

# CONTRAST MAPPING AND STATISTICAL TESTING FOR LOW-GRADE GLIOMA GROWTH QUANTIFICATION ON BRAIN MRI

*Elsa D. Angelini, Julie Delon*

*Laurent Capelle, Emmanuel Mandonnet*

Institut Telecom, Telecom ParisTech  
CNRS, LTCI  
Paris, France

Neurosurgery Department, Hopital Pitie Salpetriere,  
Neurosurgery Department, Hopital Lariboisiere,  
Paris, France

## ABSTRACT

A statistical differential analysis framework of longitudinal MRI volumes is proposed, based on difference maps after non-linear contrast midway mapping, to quantify tumor growth. This mapping was used to normalize MRI scans to a common range of values, and was adapted in this work to handle multiplicative MRI inhomogeneity fields. This led to two direct applications: (1) change detection from a statistical test on differences in midway-mapped MRI data, and (2) tumoral growth quantification. A clinical evaluation was performed on 32 clinical cases with low-grade glioma, screened with two FLAIR MRI scans, several months apart. Three growth indices (volume, maximum radius and spherical radius) were measured and evaluated in terms of accuracy, comparing to manual tracing. Millimetric growth estimation precision was achieved with the proposed method for the spherical radius growth index.

## 1. INTRODUCTION

A differential analysis of longitudinal MRI volumes is proposed, based on non-linear gray value mapping. Image change detection has been getting a lot of attention recently for applications such as remote sensing and video processing. For brain tumor change detection, an interesting approach was proposed by Liu et al [1] to select only significant changes on SPGR, T2-weighted and FLAIR longitudinal brain MRI scans using an elaborated processing framework on direct differences maps. This processing involved noise level map estimations from repeated scans on a single subject, thresholding of the difference maps above noise level and filtering of the remaining values with a learned anatomical map of artifacts (computed from spatial normalization of difference maps from 40 subjects). In the context of this study, our method avoids the needs of repeated scans and offers a framework likely to be more easily reproducible, consistent across different clinical studies, and robust to image noise. Recent studies have also focused on registration and segmentation-based longitudinal MRI brain comparison. For brain tumor change detection, a study in [2], evaluated an automated tumor segmentation performed on longitudinal series of four exams for patients with glioblastomas multiforme. The study reported very high true positive values when comparing tumor volumes to manual tracing, but also observed errors in

growth or shrinkage predictions based on tumor volume measurements from the segmented data. A study in [3] compared direct image comparison to a registration-based method, for volume growth quantification of brain meningiomas. The semi-automated methods were reasonably accurate and less impacted by intra and inter-rater manual segmentation variability than the human expert rating of growth rate. Indeed, brain tumor segmentation remains a challenging task, to include necrosis, edema, and infiltrations, with inter-observer variability for manual tracing of brain tumor that can range up to 15% [4]. It is therefore interesting to investigate tumor growth measurement methods, independent of sequential tumor segmentation.

## 2. METHOD

### 2.1. Midway Mapping

#### 2.1.1. Invariance to contrast change

The initial midway mapping framework was derived to map images generated from a common image, but modified non-linearly by different contrast changes. The fundamental idea, to enable comparison of these images, was to map them to a common histogram, that defined a reasonable intermediate gray scale distribution. This notion of a "reasonable" midway gray scale distribution was derived empirically, in the original work of Delon [5], by enforcing the following mapping rules for two test cases: (1) The midway mapping between image  $I$  and image  $I + \lambda$  should be  $I + \lambda/2$ ; (2) The midway mapping between  $f(I)$  and  $g(I)$  should be defined independently of the contrast change functions  $f$  and  $g$ , for a given image  $I$ . These test cases lead to the definition of the midway cumulative histogram as the inverse of the arithmetic mean of the two inverse cumulative histograms of the images to map. In [6] a first application of the midway mapping framework was presented to compare longitudinal FLAIR and SPGR MRI data of 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 their work, and are discussed in the following paragraphs.

#### 2.1.2. Invariance to MRI inhomogeneity

MRI image artefacts due to spatial inhomogeneities of the magnetic field  $g(x)$  are typically modeled as multiplicative

spatial fields:  $I_g(x) = I(x) \times g(x)$  [7]. This field is non-uniform in space but slowly varying. In our framework, we can model  $g_i(x)$  as a constant multiplier  $\lambda$  and apply midway image normalization on small overlapping patches. With such contrast change model, it is desirable to work with a new test case, forcing the mapping of  $\lambda I$  and  $1/\lambda I$  to be  $I$ .

This approach corresponds to defining the midway target histogram as the inverse of the geometric mean of the inverse cumulative histograms rather than an arithmetic mean as proposed in the initial framework.

To validate the capability of the multiplicative midway model to handle image differences due to spatially varying inhomogeneity fields, we generated an example for a pair of images corresponding to the test case  $(\lambda I, I/\lambda)$ , using a piecewise constant synthetic inhomogeneity field as illustrated in Figure 1. After midway mapping, an estimated inhomogeneity field was computed via the ratio of the  $\lambda I$  image and its mapped image, referring to the fact that the theoretical target common image was  $\sqrt{(\lambda I \times 1/\lambda I)} = I$ . Since the inhomogeneity field was spatially varying, we applied midway mapping on overlapping patches of  $[10 \times 10]$  pixels. Additive and multiplicative midway mapping were tested. Ideally, the difference map computed between the pair of images after midway-mapping should display very little structural information, since the images only differ by a smooth multiplicative inhomogeneity field. Results illustrated in Figure 1 showed that only the patch-based implementation of multiplicative midway mapping was able to provide difference maps with few anatomical structural information, and limited ranges of gray level variations, as desired. In addition, we observed that the inhomogeneity field spatial appearance was accurately estimated after patch-based midway mapping, and only the multiplicative model provided accurate numerical values with isolated errors.

With this example we were therefore able to confirm the ability of the patch-based multiplicative-midway mapping to cancel contrast differences due to multiplicative inhomogeneity fields and even provide an accurate estimation of a multiplicative inhomogeneity field corrupting an original MRI image. 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.

## 2.2. Statistical Thresholding of Difference Maps

In this paragraph, we propose a generic method to identify structural differences between the two longitudinal MRI acquisitions on difference maps, computed after midway mapping. We assume at this stage that we are given two images  $I_1$  and  $I_2$ , normalized with midway mapping and registered. For each voxel  $i$ , we wanted to know if the difference  $D(i) = (I_1(i) - I_2(i))$  was meaningful, *i.e.* due to a struc-

tural difference, or rather explained by the local statistics of the images. To answer this question, we relied on a statistical hypothesis testing framework. Let us denote by  $\mathcal{V}_i$  the  $(2f+1) \times (2f+1)$  centered neighborhood of the voxel  $i$ . We note  $n = (2f+1)^2$ . We defined the following *null hypothesis*  $\mathcal{H}_0^i$ : *the random variables  $\mathbf{D}(j)$ ,  $j \in \mathcal{V}_i$ , are i.i.d. and follow a centered normal distribution  $\mathcal{N}(0, \sigma_i^2)$ .* Under the hypothesis  $\mathcal{H}_0^i$ , the random variable  $\overline{\mathbf{D}}(i) = \frac{1}{n} \sum_{j \in \mathcal{V}_i} \mathbf{D}(j)$  follows a normal distribution  $\mathcal{N}(0, \sigma_i^2/n)$ . Thus, the probability that the average distance  $|\overline{\mathbf{D}}(i)|$  was larger than a given value  $\delta$  was:

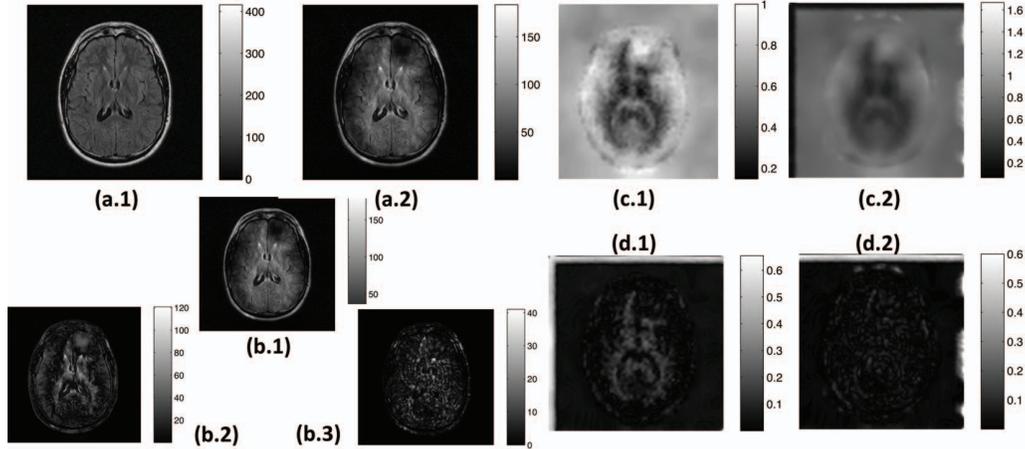
$$F_i[\delta] := P_{\mathcal{H}_0^i} [|\overline{\mathbf{D}}(i)| > \delta] = 1 - \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_i} \int_{-\delta}^{\delta} e^{-\frac{nx^2}{2\sigma_i^2}} dx. \quad (1)$$

In practice, we estimated the local standard deviation  $\sigma_i$  on  $\mathcal{V}'_i$  a  $(2t+1) \times (2t+1)$  neighborhood of  $i$ . Knowing this standard deviation, we were able to compute the distribution  $F_i[\delta]$  for any voxel  $i$  and any value of  $\delta$ . We decided to reject the hypothesis  $\mathcal{H}_0^i$  as soon as  $F_i[|\overline{\mathbf{D}}(i)|]$  was small enough, *i.e.* below a threshold value  $\alpha$ . In other words, if the mean of  $I_1$  around  $i$  was too different from the mean of  $I_2$  in the same region, the hypothesis  $\mathcal{H}_0^i$  was rejected and we considered that there was a structural difference between both images at pixel location  $i$ . This boils down to detect a structural difference as soon as  $|\overline{\mathbf{D}}(i)| > \lambda\sigma_i/\sqrt{n}$ , where  $\lambda > 0$  is a quantile of the standard normal distribution.

This simple thresholding scheme enabled us to automatically select significant changes in the difference maps, with a fixed threshold level  $\alpha$  but with two parameters controlling average window sizes  $\mathcal{V}_i$  and  $\mathcal{V}'_i$  for local mean and variance computations. These two parameters provided a very convenient framework to control the tolerance of the difference selection (in terms of levels of significance) which correlate well with an *optimistic* or *pessimistic* estimation of tumoral growth measurements, when  $(I_1, I_2)$  correspond to longitudinal MRI scans from a patient harbouring a low-grade glioma. Indeed, increasing the window size for  $\mathcal{V}_i$  decreased the difference levels, while increasing the window size for  $\mathcal{V}'_i$  increased the variance values used to normalize the differences. For example, using the sequence of averaging window sizes of  $[\mathcal{V}_i \mathcal{V}'_i]$ :  $[3 \ 5] \rightarrow [3 \ 3] \rightarrow [5 \ 5] \rightarrow [5 \ 3]$  lead to larger growth areas. To evaluate the proposed methodology, we needed to define a growth mask around the tumor at  $t_1$ . This selection was performed using a manual outline of the tumor at  $t_1$ , from a clinical expert. Obtaining an automated detection of the tumor's contours (at either  $t_1$  or  $t_2$ ) was beyond the scope of this study and still represents a very challenging task, as reviewed in [8].

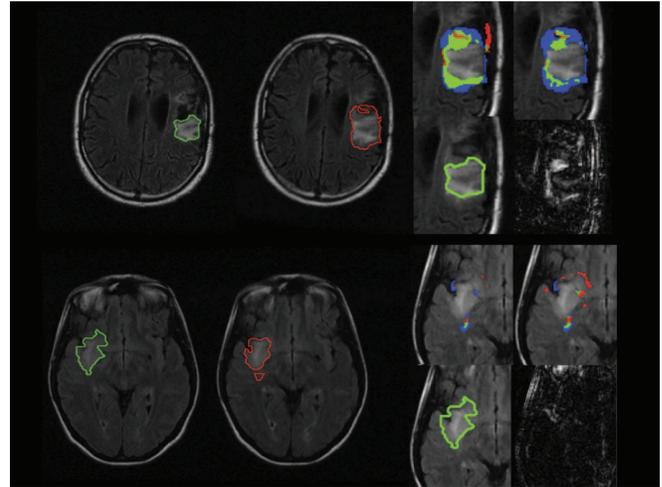
## 3. RESULTS

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. Seven patient cases included only two consecutive scans, and the remaining



**Fig. 1.** Correction of inhomogeneity field contrast changes on a simulated case: (a) *Images to map*: (a.1) Original MRI, (a.2) MRI corrupted with the multiplicative inhomogeneity field in (c.1); (b) *Differences of images*: (b.1) before mapping, (b.2-3) after multiplicative midway mapping [(b.2) global or (b.3) patch-based implementations]; (c) *Inhomogeneity fields*: (c.1) original and (c.2) estimated after patch-based multiplicative midway mapping; (d) *Errors on inhomogeneity field estimation*: (d.1) from patch-based multiplicative midway mapping and (d.2) from global multiplicative midway mapping.

six cases included 4, 7 or 9 longitudinal scans. These data sets were of dimension either  $[512 \times 512]$  or  $[256 \times 256]$  in axial slices (spatial resolution of  $[0.5 \times 0.5]$  or  $[0.9 \times 0.9]$  mm) and included 20 slices (corresponding to a standard FLAIR slice thickness of 6.5mm). All data sets were manually traced for tumor delineation by an expert neurosurgeon, without midway mapping. All experiments reported in this section were performed on pre-surgical 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 (<http://www.fmrib.ox.ac.uk/fsl/>), using correlation similarity measures and a similitude transform with 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: (1-2) Growth of the tumors volumes  $V$  in milliliters (ml) and 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  $R_s$ , defined as the radius of the sphere with the same volume as the tumor (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 [9]. In these experiments, we set  $\alpha$  to an arbitrary small value, ( $10^{-10}$ ), smaller than the chance of picking up randomly one pixel in the entire volume data set (typical axial volume dimensions were  $[256 \times 256]$  or  $[512 \times 512]$  and 20 slices). With this fixed  $\alpha$  value, we used  $[\mathcal{V}_i \mathcal{V}'_i] = [9 \ 3]$  and  $[3 \ 3]$  for small and large growth measurements. The use of a fixed  $\alpha$  value guaranteed that we never selected all pixels in the growth region of interest, even though we increased or decreased the growth area when vary-



**Fig. 2.** Illustrations of automated and manual growth estimations on 2 clinical cases with large (top) and small (bottom) tumoral progressions, showing: original FLAIR MRI data at times  $t_1$  (left) and  $t_2$  (right) with corresponding manual tracings; ROIs of data at time 2 with manual tracing at time 1, midway-based difference map and *optimistic* and *pessimistic* automated growth overlaps for TP (green), FP (red) and FN (blue).

ing the statistical neighboring sizes. In Figure 2, we illustrate the similarities and differences between the automated large and small growth estimations and manual estimations, for two clinical cases. Overlap areas (manual growth being the reference) for true-positive (TP), false-negative (FN) and false-positive (FP) measures, are color coded in the figure. In Table 1 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 can observe that the average growth volume error, lower than 8ml, is rather satisfactory, given that

**Table 1.** Average errors of tumor growth estimations, with four index measures, between automated and manual methods, for *pessimistic* (large) and *optimistic* (small) statistical setups.

|         | Tumor Growth V (ml) |       | Tumor Growth V (%) |       | Tumor Growth R (mm) |       | Tumor Growth Rs (mm) |       |
|---------|---------------------|-------|--------------------|-------|---------------------|-------|----------------------|-------|
|         | large               | small | large              | small | large               | small | large                | small |
| Average | 8.3                 | 5.1   | 22.3               | 14.3  | 2.6                 | 1.9   | 1.2                  | 0.8   |
| Max     | 32.1                | 21.4  | 62.3               | 67.4  | 13.1                | 8.4   | 3.1                  | 2.8   |
| Min     | 0.1                 | 0.0   | 0.0                | 0.1   | 0.0                 | 0.0   | 0.0                  | 0.0   |

the average tumoral volume was 48ml (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 measures provided, on average, highly accurate measures when comparing automated and manual estimations, between 1 to 3 mm. The R (maximum radius) measure generated some large errors in few cases but the histogram showed that the majority of cases provided errors below 5mm. The Rs (sphere-equivalent) radius reported much smaller maximum errors and the histogram plots showed a very peaked distribution for the *optimistic* Rs measure, confirming that the proposed method was 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 (studied in [4]) in the case of low-resolution FLAIR MRI data. On the other hand the radius-based growth measures provided the targeted 2mm average precision, given the resolution of the MRI FLAIR data. Errors higher than 2mm 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 of 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.

#### 4. DISCUSSION & CONCLUSION

In this paper, we have proposed a modification of the midway mapping framework to handle multiplicative contrast change, and a statistical framework to detect significant changes on difference maps, between two longitudinal FLAIR MRI scans, in the context of brain tumor growth monitoring. Synthetic and clinical experiments confirmed that multiplicative

midway contrast mapping was a powerful tool suited for MRI normalization and artefact correction such as magnetic field inhomogeneity or contrast change. Despite sharing some similarities with other works, using the general concept of MRI normalization via histogram mapping, our proposed method does not rely on a training set to define the target histogram shape, but rather applies a global mapping between two images. It is interesting to note that midway mapping does not necessarily correct image artefacts (as targeted when using a model histogram) but makes these artefacts similar and comparable in a pair of images. Clinical results confirmed the feasibility and accuracy of tumor growth estimation on difference maps when compared to manual tracing, with flexible parameters controlling the notion of *optimistic* or *pessimistic* growth measurement. Computational time was in the order of a minute for registration, midway mapping, and post-processing for growth measurement with the statistical tests. Finally, we point out that avoiding training for histogram learning enables our method to be applied, in a generic fashion, to any type of registered longitudinal MRI of any body part.

#### 5. REFERENCES

- [1] R. S. N. Liu, L. Lemieux, G. S. Bell, S. M. SisoDiyi, S. D. Shorvon, J. W. A. S. Sander, and J. S. Duncan, "A longitudinal study of brain morphometrics using quantitative magnetic resonance imaging and difference image analysis," *NeuroImage*, vol. 20, pp. 22–33, 2003.
- [2] L. C. Clark, L. Hall, D. B. Goldgof, R. Velthuizen, F. R. Murtagh, and M. S. Silbiger, "Automatic tumor segmentation using knowledge-based techniques," *IEEE Transactions on Medical Imaging*, vol. 17(2), pp. 187–201, 1998.
- [3] E. Konukoglu, W. M. Wells, S. Novellas, N. Ayache, R. Kikinis, P. M. Black, and K. M. Pohl, "Monitoring slowly evolving tumors," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, Paris, France, 2008, pp. 812 – 815.
- [4] M. R. Kaus, S. K. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, "Automated segmentation of MR images of brain tumors," *Radiology*, vol. 218, pp. 586–591, 2001.
- [5] J. Delon, "Midway image equalization," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 2, pp. 119–134, 2004.
- [6] E. Angelini, J. Atif, J. Delon, E. Mandonnet, H. Duffau, and L. Capelle, "Detection of glioma evolution on longitudinal MRI studies," in *International Symposium on Biomedical Imaging (ISBI)*, Arlington, VA, USA, 2007, pp. 49–52.
- [7] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 405–421, 2007.
- [8] E. D. Angelini, O. Clatz, E. Mandonnet, E. Konukoglu, L. Capelle, and H. Duffau, "Glioma dynamics and computational models: A review of segmentation, registration and in silico growth algorithms and their clinical validations," *Current Medical Imaging Review*, vol. 3, no. 4, pp. 262–276, 2007.
- [9] E. Mandonnet, J.-Y. Delattre, M.-L. Tanguy, K. R. Swanson, A. F. Carpentier, H. Duffau, P. Cornu, R. Van Effenterre, E. C. Alford, and L. Capelle, "Continuous growth of mean tumor diameter in a subset of grade II gliomas," *Ann. Neurol.*, vol. 53, pp. 524–528, 2003.