol MRI studies. A methodology for differential analysis of longitudinal MRI volumes is proposed, based on non-linear gray value mapping and detection of statistically significant differences. The proposed method is designed to avoid severe limitations from previously proposed normalization methods, including:

- Avoiding training for histogram learning so that the method can be applied in a generic fashion to any type of longitudinal MRI of different body parts.
- Avoiding the use of a template histogram shape.
- Avoiding the use of joint histograms, reported to be very sensitive to image noise.

A clinical study was performed on 32 FLAIR longitudinal clinical cases from 13 patients. Several growth indices were measured and evaluated in terms of accuracy, compared to manual tracing. Results from the clinical evaluation show that millimetric precision on a specific growth index measurement can be obtained automatically with the proposed differential analysis.

2. Materials and methods

2.1. Midway mapping theory

2.1.1. Invariance to contrast change

The initial midway mapping framework was proposed in Delon (2004) to enable comparison of a pair of images (I_1, I_2) representing

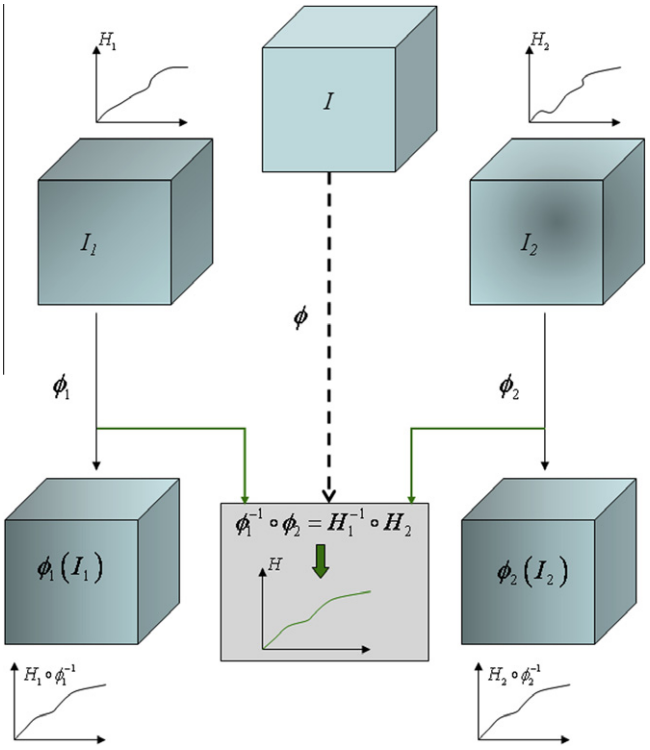

Fig. 1. Illustration of the general framework for midway-based image intensity mapping.

the same scene, while remaining independent of image contrast changes. The basic idea was to assign to the pair of images (I_1, I_2) a common gray level distribution, defined as a single target histogram corresponding to a reasonable midway (i.e. average) between the original gray level histograms of I_1 and I_2, as illustrated in Fig. 1. In the original paper (Delon 2004) the author discusses alternative normalization approaches, such as histogram mapping on a generic target, and demonstrates the superiority of the midway framework in terms of information preservation.

In order to go into further details regarding the mapping method, let us denote by h_I the gray level histogram of an image I and by H_I the corresponding cumulative histogram (or cumulative distribution). Recall that if ϕ is a continuous and strictly increasing contrast change function, the cumulative histogram of $\phi(I)$ can be written as:

$$H_{\phi(I)} = H_I \circ \phi^{-1}. \tag{1}$$

Indeed:

$$\forall \lambda, H_{\phi(I)}(\lambda) = \#\{x, \phi(I(x)) \leq \lambda\} = \#\{x, I(x) \leq \phi^{-1}(\lambda)\} \tag{2}$$

where # denotes the cardinal of a set. It follows that the histogram $h_{\phi(i)}$ of $\phi(I)$ can be written:

$$h_{\phi(I)} = \frac{h_I \circ \phi^{-1}}{\phi' \circ \phi^{-1}}. \tag{3}$$

In particular, if ϕ and H_I are both continuous and strictly increasing, the cumulative histogram of the image $\phi^{-1} \circ H_I(I)$ is ϕ, which means that I can be mapped onto any gray level distribution with a well chosen contrast change function.

In (Delon 2004), the notion of midway gray level distribution was derived empirically in the case where both images shared the same geometry, i.e. when there exists two strictly increasing functions f and g such that $I_1 = f(I)$ and $I_2 = g(I)$. The functions f and g were assumed to be increasing to preserve gray level ordering (i.e. level sets). In this particular case, a midway image was defined between I_1 and I_2 instead of just a midway histogram. For this purpose, it was assumed that there exists a function $\psi : \mathbb{R}^2 \to \mathbb{R}$, independent of f, g and I and such that the midway image between $f(I)$ and $g(I)$ can be written $\psi(f(I),g(I))$. If we add the hypothesis that the midway image between I and $I + \lambda$ should be $I + \lambda/2$, the only solution for ψ is the function:

$$\psi : (x, y) \to \frac{x + y}{2}. \tag{4}$$

This leads to a midway image defined as:

$$I_{midway} = \frac{f(I) + g(I)}{2} \tag{5}$$

The cumulative histogram of this midway image is expressed as:

$$H_{midway} = \left(\frac{f \circ H_I^{-1} + g \circ H_I^{-1}}{2} \right)^{-1} \tag{6}$$

In the general case, when we don't know the contrast change functions f and g nor the original image I, this framework leads to the following definition of the midway histogram between two observed cumulative histograms H_{I_1} and H_{I_2}:

$$H_{midway} = \left(\frac{H_{I_1}^{-1} + H_{I_2}^{-1}}{2} \right)^{-1} \tag{7}$$

We note that strictly speaking, we should use here the "pseudo inverses" of H_{I_1} and H_{I_2} when these cumulative histograms are not invertible.

The images I_1 and I_2 mapped on this midway histogram become respectively:

$$I_{midway1} = H_{midway}^{-1} \circ H_{I_1}(I_1) = \frac{I_1 + H_{I_2}^{-1} \circ H_{I_1}(I_1)}{2} \tag{8}$$

and

$$I_{midway2} = H_{midway}^{-1} \circ H_{I_2}(I_2) = \frac{I_2 + H_{I_1}^{-1} \circ H_{I_2}(I_2)}{2}. \tag{9}$$

In Angelini, Atif et al. (2007), we presented an initial application of the midway framework to compare longitudinal brain FLAIR and SPGR MRI data of patients with low-grade gliomas with promising results on the quantification of tumor's evolution. Several aspects of the specific problem of MRI normalization were not addressed in this initial paper, and are discussed in the following paragraphs.

2.1.2. Invariance to MRI inhomogeneity

As reviewed in (Vovk et al., 2007), MRI inhomogeneity fields are typically modeled as multiplicative spatial fields:

$$I_1(x) = I(x) \times \lambda(x). \tag{10}$$

In this case, the arithmetic mean $\frac{x+y}{2}$ in Equation (4) can be replaced by a geometric mean $\sqrt{xy}$. Theoretically, this corresponds to impose that given a real (non-zero) value λ, the midway-mapped image between λI and $\frac{1}{\lambda} I$ should be I.

This mapping leads to the following expression for the midway histogram:

$$H_{midway} = \left(\sqrt{H_{I_1}^{-1} H_{I_2}^{-1}} \right)^{-1} \tag{11}$$

With such mapping, images I_1 and I_2 become respectively:

$$I_{midway1} = H_{midway}^{-1} \circ H_{I_1}(I_1) = \sqrt{I_1 \times H_{I_2}^{-1} \circ H_{I_1}(I_1)} \tag{12}$$

and

$$I_{midway2} = H^{-1}_{midway} \circ H_{I_2}(I_2) = \sqrt{I_2 \times H^{-1}_{I_1} \circ H_{I_2}(I_2)} \quad (13)$$

Since inhomogeneity fields are generally assumed spatially uniform and slowly varying, $\lambda(x)$ in Equation (10) locally corresponds to a constant multiplier λ_0. Such model requires that midway mapping be applied on overlapping patches (blocks) where $\lambda(x)$ is assumed constant. Overlap of patches and fusion of mapped patches via averaging is used to enforce some spatial continuity of the mapping of gray values in the resulting image. This approach is similar to the general framework of block-based lapped transforms (Malvar, 1992). In all our experiments, patches overlap was set to 1/3 as a tradeoff between computational cost and visual quality of the mapped data.

When applied to clinical data, the patch-based multiplicative midway mapping cannot directly estimate the individual inhomogeneity fields from the two scans being mapped, but can normalize these fields to identical values. This means that the proposed midway mapping framework eliminates the need for inhomogeneity correction prior to MRI comparison. More precisely, for a pair of MRI images (I_1, I_2), we model small image patches data as $(\lambda_1 I, \lambda_2 I)$. From Equations (12) and (13) we can express the following ratios between pre and post midway mapped images:

$$\frac{I_1}{I_{midway1}} = \frac{I_1}{\sqrt{I_1 \times H^{-1}_{I_2} \circ H_{I_1}(\lambda_1 I)}} = \frac{\sqrt{I_1}}{\sqrt{H^{-1}_{I_2} \circ H_I(I)}} = \frac{\sqrt{I_1}}{\sqrt{\lambda_2 H^{-1}_I \circ H_I(I)}} = \frac{\sqrt{I_1}}{\sqrt{\lambda_2 I}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}, \quad (14)$$

and

$$\frac{I_2}{I_{midway2}} = \frac{I_2}{\sqrt{I_2 \times H^{-1}_{I_1} \circ H_{I_2}(\lambda_2 I)}} = \frac{\sqrt{I_2}}{\sqrt{H^{-1}_{I_1} \circ H_I(I)}} = \frac{\sqrt{I_2}}{\sqrt{\lambda_1 H^{-1}_I \circ H_I(I)}} = \frac{\sqrt{I_2}}{\sqrt{\lambda_1 I}} = \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}, \quad (15)$$

based on the fact that $H_{I_1}(I_1) = H_I(I)$ and $H^{-1}_{I_1} = \lambda_1 H^{-1}_I$ for $I_1 = \lambda_1 I$, (and similarly for I_2).

Therefore, if we look at the product of the two ratios from Equations (14) and (15), we should obtain a flat image equal to one. We illustrate the evaluation of this theoretical result on a clinical case with two longitudinal FLAIR MRI scans, in Fig. 2. Using a multiplicative midway mapping with 4×4 patches overlapping by 1/3, we obtained very high adequacy between the theoretical and experimental results regarding the value of the product of the two ratios, being equal or very close to one almost everywhere, especially inside the brain. Some inconsistencies occurred in the background, corrupted with acquisition noise. This result was important as it validated our data model and midway mapping framework on clinical cases. We also estimated the inhomogeneity fields on the two MRI cases with the N4ITK tool (Tustison et al., 2010) implemented in the 3DSlicer software tool (Pieper et al., 2006), using 4×4 control points for the B-Splines, which is the default recommended setting and corresponds to our patch-based midway normalization setting. We compared the inhomogeneity fields of the two longitudinal exams prior and after midway intensity normalization. We observed in Fig. 2 that midway mapping of the two clinical exams was indeed capable of mapping the two inhomogeneity fields and removing all the spatial patterns of their initial differences. We observed that for inhomogeneity fields with maximum value of 1.2, and original difference values up to 0.15, the midway mapping

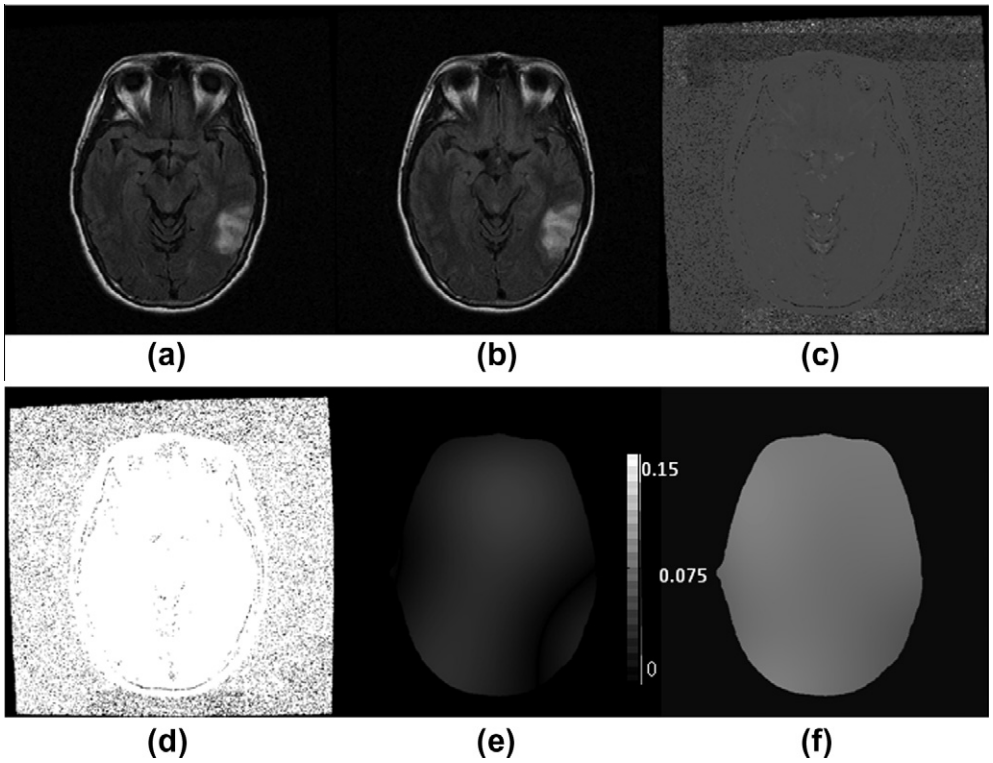

Fig. 2. Evaluation of the image data model for the multiplicative midway mapping on a clinical case. (a and b) Original corresponding axial slices of the FLAIR MRI images at time 1 and time 2. (c) Spatial map of the products of the ratios of the mapped versus the original FLAIR data from the two longitudinal times. (d) Map (c) thresholded for values between [0.5–1.5], displaying values of one almost everywhere, especially in the brain. Values of one correspond to areas where midway mapping of the inhomogeneity fields and the image models with piecewise constant multiplicative inhomogeneity fields are correct. (e and f) Absolute difference of the inhomogeneity fields of images at time 1 and time 2, displayed with the same colormap, estimated on the original data (e) and the midway-normalized data (f).

was able to homogenize the spatial outline of the differences to an almost flat value of 0.07, within the brain. Larger differences (up to 0.15) were observed in the region of the evolving brain tumor (where the working assumption of similarity between local image content is not valid). Midway mapping was therefore capable of implicitly mapping inhomogeneity fields with an error of 7%. This error level is negligible compared to the level of differences in tumoral growth areas.

2.2. Difference map processing

2.2.1. Difference map characteristics

In the experimental section, we illustrate that even though difference maps (after midway normalization) visually clearly detect tumor region growths corresponding to high difference values, the range of values of these differences is completely unpredictable since tumoral tissue is rather heterogeneous and can vary greatly between two exams. In addition, high difference values are also observed on specific structures such as CSF ventricles and surrounding tissues (skull bone, fat, gyri, etc.). In this context, simple difference map thresholding is difficult to perform in an automated fashion and provides a poor selectivity of "significant" difference locations. Therefore, in order to automatically process the midway-based difference maps, in a generic fashion, we derived a statistical test designed to select "significant" difference values corresponding to potential tumor growth. This test was based on individual Gaussian distribution models for normal brain tissues (white-gray matter and CSF), as commonly assumed, for example in Lemieux et al. (1998). In this context, because of the presence of the tumor which is not correctly modeled with a Gaussian distribution, intensity values from the two scans being compared were assumed to be either drawn from the same Gaussian distribution (i.e. there was no change in the tissue type and distributions can be directly compared thanks to midway mapping), or drawn from different distributions, because of a change in tissue type, due to mis-registration or tumoral infiltration. These tests were performed locally, for each voxel. In Lemieux et al. (1998), the authors proposed to use a Generalized Likelihood Ratio Test (GLRT) comparing the individual distributions estimated at time 1 and 2 to a joint distribution estimated on the combined voxels from both times. Values of the GLRT test were much greater than one if the assumption of tissue change between the two times was true, but typically remained around one if the assumption of similar tissue types was true (due to the intrinsic variability of local statistics estimations on the image data). Therefore, selection of voxels with "significant" changes required the identification of a threshold value, greater than one, which could be delicate to manipulate. We propose an alternative approach, with the formulation of a different statistical test, enforcing the equality of local mean values of the voxels being compared.

2.2.2. Statistical test

We assume at this stage that we are given two images I_1 and I_2, normalized with midway mapping and co-registered. For each pixel i, we want to know if the difference $D(i) = (I_1(i) - I_2(i))$ is meaningful (i.e. due to a structural difference), or on the contrary simply explained by the local statistics of the individual images. To answer this question, we rely on a statistical hypothesis testing framework, that we further detail in 2D but is straightforwardly extendable to 3D. Let us denote by Λ_i the $(2s+1)\times(2s+1)$ centered neighborhood of the pixel i. We note $n = (2s+1)^2$. We assume that the differences $D(j)$ are realizations of independent and identically distributed (i.i.d.) random variables following a given centered Gaussian (normal) distribution when j belongs to Λ_i. This local hypothesis is reasonable in practice if the images are well registered and intensity-mapped. More precisely, if we denote by X_j the random variable corresponding to the gray level difference between I_1 and I_2 at j, we define the following *null hypothesis*:

H_0^i : the random variables X_j, $j \in \Lambda_i$, are i.i.d. and follow a centered normal distribution $N(0,\sigma_i^2)$, where σ_i represents the standard deviation of the normal distribution.

Under the hypothesis H_0^i, the random variable $\overline{X}(i) = \frac{1}{n}\sum_{j\in\Lambda_i} X_j$ corresponding to the local average pixel value on the neighborhood Λ_i also follows a normal distribution $N(0,\sigma_i^2/n)$.

The probability that the random variable $|\overline{X}(i)|$ is larger than a given value δ is therefore:

$$F_i[\delta] := P_{H_0^i}[|\overline{X}(i)| > \delta] = 1 - \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_i}\int_{-\delta}^{\delta} e^{-\frac{nx^2}{2\sigma_i^2}}dx. \tag{16}$$

Assuming that we have a local estimation of the standard deviation σ_i in the neighborhood Λ_i, the null hypothesis H_0^i can thus be tested by computing $F_i[|\overline{D}(i)|]$, where $\overline{D}(i) = \frac{1}{n}\sum_{j\in\Lambda_i} D(j)$ is the observed average gray level difference in the neighborhood Λ_i. If $F_i[|\overline{D}(i)|]$ is too small, the observed difference $|\overline{D}(i)|$ cannot be explained by H_0^i and we decide to reject the null hypothesis. More precisely, the pixel i is detected as a meaningful difference if:

$$F_i[|\overline{D}(i)|] \leqslant p_0, \tag{17}$$

where p_0 is a pre-set threshold value (significance level). This is equivalent to saying that, if the mean of I_1 around i is too different from the mean of I_2 in the same region, the hypothesis H_0^i is rejected and we consider that there is a structural difference between both images at pixel location i.

Now, observe that if we note Erf() the error function of the standard normal distribution, the probability $F_i[\delta]$ can be rewritten:

$$F_i[\delta] = 1 - \frac{1}{\sqrt{\pi}}\int_{-\sqrt{\frac{n}{2}}\frac{\delta}{\sigma_i}}^{\sqrt{\frac{n}{2}}\frac{\delta}{\sigma_i}} e^{-y^2}dy = 1 - \mathrm{Erf}\left(\sqrt{\frac{n}{2}}\frac{\delta}{\sigma_i}\right). \tag{18}$$

For a given threshold p_0, Equation (17) is thus satisfied if:

$$|\overline{D}(i))| \geqslant \frac{\alpha\sigma_i}{\sqrt{n/2}} \tag{19}$$

where $\alpha > 0$ is the quantile of level $1 - p_0$ of the Erf() function (i.e. the quantity which satisfies $\mathrm{Erf}(\alpha) = 1 - p_0$). The threshold $\frac{\alpha\sigma_i}{\sqrt{n/2}}$ depends on the neighborhood size n used to estimate the local mean $\bar{D}(i)$ and on the local standard deviation σ_i of the differences. The first parameter corresponds to the random variable being tested, the second one corresponds to the parameter of the distribution model. In practice, we estimate the local standard deviation σ_i on a neighborhood Λ_i^* centered on pixel i and of size $(2t+1)\times(2t+1)$ pixels. Working with a pre-set p_0 value enables us to tune the statistical test via the two parameters n and σ_i. Regarding the estimation of the local average $\bar{D}(i)$, increasing the size n of Λ_i decreases the threshold in Equation (19), as well as extends the range of influence of $\bar{D}(i)$ around locally high differences. The first effect is more significant and makes the thresholding test more permissive (i.e. more easily verified and at more locations). This leads to systematic larger growth areas, and a more *pessimistic* growth estimation. Regarding the local variance, increasing the size t of Λ_i^* leads to larger variance values of $D(i)$ and therefore increases the threshold in Equation (19). This results in a more restrictive thresholding test, identifying smaller growth areas and providing a more *optimistic* growth estimation. For example, using the sequence of [Λ_i, Λ_i^*] window sizes $[s,t] = [1,4] \to [1,2] \to [1,1]$ leads to increasing growth areas while using the sequence of [Λ_i, Λ_i^*] window sizes $[s,t] = [4,1] \to [2,1] \to [1,1]$ leads to decreasing growth areas. We chose to work with a fixed Λ_i^* neighborhood size of $t = 1$ for the variance estimation and two Λ_i neighborhood sizes $s = 1, 4$ to define

optimistic growth estimates with $[s, t] = [1, 1]$ and *pessimistic* growth estimates with $[s, t] = [4, 1]$.

Similarly to standard statistical thresholding tests, we fixed the significance level (in this study $p_0 = 10^{-10}$ was set empirically and was robust to all cases being tested) to threshold the p-values from the test in Equation (16), which is equivalent to selecting $\alpha \simeq 4.6$ in Equation (19). With this fixed thresholding framework, we manipulate the size of the local patches used to estimate the statistics of the tested variables and the observations, to implicitly manipulate patient-specific thresholds. The use of local average rather than direct difference values enforces spatial coherence in areas of large values and a zero-mean Gaussian distribution in noisy areas. This averaging is similar to the pre-smoothing tasks classically performed in statistical parametric mapping (SPM) of noisy data (Marchini and Presanis 2004). The size of the smoothing window Λ_i is used as a parameter of the statistical test, and implicitly affects the thresholding level in Equation (19), similarly to the Bonferroni correction used to weight the false discovery rate in SPM.

3. Experiments and results

We first performed a simple experiment to illustrate the aptitude of the proposed midway normalization and statistical testing on midway-based difference maps to handle MRI field inhomogeneity and localize tumoral growth. We then performed a clinical study on 32 FLAIR MRI cases, testing several growth indices. Preliminary results from this study were presented in Angelini et al. (2010).

3.1. Clinical examples of difference maps

To process our series of clinical cases, we used multiplicative-midway mapping, with a 2D 4×4 patch-based implementation (with overlaps set to 1/3 of patch size), without pre-correction of the inhomogeneity field. With this setting, we guaranteed that the patch size was bigger than the tumor, limiting the influence of evolving tumoral tissues in the mapping process. We chose to

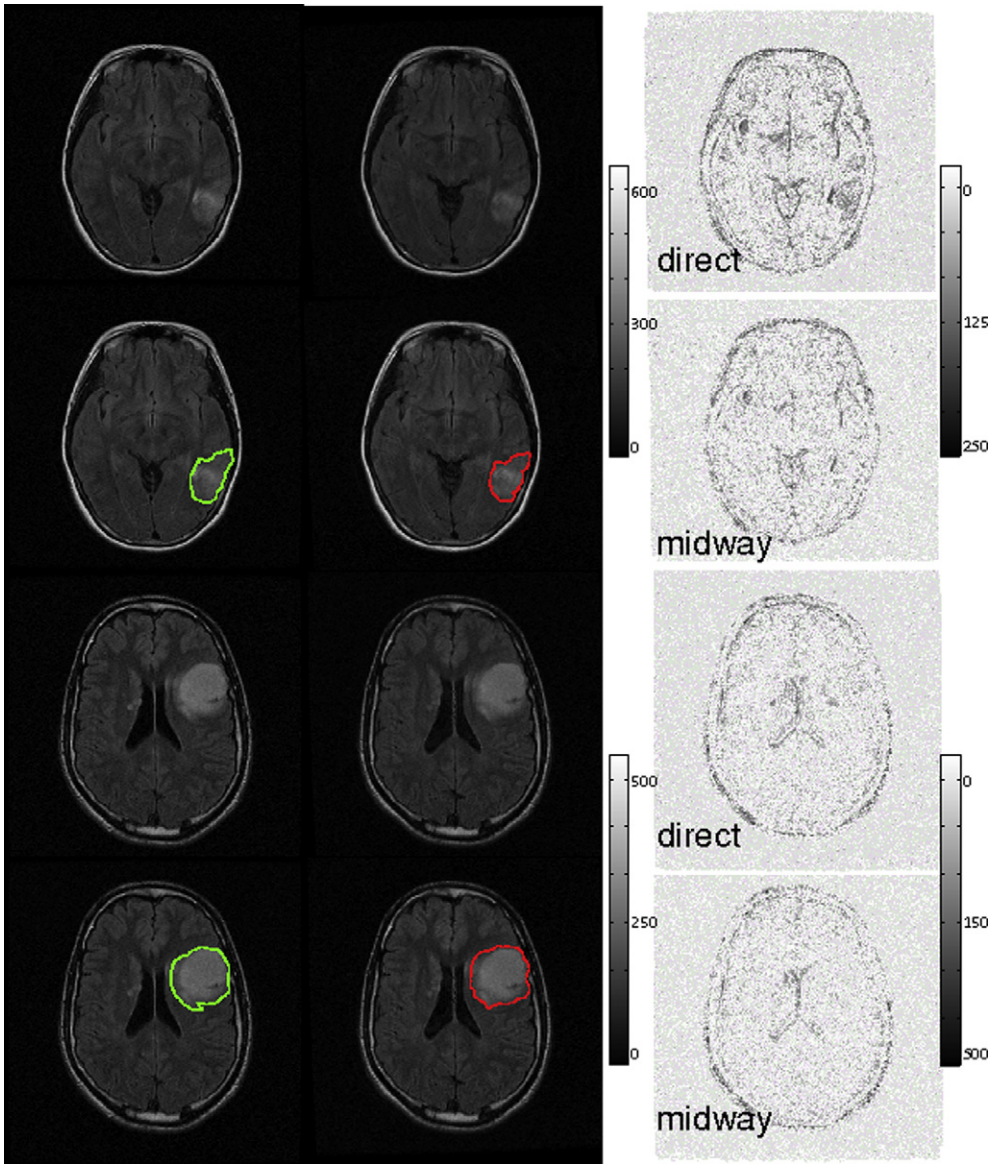

Fig. 3. Illustrations of direct and midway-based difference maps between two longitudinal FLAIR MRI data sets, on two clinical cases with low-grade glioma. Corresponding axial slices from two longitudinal data sets acquired on the same patient are displayed along with the manually traced contours at time 1 (green) and time 2 (red). Direct difference maps (labeled as direct) and midway-based difference maps (labeled as midway) are also displayed. For these two cases, with no tumor evolution, the midway-based difference maps appear more selective, inside and around the tumor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

perform midway mapping in 2D rather than in 3D. This choice increases the computational time of intensity mapping but was found to provide slightly more accurate results when computing the difference maps and extracting statistically significant differences. This is likely due to the fact that the assumption of a piecewise constant multiplicative inhomogeneity field is more correct on 2D patches than on 3D patches, with large variations of the inhomogeneity field across thick slices. This issue was discussed in (Leemput et al., 1999), who proposed a 2D polynomial bias field estimation on each slice separately for optimal correction.

3.1.1. Midway-based difference maps

To illustrate the challenges raised by longitudinal scan comparison, we computed in Fig. 3, on two clinical cases with low-grade gliomas (with image properties detailed in the Clinical Study Section), difference maps from direct subtractions of two pre-registered FLAIR MRI exams for the same patient.

The dynamic ranges of the longitudinal data sets were similar between the two scans (up to 100 difference in maximal intensity value). A direct comparison of the two exams, via computation of difference map representing their absolute difference values, displayed large variations of intensity inside the tumor, which would be misleading for any visual inspection. On the other hand, the difference map between the two midway-mapped scans provided a much more selective difference map, with almost no difference inside the tumor. In both cases, no evolutions were measured on the illustrated slices, based on manual tracings from a clinical expert, and therefore the difference maps should not display any important difference values around the tumoral border at time 1 (which was the case).

This simple experiment confirmed that linear gray-scale mapping was not sufficient to bring longitudinal MRI exams to a common gray-scale range of values permitting direct comparison. On the other hand, patch-based multiplicative midway mapping of the two longitudinal MRI exams enabled direct comparison of the data via the computation of a difference map where high difference values were only observed on structure borders (especially on the skull, the CSF ventricles and some gyri) and on the tumor growth area, but not inside homogeneous tissues such as the white and gray matters or similar tissues such as the initial tumoral area.

3.1.2. Selection of significant differences

An illustration of manual tumor delineation and growth estimations is provided in Fig. 4 for a clinical case, displaying corresponding slices from two original longitudinal image data sets, manual segmentation of the tumor from an expert clinician, the statistical thresholding results for window sizes $\{[s,t]_{opt}, [s,t]_{pes}\}$ equal to $\{[1,1],[4,1]\}$ corresponding to small *optimistic* and large *pessimistic* estimations of tumor growth, targeted to encompass the manual tracing estimation.

Some significant differences were still selected inside the tumor, and on other tissues, illustrating the great challenge in handling the variability in MRI appearance for longitudinal selection of differences in tissue types, along with registration errors. Therefore, to evaluate the proposed methodology, we needed to define a growth mask around the tumor at time 1. This selection was performed using a manual outline of the tumor at time 1, defined by a clinical expert. Obtaining an automated detection of the tumor's contours (at either time 1 or time 2) was beyond the scope of this study and still represents a very challenging task, as reviewed in Angelini, Clatz et al. (2007).

It is important to point out that the proposed test required that the MRI data sets were mapped to a common range of gray scale intensity values (as performed by the midway mapping) to be compared in terms of pixel intensity distributions or pixel intensity difference values.

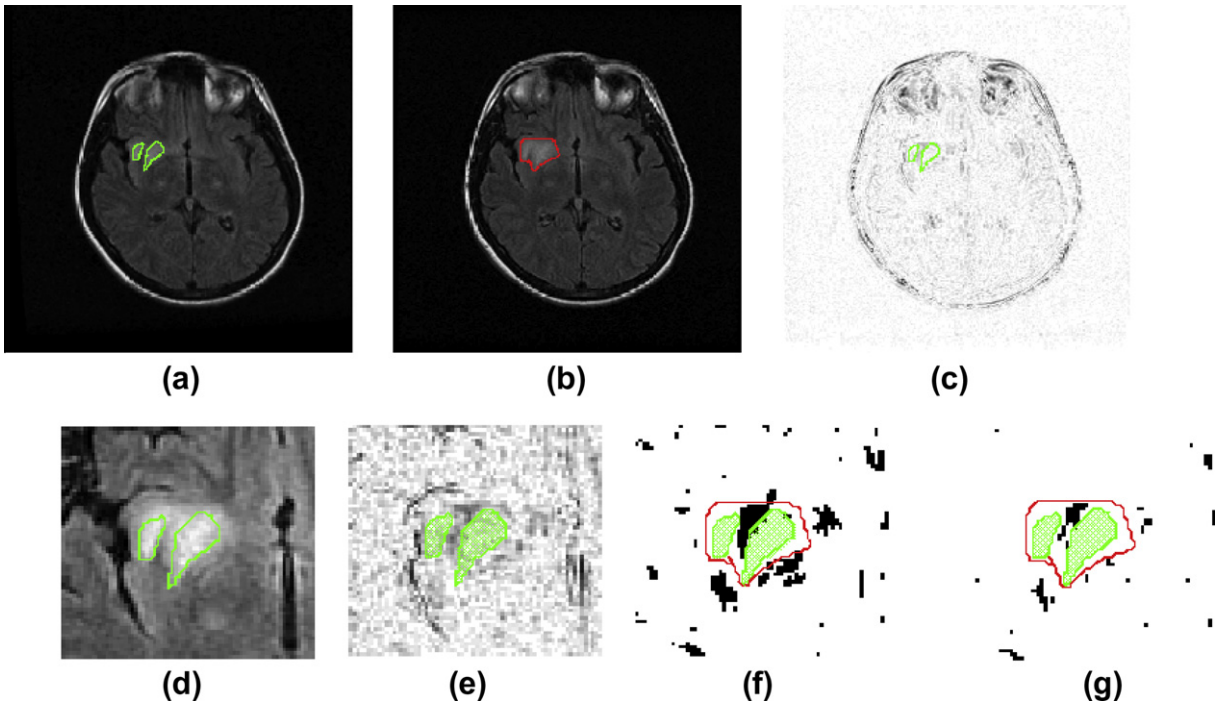

Fig. 4. Selection of "significant" difference values between two longitudinal MRI FLAIR scans with a low-grade glioma, based on statistical testing of the midway-based difference map. (a and b) Original MRI FLAIR data at time 1 (a) and time 2 (b), along with the manual segmentation of the tumor at time 1 (green), and at time 2 (red). (c) Difference map after midway mapping. (d and e) Region of interest around the tumor on the MRI FLAIR data at time 2 (d) and on the difference map (e). (f and g) Thresholded statistical maps for parameter settings corresponding to *pessimistic* (large) growth estimations (f) and *optimistic* (small) growth estimations (g). The shaded area inside the segmentation at time 1 is masked out for growth estimation. N.b. The difference maps are displayed in inverted contrast for better visualization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Clinical study for tumoral growth quantification

3.2.1. Setup

A clinical study was performed on 32 clinical FLAIR pairs of longitudinal scans from 13 patients, to automatically measure tumor growth, given an initial segmentation. The FLAIR MRI exams were performed on a 1.5 T GE SIGNA scanner, with a 2D IR brain scanning protocol with the following parameters for fluid attenuation: TR = 10,000 ms, TE = 149 ms, TI = 2400 ms, Flip angle = 90°, FOV = 240 cm × 240 cm) with or without injection of Gadolinium contrast agent. Seven patient cases included only two consecutive scans, and the remaining six cases included 4, 7 or 9 longitudinal scans. These data sets were of dimension either 512 × 512 or 256 × 256 in axial slices (spatial resolution of 0.5 × 0.5 or 0.9 × 0.9 mm) and included 20 slices (corresponding to a standard FLAIR slice thickness of 6.5 mm). All data sets were manually traced for tumor delineation by an expert neurosurgeon, without midway mapping. A second expert neurosurgeon traced three additional cases (one with 5 times and two with 2 times) to evaluate inter user variability.

Experiments reported in this section were performed on presurgical or post-surgical data, in the context of monitoring the evolution of low-grade gliomas as a primary tumor or as a residual tumor after surgery.

MRI registration was performed with the free software tool FSL (Smith et al., 2004), using correlation similarity measures and 7 degrees of freedom (translation, rotation and one global scale parameter), to constrain intra-patient deformations.

Four tumor growth indices were tested to compare midway-based automated and manual estimations:

1. Growth of the tumor's volume *V* in milliliters (ml).
2. Growth of the tumor's volume in %.
3. Growth of maximal 2D radius *R*, estimated as the radius of maximum erosion of the growth area, over each slice.
4. Growth of the radius of the equivalent sphere *Rs*, defined as the sphere with the same volume as the initial tumor plus the growth area (in 3D).

The 2nd growth index is related to the RECIST and WHO tumor growth measures, while the 3rd index was shown to be a reliable and sensitive measure to evaluate growth rate on patients with low-grade gliomas (Mandonnet et al., 2003).

All cases were processed automatically, with the following sequence of processing steps:

1. Intensity-based affine registration of the two longitudinal MRI data sets (in 3D). Simultaneous registration of the manual segmentations.
2. Midway mapping with multiplicative image model on 4 × 4 patches (2D).
3. Statistical selection of significant differences between the midway mapped data with two sets of parameter settings for *optimistic* and *pessimistic* growth estimations (2D).
4. Post-processing of the selected significant differences to preserve only components inside a growth region of interest (ROI) defined around the contours of the tumor at time 1, and remove components corresponding to dark intensity values (2D).
5. Definition of the tumoral growth areas by removing the area of the tumor segmented at time 1 from either the area of the tumor segmented at time 2 (manual growth area) or the selected significant differences (automated growth area).
6. Computation of the four growth indices for the manual and the automated growth areas.

3.2.2. Localization of tumoral growth

In step 4, when selecting statistical significant difference values, we needed to automatically select values only related to tumor growth, and not to registration errors. Tumor growth patterns are not necessarily connected to the tumor borders at time 1 and can spread by a certain amount beyond the initial tumor margins. To localize the tumor growth patterns, we created a growth region of interest (ROI) by dilating the segmentation from time 1, with a circular structuring element of radius 4 mm, defining a ROI around the initial tumor borders. This value was defined based on the facts that average spontaneous growth-rates and minimal post-operative growth rates of low-grade gliomas were manually estimated around 4 mm/year (Mandonnet et al., 2003; Mandonnet et al., 2010). For each slice, 2D connected components from the selected significant differences masks were extracted and their connexity with the growth ROI was tested.

Rather than relying on 3D connectivity with the growth ROI mask, we combined mask shapes from adjacent slices to handle the large slice thickness in FLAIR MRI data. The fact that we were studying low-grade gliomas with slow growth patterns was critical here, to rely on regions of interest close to the tumor.

A final correction of the selected connected components of the difference mask was performed to remove dark components corresponding to large registration errors in the CSF (especially in the ventricles and on the gyri). Based on the weak assumption of a Gaussian distribution of the pixel intensities inside the tumor, we computed tumor statistics under the manually segmented area at time 1 and removed difference mask components with intensities at time 2 below the mean tumor value minus four standard deviations (corresponding to pixels much darker than inside the tumor).

3.2.3. Tumoral size and growth measurements

Average, minimal and maximal growth measurements are reported in Tables 1 and 2, along with the initial tumor volumes and radii, estimated from the manually traced contours. As can be inferred from these initial values, the tumors were quite large, with a large range of values for the volume and a narrow range of values, well centered around the average value, for the radii. Growth measurements are reported based on manual tracing (man.) and based on statistical selection of significant differences (auto). The two tables report average growth values for large *pessimistic* estimations (Table 1) and small *optimistic* estimations (Table 2), corresponding to the two sets of parameters proposed for local statistics estimations on the midway-mapped data.

We obtained a large range of values for volume growth measurements (especially in %), and a smaller range of values for the radii. The automated method behaved as expected, providing upper and lower bounds of manual average growth estimations. These upper and lower bounds were well constrained. There was also a very high similarity between manual maximal and minimal growth values and, respectively, the maximal *pessimistic* estimations and the minimal *optimistic* estimations.

Inter-user variability was evaluated for the measurements of tumoral size and growth with manual tracing and for the sensitivity of the growth indices measures with respect to the manual delineation used for time 1. For the tumoral size, inter-user variability was measured equal to: 3.1 ± 1.9 ml for *V*, 0.7 ± 0.8 mm for *R* and 0.9 ± 0.4 mm for *Rs*. The volume variability corresponds to 15% of the average volume value. For the manual growth estimation, inter-user variability was measured equal to: 1.7 ± 1.3 ml for *V*, 1.9 ± 1.3 mm for *R* and 0.8 ± 0.3 mm for *Rs*. These results showed that both manual volumes and maximum radius measures were quite sensitive to manual tracing, even for growth estimations which might cancel potential tracing bias of under or over estimation of the tumoral region. Comparing manual radius-based growth measures, inter-user variability was higher for maximum radius growth than for instantaneous values, while the two were comparable for the sphere equivalent radius. Finally, regarding

Table 1
Average *pessimistic* (i.e. large) tumoral growth measurements over 32 cases, between two longitudinal FLAIR MRI data sets. Growth indices are: volume (*V*) in ml, value of maximal radius (radius *R*) and radius of equivalent sphere (radius *Rs*) in mm. Average initial volume and radii values at time 1 are provided, based on manual tracing.

	Tumor volume, *V* (ml)	Tumor radius, *R* (mm)	Tumor radius, *Rs* (mm)	Tumor growth, *V* (ml)		Tumor growth, *V* (%)		Tumor growth, *R* (mm)		Tumor growth, *Rs* (mm)	
				Auto	Man.	Auto	Man.	Auto	Man.	Auto	Man.
Average	48.1	12.6	21.1	15.0	8.1	37.7	21.8	6.4	4.7	2.2	1.2
Std	35.0	4.1	5.7	9.8	10.2	19.5	26.6	2.8	2.8	0.9	1.3
Max	123.6	20.6	30.9	36.7	39.8	98.5	111.8	15.9	15.0	4.4	5.0
Min	6.2	5.6	11.4	1.3	0.3	13.5	0.5	3.3	0.9	0.7	0.1

Table 2
Average *optimistic* (i.e. small) tumoral growth measurements over 32 cases, between two longitudinal FLAIR MRI data sets. Growth indices are: volume (*V*) in ml, value of maximal radius (radius *R*) and radius of equivalent sphere (radius *Rs*) in mm. Average initial volume and radii values at time 1 are provided, based on manual tracing.

	Tumor volume, *V* (ml)	Tumor radius, *R* (mm)	Tumor radius, *Rs* (mm)	Tumor growth, *V* (ml)		Tumor growth, *V* (%)		Tumor growth, *R* (mm)		Tumor growth, *Rs* (mm)	
				Auto	Man.	Auto	Man.	Auto	Man.	Auto	Man.
Average	48.1	12.6	21.1	5.7	8.1	13.9	21.8	3.4	4.7	0.9	1.2
Std	35.0	4.1	5.7	4.9	10.2	9.9	26.6	1.4	2.8	0.6	1.3
Max	123.6	20.6	30.9	19.5	39.8	44.4	111.8	6.6	15.0	2.5	5.0
Min	6.2	5.6	11.4	0.3	0.3	2.0	0.5	1.4	0.9	0.1	0.1

the sensitivity of the growth indices measures with respect to the manual delineation used for time 1, variability of the growth indices measures was measured equal to: 3.9 ± 1.3 ml for *V*, 1.7 ± 0.8 mm for *R* and 0.4 ± 0.4 mm for *Rs* for *pessimistic* measures, and 2.5 ± 1.6 ml for *V*, 3.9 ± 5.1 mm for *R* and 0.6 ± 0.5 mm for *Rs* for *optimistic* measures. It appears that only the radius of the equivalent sphere *Rs* provides satisfactory robustness to inter-user variability of the tracing of the tumor at time 1.

In Table 3 we report average errors for the automated estimations of the four growth indices, comparing to manual tracing, with *pessimistic* (large) and *optimistic* (small) estimation frameworks. We also report the minimal errors from the pair of (*pessimistic, optimistic*) estimations. The histograms of automated and manual growth measures errors are plotted in Fig. 5. We can observe that the average growth volume error, lower than 8 ml, is rather satisfactory, given that the average tumoral volume was 48 ml (16% of error). This should be appreciated with precaution though, given the large values of maximum growth volume discrepancies. The distributions of the growth volume errors showed that the *optimistic* estimations were more consistent with respect to manual tracing estimations than *pessimistic* estimations. Radius growth indices provided, on average, highly accurate measures when comparing automated and manual estimations (errors between 1 to 3 mm). Evaluation of the minimum errors between pairs of (*pessimistic, optimistic*) estimations shows a systematic decrease of the maximum and average errors, as well as a decrease by half of the standard deviation of the errors distribution. This confirms that for all cases with a large error with one estimation, the other one was

Table 3
Average errors of tumor growth estimations between automated and manual methods, for *pessimistic* (large) and *optimistic* (small) midway-based index measures, as well as the minimum error (*Min* (l, s)) between the two estimations, where l refers to large and s to small estimations.

	Tumor growth, *V* (ml)			Tumor growth, *V* (%)			Tumor growth, *R* (mm)			Tumor growth, *Rs* (mm)		
	Large	Small	*Min* (l, s)	Large	Small	*Min* (l, s)	Large	Small	*Min* (l, s)	Large	Small	*Min* (l, s)
Average	8.4	5.2	3.0	22.5	14.3	8.8	2.6	1.9	1.2	1.2	0.8	0.5
Std	7.5	5.5	2.7	16.7	17.6	9.6	2.5	1.9	1.1	0.8	0.7	0.4
Max	32.1	21.4	10.2	62.3	67.4	42.4	13.1	8.4	4.2	3.1	2.8	1.6
Min	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

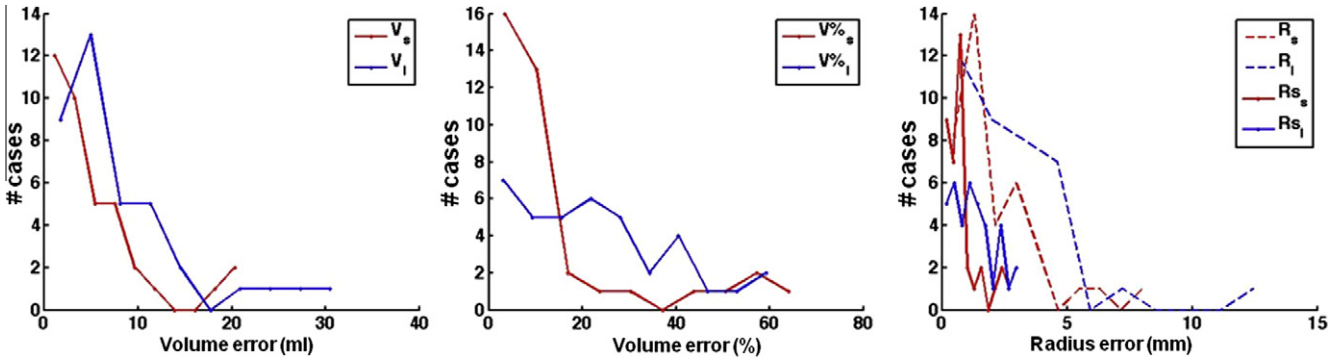

Fig. 5. Histograms of growth measurements errors for volume measures (left), volume percentage measures (center) and radius-based measures (right). Pessimistic ($V\%_l$, V_l, R_l, Rs_l) histograms are plotted in blue and optimistic ($V\%_s$, V_s, R_s, Rs_s) histograms are plotted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

much more accurate. For some cases, both estimations made similar errors (i.e. the ground truth corresponded to an intermediate growth between the small and large estimations). This result suggests that the development of a method to select the "best" growth estimation, might lead to even higher precision in growth indices, compared to manual tracing.

The *R* (maximum radius) measure generated some large errors in few cases but the histogram plots show that the majority of cases provided errors below 5 mm. The *Rs* (sphere-equivalent) radius reported much smaller maximum errors and the histogram plots show a very peaked distribution for the *optimistic Rs* measure, confirming that the proposed method is very efficient at accurately measuring very small tumoral growths with this index.

The overall analysis of these errors suggests that volume measurements might not be recommended as a reproducible tumoral growth index, being too sensitive to the variability in manual tracing (Kaus et al., 2000) in the case of low-resolution FLAIR MRI data. On the other hand the *Rs* radius-based growth measures provided

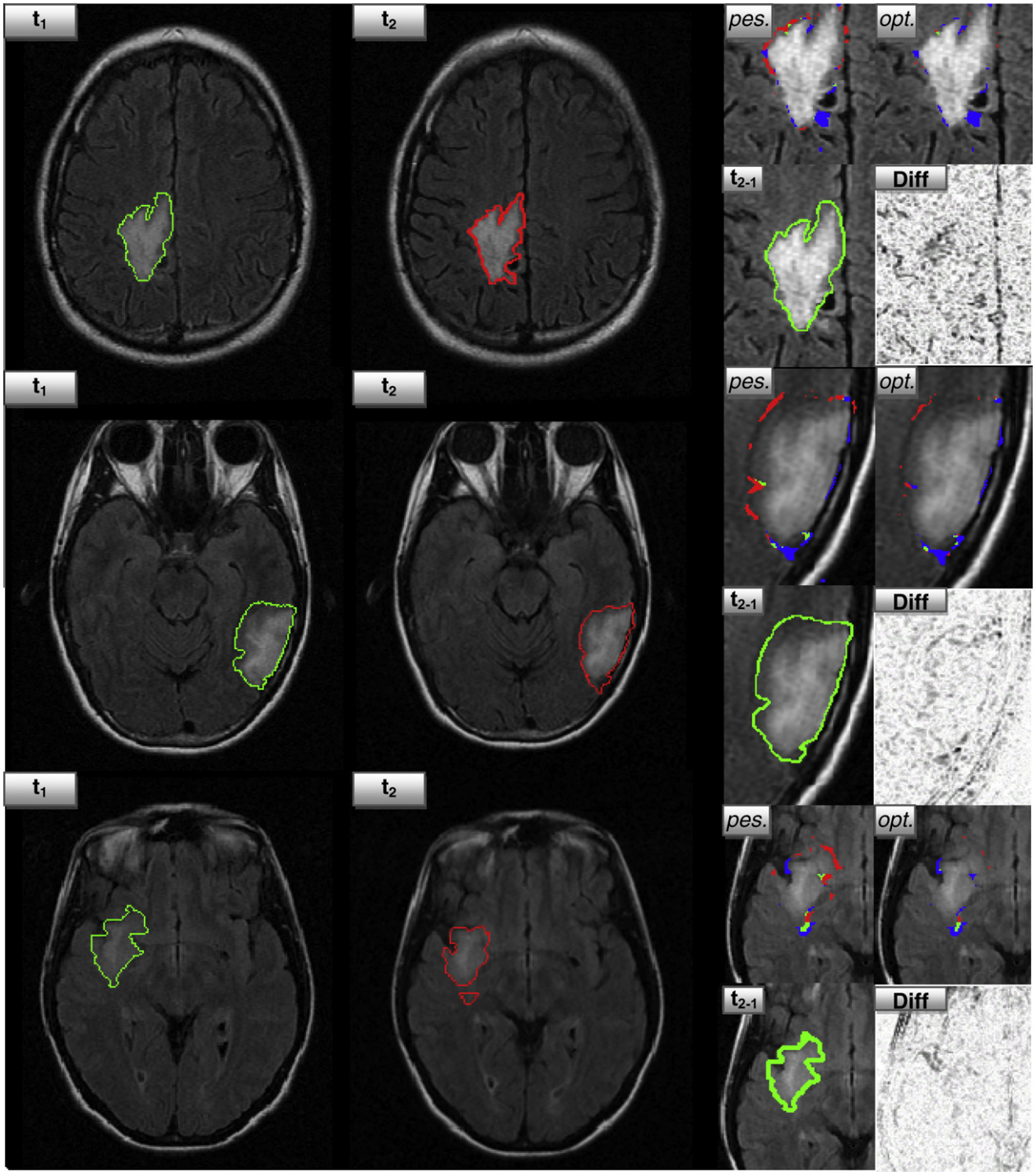

Fig. 6. Illustrations of automated and manual growth estimations on three clinical cases with small progressions, showing: original FLAIR MRI data at times time 1 (t1) and time 2 (t2) with corresponding manual tracings; ROIs of data at time 2 with manual tracing at time 1 (t_{2-1}), midway-based difference map (Diff) and *optimistic* (*opt.*) and *pessimistic* (*pes.*) automated growth overlaps for TP (green), FP (red) and FN (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.5 ± 0.4 mm average precision from the minimal errors between *optimistic* and *pessimistic* estimations, comparable to the 0.8 ± 0.3 mm target precision provided by the manual tracing variability study. Errors higher than 2 mm in *Rs* occurred in 7 cases when either the *optimistic* or *pessimistic* estimation almost perfectly matched the manual estimation. The major difficulty of the neurosurgeon being to evaluate slow progression for these low-grade gliomas, it was very satisfactory to confirm that small or null tumoral progressions were accurately detected by the automated method, with a better precision and robustness with the *Rs* (sphere-equivalent) radius measures.

In Figs. 6 and 7, we illustrate the results of midway-based automated growth tumor detection on respectively small and large growth cases. We included longitudinal midway-mapped FLAIR

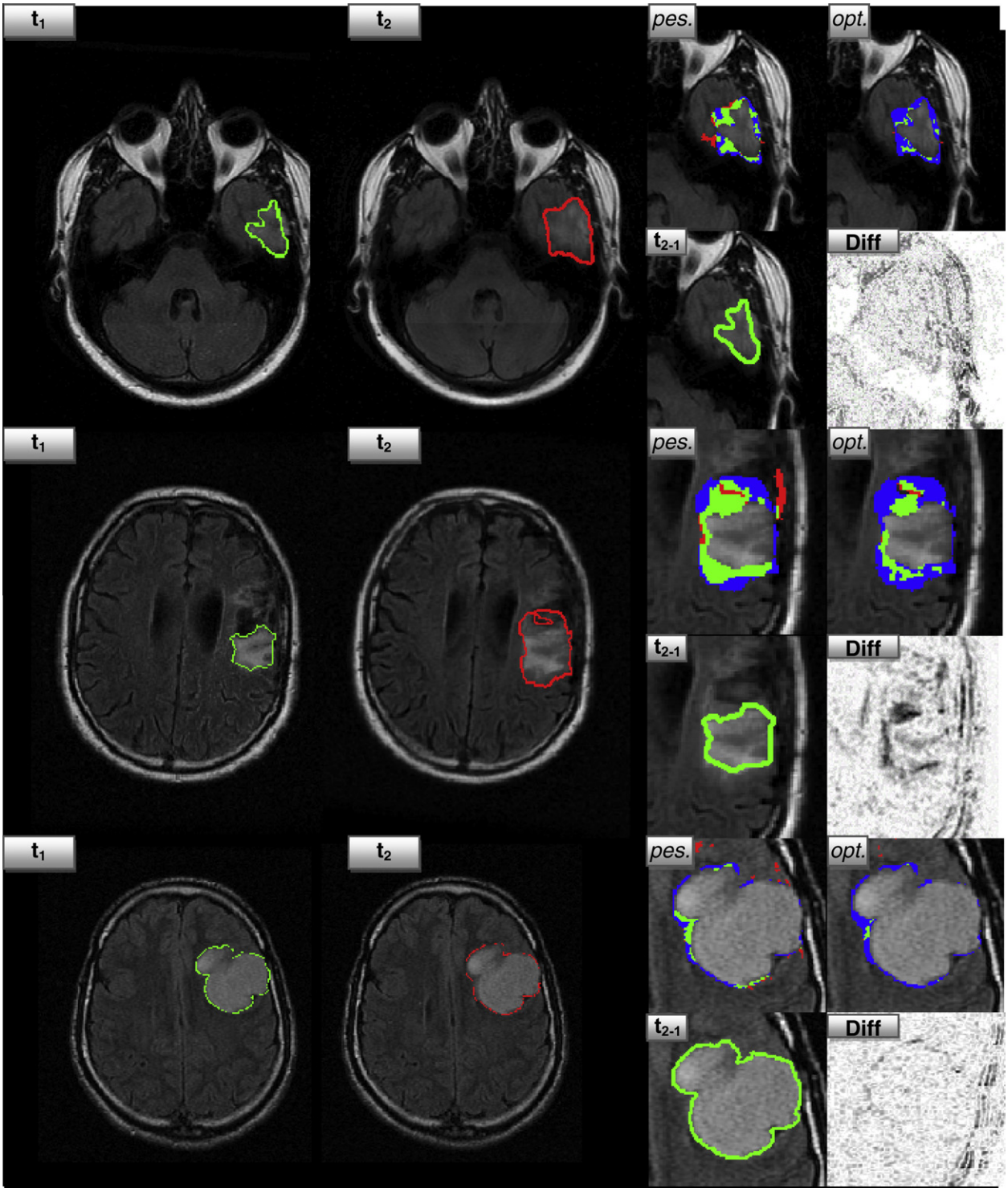

Fig. 7. Illustrations of automated and manual growth estimations on three clinical cases with large progressions, showing: original FLAIR MRI data at times time 1 (t1) and time 2 (t2) with corresponding manual tracings; ROIs of data at time 2 with manual tracing at time 1 (t_{2-1}), midway-based difference map (Diff) and *optimistic* (*opt.*) and *pessimistic* (*pes.*) automated growth overlaps for TP (green), FP (red) and FN (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data along with manual tracing of the tumors, and zoomed views in the ROI around the tumor for: (1) MRI data at time 2 overlaid with the manual contour at time 1 (to visualize potential tumor growth areas as hyperintense MRI signal), (2) midway difference maps (to visualize potential tumor growth areas as large differences), and (3) growth tumor regions, color-coded with respect to true positive (TP) in green, false negative (FN) in blue and false positive (FP) in red, comparing to manual growth estimation. On these illustrations we can observe the limitations of pixel-wise tumor growth comparison due to the registration process. Indeed, manual tracing was performed on non-registered data and the registration process can create some regions from the segmentation at time 1 not included in the segmentation at time 2. These regions could correspond to growth areas detected by the automated method which were not identified by the manual tracing process, since we did not define the notion of regression area, but rather ignored regions segmented at time 1 and not at time 2.

4. Discussion and conclusions

In this study, we presented a novel differential analysis method of longitudinal MRI data with low-grade gliomas. The proposed method combines midway-based MRI intensity normalization, statistical selection of significant differences between the two scans and quantification of tumoral growth. A novel multiplicative midway mapping framework was proposed, based on an image model of MRI data corrupted with local multiplicative inhomogeneity fields. This mapping provided very selective difference maps, insensitive to disparities in the dynamic range of the MRI data sets being compared. Overall, regarding the problem of longitudinal MRI comparison, midway brain MRI mapping provided a very efficient computational framework for gray scale ''standardization'', eliminating the need for tumor segmentation at both times or the use of an image model. Addressing a similar problem, Nyul et al. (2000) proposed a piecewise linear histogram transformation, to map a ''model'' shape defined with landmark points corresponding to percentile levels and median values, whose positions were learned on a training database of MRI images. In our framework, issues related to the use of average landmark points and their unknown ''anatomical'' significance was avoided with global midway mapping transform. Still, some improvement could be gained from investigating the effect of gray-value quantification on the midway-based difference map selectivity. An exact histogram specification method was proposed by Coltuc et al. (2006), ordering all image pixels depending on their gray levels and local averages. In our case though, limitations would arise from high computational cost of 3D pixel ordering and image-dependent pixel ordering not preserving the overall hierarchy of tissue gray-levels, in the two MRI data sets being compared.

Tumor growth quantification with the proposed method was directly dependent on the efficiency of the difference map automated post-processing in selecting ''significant'' difference values. A statistical thresholding framework was proposed. This framework shares some similarities with statistical parametric mapping approaches used in fMRI to detect brain activation sites, reviewed in Marchini and Presanis (2004). With respect to the classification of the methods in this review, our method falls into the type I error class of method. FDR methods were not suited since we had to handle cases with no tumoral evolution, while posterior probability thresholding would have requested the design of a tumoral intensity evolution model in the growth area. Sophistication of the proposed post-processing framework could be investigated, such as the use of weighted kernels for local statistics estimations. Spatial selectivity of the significant differences selection could also be investigated, by combining some aspects of the method proposed on a similar problem by Liu et al. (2003) to select significant changes on SPGR, T2-weighted and FLAIR longitudinal brain MRI scans. Their method used simple difference maps, combined with a ''noise level map'' learned from repeated scans of a single subject. Structured difference images were then generated by thresholding the difference maps and spatially clustering growth features. Significance of the detected changes was assessed by thresholding the difference maps against an anatomical map of artifacts computed from spatial normalization of difference maps from 40 subjects. This final threshold used a probabilistic experimental level of significance. Our proposed method described an alternative computational framework to compare single-protocol repeated scans without the need for integrating image-based noise level evaluation. Indeed the use of global midway mapping and local Gaussian statistics was able to remove from the difference maps random MRI signal variation in repeated scans, due to physiological and scanner related artifacts. Moreover, instead of providing fixed values for the parameters of our statistical test, we proposed two types of parameters setting, corresponding to predictable *optimistic* (small) and *pessimistic* (large) tumoral growth estimations. Learning of the typical large differences due to mis-registration and noise variability (as in the noise level map) could help us eliminate the need for a pre-localization of the tumoral area as the region of interest.

In the case of tumoral growth quantification, some concerns remain regarding the influence of the registration quality on the overall quantification, beyond the influence of the large MRI slice thickness in the particular case of the present study. Brain tumors can induce significant modifications on the surrounding cerebral structures such as the lateral CSF ventricles, due to tumoral mass-effect. There are also inherent physiological changes within the patient's brain between two scan times such as CSF ventricle growth, resorption of the post-surgical cavity or brain shift altering the shape of gyri. These cerebral structures, where we observed the largest errors in overestimating tumoral growth, contribute to the alignment optimization process performed during registration of longitudinal MRI data sets. Some registration works have focused on developing specific registration methods for brain MRI data with tumors (Zacharaki et al., 2008), and on developing image-based tumor growth models (Clatz et al., 2005; Angelini, Clatz et al., 2007). It is very likely that future use of dedicated registration methods or dedicated tumor segmentation and growth model methods could greatly improve the selectivity of the difference maps provided with midway mapping. Nevertheless, registration of longitudinal MRI images of patients with tumors will remain an ill-posed problem, and we plan on studying specifically the spatial distribution and characteristics of large difference values observed on difference maps solely due to registration errors artifacts.

Regarding the large slice thickness of the FLAIR data used in our clinical study, it certainly limited the accuracy of tumoral volume estimations. This limitation led us to study of alternative volumetric radius growth measures, more robust to large slice thickness. The large slice thickness has also driven our choice to perform intensity mapping and statistical testing in 2D. Optimization of the accuracy of the proposed method, which was assessed based on 2D manual delineation of tumoral contours, might have naturally been biased towards 2D processing. It remains therefore unclear if 3D processing could have provided more accurate growth estimates. We expect that, with the advent of 3T MRI, 3D-FLAIR will be widely available in clinical routine, thus circumventing this issue and enable to compare the accuracy of 2D versus 3D tumoral growth estimates.

In conclusion, this paper provided a ''proof-of-concept'' for a new clinical tool to detect and quantify tumor evolution on brain MRI data, evaluated on a large cohort of 32 longitudinal clinical

cases from 13 patients. The proposed method was validated in the context of replicating a tumoral growth quantification task performed in clinical routine via manual tracing on large-thickness FLAIR data. A growth index based on a volumetric radius achieved an average 0.5 ± 0.4 mm precision, which corresponded to the targeted precision provided by the manual tracing variability study. The proposed method was designed to generate two types of tumoral growth estimations: *optimistic* and *pessimistic*. The results have shown that this framework was pertinent for the clinical problem at hand, where in the majority of cases, one of the estimation was highly accurate. The selection of the best estimator therefore seems to be patient-dependent rather than rater-dependent, and will be investigated in our future work. The proposed growth quantification framework will also be evaluated in a near future on a larger series of patients with brain tumors of different types and grades. Such study will evaluate automated tumor growth index measurements as well as potential benefits in limiting manual tracing variability when using the difference maps, after midway mapping. Finally, it is important to note that the method is not specific to the quantification of brain tumor evolution and could be applied to other reviewing tasks on follow-up MRI exams.

Acknowledgment

This study was partially funded by the Institut Telecom and by the French National Cancer Institute (INCA).

References

Angelini, E., Atif, J., Delon, J., Mandonnet, E., et al., 2007. Detection of glioma evolution on longitudinal MRI studies. In: International Symposium on Biomedical Imaging (ISBI), Arlington, VA, USA, pp. 49–52.

Angelini, E., Delon, J., Capelle, L., Mandonnet, E., 2010. Contrast mapping and stastitical testing for low-grade glioma growth quantification on brain MRI. In: IEEE International Symposium on Biomedical Imaging (ISBI), Rotterdam, NL, pp. 872–875.

Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., et al., 2007. Glioma dynamics and computational models: a review of segmentation, registration and in silico growth algorithms and their clinical validations. Current Medical Imaging Review 3 (4), 262–276.

Bosc, M., Heitz, F., Armspach, J.-P., Namer, I., et al., 2003. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. NeuroImage 20 (2), 643–656.

Bromiley, P.A., Thacker, N.A., Courtney, P., 2002. Non-parametric image subtraction using grey level scattergrams. Image and Vision Computing 20 (9–10), 609–617.

Bynevelt, M., Britton, J., Seymour, H., MacSweeney, E., et al., 2001. FLAIR imaging in the follow-up of low-grade gliomas: time to dispense with the dual-echo? Neuroradiology 43 (2), 129–133.

Caseiras, G.B., Ciccareli, O., Altmann, D.R., Benton, C.E., et al., 2009. Low-grade gliomas: six-month tumor growth predicts patient outcome better than admission tumor volume, relative cerebral blood volume, and apparent diffusion coefficient. Radiology 253 (2), 505–512.

Clark, M.C., Lawrence, L.O., Golgof, D.B., Velthuizen, R., et al., 1998. Automatic tumor segmentation using knowledge-based techniques. IEEE Transactions on Medical Imaging 17 (2), 187–201.

Clatz, O., Sermesant, M., Bondiau, P.-Y., Delingette, H., et al., 2005. Realistic simulation of the 3D growth of brain tumors in MR images coupling diffusion with mass effect. IEEE Transactions on Medical Imaging 24 (10), 1334–1346.

Coltuc, D., Bolon, P., Chassery, J.-M., 2006. Exact histogram specification. IEEE Transactions on Image Processing 5 (15), 1143–1152.

Delon, J., 2004. Midway image equalization. Journal of Mathematical Imaging and Vision 21 (2), 119–134.

Gerig, G., Welti, D., Guttmann, C.R.G., Colchester, A.C.F., et al., 2000. Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. Medical Image Analysis 4 (1), 31–42.

Hogea, C.S., Biros, G., Davatzikos, C., 2008. An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. Journal of Mathematical Biology 56 (6), 793–825.

Jager, F., Hornegger, J., 2010. Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE Transactions on Medical Imaging 28 (1), 137–150.

Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., et al., 2000. Automated segmentation of MR images of brain tumors. Radiology 218, 586–591.

Konukoglu, E., Clatz, O., Bondiau, P.-Y., Delingette, H., Ayache, N., 2010a. Extrapolating glioma invasion margin in brain magnetic resonance images: Suggesting new irradiation margins. Medical Image Analysis 14 (2), 111–125.

Konukoglu, E., Clatz, O., Menze, B.H., Stieltjes, B., Weber, M.A., Mandonnet, E., Delingette, H., Ayache, N., 2010b. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic Eikonal equations. IEEE Transactions on Medical Imaging 29 (1), 77–95.

Konukoglu, E., Wells, W. M., Novellas, S., Ayache, N., et al., 2008. Monitoring slowly evolving tumors. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 812–815.

Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction of MR images of the brain. IEEE Transactions on Medical Imaging 18 (10), 885–896.

Lemieux, L., Wieshmann, U.C., Moran, N.F., Fish, D.R., et al., 1998. The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization. Medical Image Analysis 2 (3), 227–242.

Liu, R.S.N., Lemieux, L., Bell, G.S., Sisodiya, S.M., et al., 2003. A longitudinal study of brain morphometrics using quantitative magnetic resonance imaging and difference image analysis. NeuroImage 20 (1), 22–33.

Malvar, H.S., 1992. Signal Processing with Lapped Transforms. Boston, London, Artech House.

Mandonnet, E., Delattre, J.-Y., Tanguy, M.-L., Swanson, K.R., et al., 2003. Continuous growth of mean tumor diameter in a subset of grade II gliomas. Annals of Neurology 53 (4), 524–528.

Mandonnet, E., Pallud, J., Clatz, O., Taillandier, L., et al., 2008. Computational modeling of the WHO grade II glioma dynamics: principles and applications to management paradigm. Neurosurgical Review 31 (3), 263–269.

Mandonnet, E., Pallud, J., Fontaine, D., Taillandier, L., et al., 2010. Inter- and intra-patients comparison of WHO grade II glioma kinetics before and after surgical resection. Neurosurgical Review 33 (1), 91–96.

Marchini, J., Presanis, A., 2004. Comparing methods of analyzing fMRI statistical parametric maps. NeuroImage 22, 1203–1213.

Meier, D.S., Guttmann, C.R.G., 2003. Time-series analysis of MRI intensity patterns in multiple sclerosis. NeuroImage 20 (2), 1193–1209.

Miller, M.I., 2004. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage 23 (Suppl. 1), S19–S33.

Nyul, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19 (2), 143–150.

Pallud, J., Capelle, L., Taillandier, L., Fontaine, D., et al., 2009. Prognostic significance of imaging contrast enhancement for WHO grade II gliomas. Neuro-Oncology 11 (2), 176–182.

Pallud, J., Mandonnet, E., Duffau, H., Kujas, M., et al., 2006. Prognostic value of initial magnetic resonance imaging growth rates for World Health Organization grade II gliomas. Annals of Neurology 3 (60), 380–383.

Pieper, S., Lorensen, B., Schroeder, W., Kikinis, R., 2006. The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community. In: IEEE International Symposium on Biomedical Imaging (ISBI), Arlington, VA, USA, pp. 698–701.

Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., et al., 2003. Automatic brain tumor segmentation by subject specific modifcation of atlas priors. Academic Radiology 10, 1341–1348.

Prima, S., Ayache, N., Janke, A., Francis, S.J., et al., 2002. Statistical analysis of longitudinal MRI data: applications for detection of disease activity in MS. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 363–371.

Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B., 2005. Image change detection algorithms: a systematic survey. IEEE Transactions on Image Processing 14 (3), 294–307.

Rey, D., Subsol, G., Delingette, H., Ayache, N., 2002. Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis. Medical Image Analysis 6, 163–179.

Ricard, D., Kaloshi, G., Amiel-Benouaich, A., Lejeune, J., et al., 2007. Dynamic history of low-grade gliomas before and after temozolomide treatment. Annals of Neurology 61 (5), 484–490.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (S1), 208–219.

Therasse, P., Arbuck, S.G., Eisenhauer, E.A., Wanders, J., et al., 2000. New guidelines to evaluate the response to treatment in solid tumors. Journal of the National Cancer Institute 92 (3), 205–216.

Thompson, P.M., Hayashi, K.M., de Zubicaray, G.I., Janke, A.L., et al., 2004. Mapping hippocampal and ventricular change in Alzheimer disease. NeuroImage 22 (4), 1754–1766.

Tustison, N.J., Avants, B.B., Cook, P.A., Yuanjie, Z., et al., 2010. N4ITK: improved N3 bias correction. IEEE Transactions on Medical Imaging 29 (6), 1310–1320.

Vovk, U., Pernus, F., Likar, B., 2007. A Review of methods for correction of intensity inhomogeneity in MRI. IEEE Transactions on Medical Imaging 26 (3), 405–421.

Xue, Z., Shen, D., Davatzikos, C., 2006. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. NeuroImage 30 (2), 388–399.

Zacharaki, E.I., Shen, D., Lee, S.-K., Davatzikos, C., 2008. Orbit : A multiresolution framework for deformable registration of brain tumor images. IEEE Transactions on Medical Imaging 27 (8), 1003–1017.