MONITORING SLOWLY EVOLVING TUMORS

E. Konukoglu[1]*, W. M. Wells*[2]*, S. Novellas*[1]*, N. Ayache*[1]*, R. Kikinis*[2]*, P. M. Black*[2]*, K. M. Pohl*[2]

[1] Asclepios Research Project, INRIA
Sophia Antipolis , France
{ekonukog,snovella,na}@sophia.inria.fr

[2] Brigham & Women's Hospital, Boston, MA
{sw,kikinis,pohl}@bwh.harvard.edu
pblack@partners.org

ABSTRACT

Change detection is a critical task in the diagnosis of many slowly evolving pathologies. This paper describes an approach that semi-automatically performs this task using longitudinal medical images. We are specifically interested in meningiomas, which experts often find difficult to monitor as the tumor evolution can be obscured by image artifacts. We test the method on synthetic data with known tumor growth as well as ten clinical data sets. We show that the results of our approach highly correlate with expert findings but seem to be less impacted by inter- and intra-rater variability.

***Index Terms*—** tumor, follow-up, time series analysis

1. INTRODUCTION

Meningiomas are the most common type of primary brain tumor. Most of these tumors are categorized as benign pathology that grows slowly between brain tissue and dura. To avoid the risk of surgery, neurosurgeons carefully monitor patients with benign meningiomas by having the patient regularly undergo Magnetic Resonance (MR) scanning. An expert assess the tumor growth through visual inspection of consecutive 3D scans. A precise analysis, however, is extremely difficult as slow growth is often obscured by changes in the head position or intensity profile between the two scans. We address this issue by describing a relatively fast and robust method that semi-automatically analyzes the tumor evolution.

One can compute volume change by separately segmenting each scan in the time series via automatic tumor segmenters, such as [3, 4]. This type of analysis, however, is very sensitive to intra-rater variability. Rey *et al.* [1] therefore proposed a pipeline consisting of non-rigidly aligning brain scans of the same patient and then analyzing the resulting deformations. The approach is tested on multiple sclerosis lesions where the growth is clearly visible. The method by Angelini *et al.* [2] rigidly aligns the sequences of images, normalizes their intensities, and then computes difference maps between them. They apply the approach to low grade gliomas, where the growth is above 30% and also clearly visible. We are aware of the incompleteness of this review. Nevertheless, we note that most automatic methods for monitoring slowly evolving pathology focus on visible growth.

To the best of our knowledge, our approach is the first in medical vision to automatically measuring growth of brain tumors that is difficult to determine via visual inspection. Our approach first semi-automatically segments the tumor in the initial patient scan. It then aligns the second scan of the patient to the first using a hierarchical rigid registration approach. Finally, it measures growth or shrinkage from these images, for which we suggest two different types of metric. Motivated by [2], the first metric detects change through analyzing differences in intensity distributions. Unlike [2], we relate the analysis to hypothesis testing. The second metric is motivated by the work of [1], which detects change by analyzing the deformation field between two scans. Unlike [1], our metric returns a quantitative analysis of the differences in volume (mm^3). Our approach also differs significantly from works on segmentation, such as [3, 4], as we estimate volume change by simultaneously analyzing the sequence of scans.

The approach is described in detail in Section 2 and the accuracy of the method is evaluated in Section 3. We first show on synthetic data that our analysis is more reliable than visual inspection. We then apply our method to 10 clinical MR scans and compare them again to expert determinations.

2. SEMI-AUTOMATIC CHANGE DETECTION

The software pipeline is composed of three steps, which are: tumor segmentation, image registration, and change detection. The first step semi-automatically identifies the tumor in the first scan only. In order to avoid issues of intra-rater variability, we use manual supervision in one time point only. We also prefer semi-automatic over automatic segmentation as the accuracy of semi-automatic approaches in identifying pathology is generally higher. Our pipeline relies on accurate segmentation in order to detect changes of a few voxels in pathology. A more in depth discussion of current segmentation technology is outside the scope of this paper.

The semi-automatic segmentation is based on a user-defined bounding box (BB) around the tumor and a lower bound of the intensities that characterize the pathology. From these indicators, the pipeline can reliably extract most of the pathology because meningioma generally show up on post-gadolinium, T1-weighted MR images as homogeneous, bright objects [5]. The pipeline post-processes the resulting binary map by removing small islands and holes caused by

This work has been partly supported by the European Health-e-Child project (IST-2004-027749), the CompuTumor associated team, the Brain Science Foundation and the NIH grants P41RR13218, U54EB005149, and U41RR019703. We also would like to thank Dr. Zamani, Department of Radiology, Brigham and Women's Hospital, for scanning the patients.

the noise in the MRs. We note that the resulting map $\mathcal{I}_{seg}^{(1)}$ will also include part of the dura and vessels since these structures have intensity patterns that are similar to pathology. However, because they are static, these additional structures should not substantially impact the analysis.

The second step automatically aligns the pathology of the second scan to the first. It does so by rigidly registering [6] the scans. Now, we assume that the previous BB is large enough so that its coordinates also define a BB around the pathology in the second scan. We then accommodate partial voluming in images by increasing the resolution of both BBs. Finally, the framework addresses non-linear perturbation artifacts caused by the MR acquisition by rigidly aligning the contents of the BBs with each other. This results in two images where, in theory, barring temporal changes, the pathology is well aligned.

The final step of our approach measures the tumor evolution based on the initial segmentation and the BBs described above. We now propose three metrics for detecting change.

2.1. Statistical Analysis of Change

The first metric is based on a statistical model that differentiates changes in the intensity patterns that are due to evolving pathology rather than image artifacts. From this model, we can infer a label map $\mathcal{T}$, where each voxel in the image $\mathbb{I}$ is labeled as dormant or non-modified tissue (*N*), or growing (*G*) or shrinking (*S*) pathology. The change in tumor size is the difference in the number of voxels labeled as *G* versus *S*.

Let $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ be the images resulting from the second step and $\mathcal{I}^{(abs)} \triangleq abs(\mathcal{I}^{(1)} - \mathcal{I}^{(2)})$ the absolute difference between the two intensities. In general, $\mathcal{I}^{(abs)}$ is much larger at regions of changing pathologies (*G* and *S*) versus dormant tissue *N*. Thus, we can compute $\mathcal{T}$ more easily if we know the threshold that separates those two regions in $\mathcal{I}^{(abs)}$. This threshold can be estimated via a hypothesis test.

The test is specified by the null hypothesis that image artifacts caused the changes in intensity, and the false positive, which is the error for rejecting the null hypothesis given that it is actually true. Furthermore, the training data $\mathbb{N}$ is composed of voxels of the image domain $\mathbb{I}$ certainly assigned to *N* and the test data $\overline{\mathbb{N}} \triangleq \mathbb{I} \backslash \mathbb{N}$ are the remaining voxels in the image. If we now assume that both images are some what well aligned and the evolution of pathology is fairly small then $\mathbb{N}$ can be simply estimated from $\mathcal{I}_{seg}^{(1)}$. We then approximate $P(\mathcal{I}_x^{(abs)} | \mathcal{T}_x = N)$, the probability of intensity differences due to image artifacts at voxel x, by the normalized histogram of $\mathcal{I}^{(abs)}$ confined to $\mathbb{N}$. The false positive is denoted as $s \in [0,1]$, where a voxel $x \in \overline{\mathbb{N}}$ is assigned to *N* if $\mathcal{I}_x^{(abs)} < \sigma^{-1}(s)$ with the inverse of the cumulative function defined as $\sigma^{-1}(s) \triangleq \arg min_l \{P(\mathcal{I}_x^{(abs)} < l | \mathcal{T}_x = N) = s\}$.

We know that $\mathcal{I}_x^{(1)} \gg \mathcal{I}_x^{(2)}$ indicates growth and $\mathcal{I}_x^{(1)} \ll \mathcal{I}_x^{(2)}$ shrinkage. Thus, the map $\mathcal{T}$ at voxel $x \in \mathbb{I}$ is defined as

$$\mathcal{T}_x(s) \triangleq \begin{cases} G\,,\ \mathcal{I}_x^{(2)} > \mathcal{I}_x^{(1)} + \sigma^{-1}(s)\ \wedge\ x \in \overline{\mathbb{N}} \\ S\,,\ \mathcal{I}_x^{(1)} > \mathcal{I}_x^{(2)} + \sigma^{-1}(s)\ \wedge\ x \in \overline{\mathbb{N}} \\ N,\ \text{ otherwise,} \end{cases}$$

and $V_{sta}(s) \triangleq \sum_{x \in \mathbb{I}} (\mathcal{T}_x(s) \& G) - (\mathcal{T}_x(s) \& S)$ captures change.

2.2. Deformable Model Based Analysis

The second type of metric measures the tumor evolution by analyzing the deformation field $D : \mathcal{I}^{(1)} \mapsto \mathcal{I}^{(2)}$ that maps the first BB to the second one. We further confine the analysis to the region $\mathbb{L}$ defined by the segmentation of the first step to omit any information in $D(\cdot)$ unrelated to the tumor.

We determine $D(\cdot)$ by the algorithm explained in [7] that is based on the well known demons method. This non-rigid registration algorithm computes $D(\cdot)$ by finding the *diffeomorphic* transformation that best aligns the intensity information of the first BB to the second. Any evolving tumor tissue causes differences in the local intensity pattern of the scans that are captured by $D(\cdot)$ as the registration algorithm brings the images into agreement. The diffeomorphic constraint further confines $D(\cdot)$ to be smooth so that artifacts due to noise are reduced. While other approaches are well suited for this task, a proper review is outside the scope of this paper.

After computing $D(\cdot)$, we can choose from a variety of methods to determine $V^{(2)}$, the volume of pathology in the second scan. Here, we introduce $V_{jac}^{(2)}$, a metric based Jacobian determinants and $V_{def}^{(2)}$, which measures the volume by applying the deformation field to the binary map $\mathcal{I}_{seg}^{(1)}$:

$$V_{jac}^{(2)} \triangleq \sum\nolimits_{x \in \mathbb{L}} |J_D(x)| \quad (1)$$

$$V_{def}^{(2)} \triangleq \sum\nolimits_{x \in \mathbb{I}} (\mathcal{I}_{seg}^{(1)} \circ D)(x). \quad (2)$$

$|J_D(x)|$ denotes the Jacobian determinant of $D(\cdot)$ at voxel x. $V_{reg} \triangleq V^{(2)} - V^{(1)}$, the volumetric difference between the pathology in the second $V^{(2)}$ and first scan $V^{(1)} \triangleq \sum_x \mathcal{I}_{seg}^{(1)}(x)$, quantifies the evolution of the pathology.

We provide an implementation of the methodology as part of the 3D Slicer (www.slicer.org). The semi-automatic analysis is easy to calibrate and completes in less than five minutes returning the total area of tumor change in mm^3.

3. TESTING OUR METRICS

In the following we characterize our methods in two experiments. We first measure the accuracy of our approach using synthetic tumor data, where we artificially enlarge the meningioma in the MR scan of a single patient. The synthetic data is also visually inspected by an expert to test the accuracy of human raters. The second experiment uses a series of MR scans from 9 patients with meningiomas. The same expert labels the series of scans as tumor growing or dormant. Our tool is then applied to data and compared to manual findings.

Both experiments are based on clinical MR scans that were acquired through the Department of Radiology, Brigham and Women's Hospital, Boston under normal hospital environment using a 1.5 T General Electric Scanner (axial slice direction, post-gadolinium, matrix: 256×256×130, voxel dimension: 0.9375 mm × 0.9375 mm × 1.2 mm).

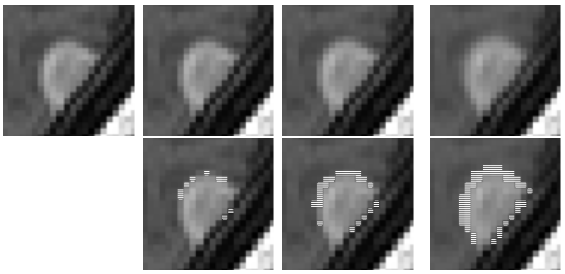

Fig. 1. The real scan of the patient with synthetically evolving pathology of 1%, 5% and 22% volume growth. Notice, the similarities of images with 1% and 5% growth to the real one.

True (%)	INTENSITY (avg±std)	JACOBIAN (avg±std)	SEGMENT (avg±std)	EXPERT (Ratio)
1.1	0.14 ± 0.13	3.5 ± 1.11	0.6 ± 0.94	0/5
5.4	2.2 ± 0.48	7.0 ± 1.22	4.4 ± 1.30	1/5
22.0	13.6 ± 0.32	21.0 ± 0.80	19.0 ± 0.87	5/5

Table 1. The mean and standard deviation of the three metrics (INTENSITY, JACOBIAN, SEGMENT) in detecting the growth (True) and the success ratio of a clinician (EXPERT).

3.1. Synthetic Example

We quantitatively assess our approach by analyzing its results on a synthetic data set, where the amount of growth between two scans is known. Furthermore, we show the difficulties of this task for humans by having a radiologist detect growth through visual inspection of the scans.

We construct our test data by virtually growing the pathology in a brain MR scan of a meningioma patient (see Figure 3.1). The growth is simulated via a simple mechanical model, which performs homogeneous expansion by deforming the surrounding tissue around the tumor while keeping the skull fixed [8]. The model also assumes that the enhancing agent uptake of the grown tumor is similar to the initial scan.

We use the simulator to create eight different tumor scans ranging from 1% to 22% growth in volume of the pathology. We then apply ten randomly generated rigid transformations to each of the scans that simulate changes in head positions between consecutive scans. This results in 80 test data sets in which the exact tumor growth is known.

Figure 2 and Table 1 summarize the experiment. INTENSITY corresponds to results of the statistically motivated metric of Section 2.1, while the entries of JACOBIAN (Equation 1) and SEGMENT (Equation 2) analyze the deformation field as described in Section 2.2. The mean and standard deviation of each metric is based on the corresponding ten measurements for each true growth value (True). In addition, the table shows the success ratio of a radiologist in correctly identifying growth. To determine this ratio, the radiologist looked at 15 of 80 scans to decide whether the tumor was dormant or growing in correspondence to the first scan.

In this experiment, all three metrics successfully detect even small growth. Although, in the 1% case the INTENSITY and the SEGMENT analyses show intervals of confidence including zero growth in the Table 1, when we look at the individual cases they both detect growth for 7 out of 10 cases. Based on the mean value, SEGMENT most accurately recovers the true growth (represented by the dotted line) in most cases. INTENSITY seems to be least impacted by the head position as shown by the relatively small standard deviation. Both INTENSITY and SEGMENT generally underestimate the growth and all three metrics are significantly different from each other. Furthermore, the very low success rate of the radiologist for 1% and 5% growth indicates the difficulties in visually detecting small growth. The results suggests that our tool successfully detects growth even when visual inspection is very difficult.

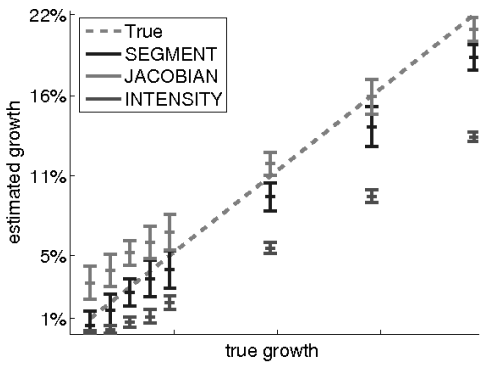

Fig. 2. The results of the three metrics based on synthetic data set where the growth in known (dotted line). The x-axis represents the true growth while the y-axis corresponds to the growth computed by the metrics.

3.2. Real Images

In the second experiment, we apply our tool to MR scans taken from nine patients with slowly evolving meningiomas. Eight out of nine data sets consisted of two consecutive scans while one contained three scans. In addition to the automatic measurements, a radiologist determined the volume growth by manually segmenting the tumor in each scan.

The result of this experiment are summarized in Table 2. Based on the manual measurements, the data set consists mostly of cases with less than 100 mm^3 growth, which is generally very difficult to detect. The automatic measurements generally agree with the manual ones in all but two cases (Case 4 and Case 5). Figures 3 show a sample slice of the first and second scan of Case 4. The intensity pattern of the tumor varies considerably between the acquisitions, which is unusual from our experience, and violates one of the underlying assumptions of the pipeline. These intensity differences may be caused by necrosis or fibrosis, which are very rare in meningioma [5], or changes in the acquisition protocol, which seems to be the case here. The inconsistencies observed in Case 5 are to due to the relatively small tumor volume, which is 334 mm^3 (270 voxels). For such small tumors, distinguishing growth from interpolation artifacts is relatively difficult.

Case	EXPERT		INTENSITY		JACOBIAN		SEGMENT	
	%	mm^3	%	mm^3	%	mm^3	%	mm^3
1	4.8	596	9.7	1202	3.6	447	3.5	433
2	5.4	883	4.2	677	8.16	1325	9.14	1483
3	-0.4	-2.5	0.7	3.9	-1.1	-6.3	-2.3	-13
4	-7.3	-743	2.9	299	16.2	1660	20.1	2063
5	-23.3	-77	17.1	57	38.1	127	36.7	122
6	-6.3	-323	0.26	13.3	-11.8	-608	-13.7	-706
7	0.1	6.33	2.7	174	-0.9	-58	0.8	53
8a	5.3	71	0.4	5.7	5.1	69	5.0	68
8b	-1.9	-25	0.1	1.3	1.5	20	1.4	18
9	38.9	1165	25.1	751	29.7	887	25.6	764

Table 2. The tables list manual and automatic volume differences of 10 MR sequences. Case 8a and 8b represent the volume differences between the first and second scan, and first and third scan. The relatively large agreement across all metrics between growth and dormant cases is striking. The two cases (4 and 5) where the expert results differ from the algorithm estimations are particular cases explained in the text.

We would like to note, however, that not all radiologists we consulted diagnosed these two cases as meningiomas.

A more typical sample of our data set is shown in the first row of Figures 3 corresponding to the scans of Case 2. The second row shows the differences in detection illustrating again the difficulties in visual analysis. The close correlation between expert and automatic findings for tumors with less than 2% growth is therefore very satisfying.

We also test the metrics for intra-rater variability. In the first test, the expert repeated the segmentation of Case 2 three times with a week of separation each time. These measurements varied substantially (first: $883.8mm^3$, second: $545.8mm^3$, third: $-99.8mm^3$), which demonstrates the large variability in manual measurements of small tumor growth (less 1 cm^3 in volume). It also shows that the differences between the automatic and manual measurements are within the variations expected from manual measurements. In the second test, we analyze the intra-rater variability by comparing the growth measurements between first and second scan (case 8a), and first and third scan (case 8b). The time lapse between scan two and three was very small so that we would expect small differences between the measurements. The volume changes estimated by the radiologist, $71.7mm^3$ and $-25.3mm^3$, differ, while the automatic measurements, especially of INTENSITY, only slightly change.

4. CONCLUSION

We have discussed an approach that successfully measured the volume change of slowly evolving meningioma from two successive MR scans. The correlation to expert findings shows the potential of the tool. Moreover, the tool seems to be less impacted by intra and inter-rater variability than the human expert as shown by the variance analysis in Section 3. In summary, this experiments not only stressed the need for observer independent tools for detecting volume change in slowly growing tumors but also provided a possible solution.

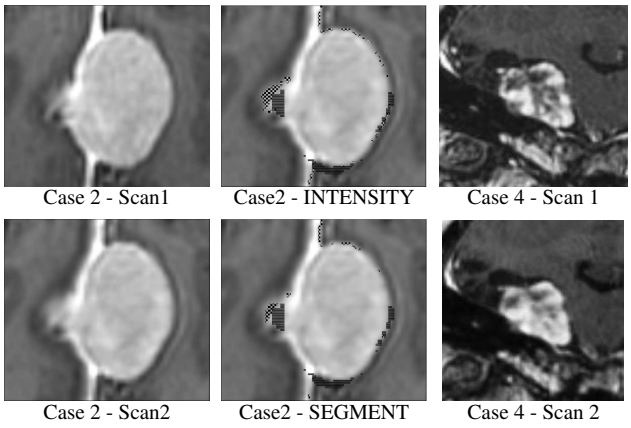

Fig. 3. The figure shows sample slices of Case 2 with the corresponding growth detection. Striped regions indicate manually detected growth, while checkerbox is the result of INTENSITY and SEGMENT. The results of JACOBIAN are omitted as they are hard to visualize. Also note, that some of the "displaced" manually detected growth is due to alignment issues between scans. We observe that the evolution of the tumor is indeed very small, which our approach successfully captures. The third row shows Case 4, which was difficult to analyze due to changes in the intensities between the scans.

5. REFERENCES

[1] D. Rey, G. Subsol, H. Delingette, and N. Ayache, "Automatic detection and segmentation of evolving processes in 3d medical images: Application to multiple sclerosis," *Medical Image Analysis*, vol. 6, pp. 163–179, 2002.

[2] E.D. Angelini, J. Atif, J. Delon, E. Mandonnet, H. Duffau, and L. Capelle, "Detection of glioma evolution on longitudinal MRI studies," in *ISBI*, 2007, pp. 49 – 52.

[3] M. Prastawa, E. Bullitt, N. Moon, K.V. Leemput, and G. Gerig, "Automatic brain tumor segmentation by subject specific modification of atlas priors," *Academic Radiology*, vol. 10, pp. 1431–1348, 2003.

[4] J. Liu, J. Udupa, D. Odhner, D. Hackney, and G. Moonis, "A system for brain tumor volume estimation via MR imaging and fuzzy connectedness," *Computerized Medical Imaging and Graphics*, vol. 29, pp. 21–34, 2005.

[5] S.O'Leary, W.M. Adams, R.W. Parrish, and W. Mukonoweshuro, "A typical imaging appearances of intracranial meningiomas," *Clinical Radiology*, vol. 62, pp. 10–17, 2007.

[6] P. Viola and W.M. Wells, "Alignment by maximization of mutual information," *IJCV*, vol. 24, pp. 137 – 154, 1997.

[7] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Non-parametric diffeomorphic image registration with the demons algorithm," in *MICCAI*, 2007, pp. 319–326.

[8] O. Clatz, M. Sermesant, P.Y. Bondiau, H. Delingette, S.K. Warfield, G. Malandain, and N. Ayache, "Realistic simulation of the 3D growth of brain tumors in MR images coupling diffusion and biomechanical deformation," *Trans. of Medical Imaging*, vol. 24, pp. 1334–1346, 2005.