Int. J. Radiation Oncology Biol. Phys., Vol. 61, No. 1, pp. 289–298, 2005

0360-3016/05/$–see front matter

doi:10.1016/j.ijrobp.2004.08.055

PHYSICS CONTRIBUTION

ATLAS-BASED AUTOMATIC SEGMENTATION OF MR IMAGES: VALIDATION STUDY ON THE BRAINSTEM IN RADIOTHERAPY CONTEXT

Pierre-Yves Bondiau, M.D., Ph.D.,*† Grégoire Malandain, Ph.D.,* Stéphane Chanalet, M.D.,‡ Pierre-Yves Marcy, M.D.,† Jean-Louis Habrand, M.D., Ph.D.,§ François Fauchon, M.D.,‖ Philippe Paquis, M.D., Ph.D.,¶ Adel Courdi, M.D.,† Olivier Commowick, Ph.D.,* Isabelle Rutten, M.D.,# and Nicholas Ayache, Ph.D.*

*Projet Epidaure, Institut National de Recherche en Informatique et Automatique, Sophia Antipolis, France; †Département de Radiothérapie, Centre Antoine Lacassagne, Nice, France; Services de ‡Radiologie and ¶Neurochirurgie, CHU Pasteur, Nice, France; §Département de Radiothérapie, Institut Gustave Roussy, Villejuif, France; ‖Centre des Hautes Energies, Nice, France; #Service de Radiothérapie, Centre Hospitalier Universitaire de Liège, Domaine Universitaire du Sart Tilman, Liège 1, Belgium

Purpose: Brain tumor radiotherapy requires the volume measurements and the localization of several individual brain structures. Any tool that can assist the physician to perform the delineation would then be of great help. Among segmentation methods, those that are atlas-based are appealing because they are able to segment several structures simultaneously, while preserving the anatomy topology. This study aims to evaluate such a method in a clinical context.

Methods and Materials: The brain atlas is made of two three-dimensional (3D) volumes: the first is an artificial 3D magnetic resonance imaging (MRI); the second consists of the segmented structures in this artificial MRI. The elastic registration of the artificial 3D MRI against a patient 3D MRI dataset yields an elastic transformation that can be applied to the labeled image. The elastic transformation is obtained by minimizing the sum of the square differences of the image intensities and derived from the optical flow principle. This automatic delineation (AD) enables the mapping of the segmented structures onto the patient MRI. Parameters of the AD have been optimized on a set of 20 patients. Results are obtained on a series of 6 patients' MRI. A comprehensive validation of the AD has been conducted on performance of atlas-based segmentation in a clinical context with volume, position, sensitivity, and specificity that are compared by a panel of seven experimented physicians for the brain tumor treatments.

Results: Expert interobserver volume variability ranged from 16.70 cm^3 to 41.26 cm^3. For patients, the ratio of minimal to maximal volume ranged from 48% to 70%. Median volume varied from 19.47 cm^3 to 27.66 cm^3 and volume of the brainstem calculated by AD varied from 17.75 cm^3 to 24.54 cm^3. Medians of experts ranged, respectively, for sensitivity and specificity, from 0.75 to 0.98 and from 0.85 to 0.99. Median of AD were, respectively, 0.77 and 0.97. Mean of experts ranged, respectively, from 0.78 to 0.97 and from 0.86 to 0.99. Mean of AD were, respectively, 0.76 and 0.97.

Conclusions: Results demonstrate that the method is repeatable, provides a good trade-off between accuracy and robustness, and leads to reproducible segmentation and labeling. These results can be improved by enriching the atlas with the rough information of tumor or by using different laws of deformation for the different structures. Qualitative results also suggest that this method can be used for automatic segmentation of other organs such as neck, thorax, abdomen, pelvis, and limbs. © 2005 Elsevier Inc.

Brain tumors, Radiotherapy, Magnetic resonance imaging, Segmentation matching.

INTRODUCTION

Brain tumors

The majority of the primary tumors of the central nervous system are from astrocytic or oligodendroglial origin. The treatment of these tumors may require surgery, radiotherapy, or chemotherapy. Since the 1980s, the incidence of the primary and secondary tumors of the central nervous system has been increasing among all age categories, whereas mortality has been decreasing for patients younger than 65 years old (1).

The treatments of high-grade glioma are disappointing. Anaplastic astrocytomas and glioblastomas represent the most aggressive tumors. Despite the various research treatment protocols, overall survival is about 1 year. The overall survival is similar for combined treatments and is quite comparable to results of other series that used radiotherapy alone (2). The efficiency of chemotherapy

Reprint requests to: Pierre-Yves Bondiau, Département de Radiothérapie, Centre Antoine Lacassagne, 33 av de Valombrose, 06189, Nice, France; Tel: 33 4 92 03 12 61; Fax: 33 4 92 03 15 70; E-mail: pierre-yves.bondiau@cal.nice.fnclcc.fr

Received Feb 3, 2004, and in revised form Aug 9, 2004. Accepted for publication Aug 16, 2004.

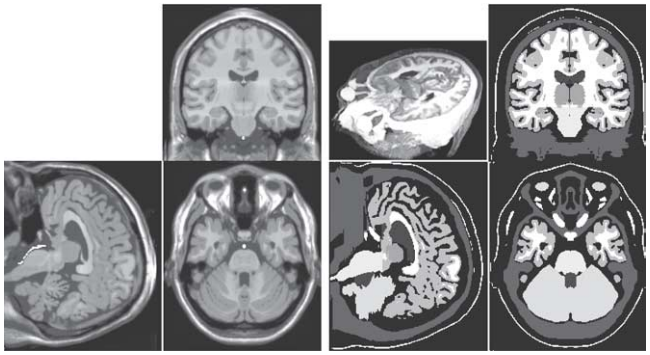

Fig. 1. Atlas purpose-built for radiotherapy. (Left) Symmetrical artificial magnetic resonance imaging (MRI) (data from BrainWeb); basilar artery was added to avoid matching this structure with the brainstem. (Right) Result of the segmentation of the artificial MRI and the three-dimensional view. Main organs at risk are: eyes, optic nerve, optic chiasm and optic tract, pituitary gland, brainstem, and basal ganglia (thalamus, caudate ganglia, putamen, and pallidum). Fat, skull, and skin are segmented for radiotherapy purposes.

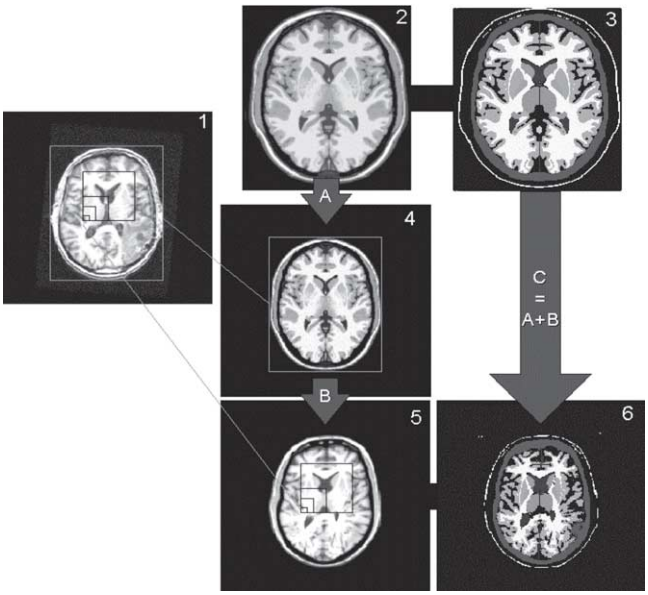

Fig. 2. Flowchart of the automatic delineation. (1) Magnetic resonance imaging (MRI) to be segmented. (2) Artificial MRI, high-definition. (3) Atlas of (2). First, (2) is put in the geometry of (1) with a rigid transformation A: this yields (4). There is no deformation of internal structures in (4). Then, a nonrigid transformation B is used to match (4) on (1); that yields (5). There is deformation of internal structures; for example, deformation of the left ventricle by the tumor. (6) Automatic delineation segmented image. We deformed (3) by C, which is the A+B transformation.

seems to be low (3, 4). Patients with high-grade astrocytomas are treated by radiotherapy protocols using three-dimensional (3D) treatment planning and conformal radiotherapy. Other radiation methods used to treat this kind of tumor are hadron therapy (5), radiosurgery, gamma knife (6), or CyberKnife treatment. These high-dose conformal radiotherapies could improve disease survival and change the pattern of failure (7). Further dose escalation seems reasonable, based on the same target volume definition criteria.

3D conformal radiotherapy and 3D reconstruction

Conformal radiotherapy (CRT) consists of a set of strict procedures allowing high-precision radiotherapy for improvement of local tumor control and reduction of complications. Recent data from the literature show a significant reduction of adverse effects as compared with conventional radiotherapy and the possibility of increasing tumoral control by using conformal techniques (8, 9). Moreover, conformal radiotherapy enables a dose distribution of high precision in cerebral structures. However, it requires the accurate delineation of structures of interest in the brain and the tumor to determine the best characteristics of the treatment plans. A deeper interpretation of the results of the CRT also requires the accurate localization of other cerebral structures to estimate the irradiation dose of each of them. Furthermore, the localization of the same cerebral structures is still needed for patient follow-up. Long-term evaluation of adverse effects should take into account the functional results, which could be related to the dose to normal cerebral structures. Thus the knowledge of doses delivered in each cerebral structure is also necessary to evaluate the functional results.

Before now, the delineation of brain structures was a manual procedure structure with each structure drawn in each slice of the 3D magnetic resonance imaging (MRI) dataset, which yields a poor 3D coherence. The mean time for the analysis and delineation of a typical brain MRI dataset has been evaluated to 86 min, with poor reproducibility. The interobserver coefficient of variation estimate is 11 ± 5.8% and the intraobserver coefficient of variation is 9 ± 5.2% (10, 11). The treatment team spends significant time delimiting the various structures of interest. Moreover, this delineation is not reproducible, which is a drawback for a consistent statistical analysis and patient follow-up.

Obviously, any image processing tool that can assist the physician to perform the delineation of the cerebral structures will be of great value because it will both reduce the delineation time and ensure better reproducibility of the segmentation results. Image segmentation has already been widely studied in the literature. One choice is to segment each structure of interest separately with an ad hoc approach. However, inconsistencies (e.g., overlap of adjacent structures) are likely to occur. A more appealing approach is an atlas-based segmentation that allows simultaneous mapping of all the structures in the patient's image dataset. Software able to accurately delineate and label all the cerebral structures of interest for any patient is very attractive. It allows clinical benefits: automatic evaluation of the radiotherapy doses delivered in the cerebral structures of interest, better treatment planning by minimizing irradiation of normal tissue and critical structures, and better patient follow-up and pro-

Table 1. Image characteristics

MRI	T1W, T1W with contrast	MR1 T2W	Artificial MRI T1WRR
XDIM	256	256	181
YDIM	256	256	217
ZDIM	60	64	181
TYPE	Unsigned fixed	Unsigned fixed	Signed fixed
PIXSIZE	16 bits	16 bits	16 bits
VX (mm)	1.015623	1.015641	0.6
VY (mm)	1.015625	1.015625	0.6
VZ (mm)	2.000000	1.900000	0.6
Field (mm)	173 * 230	173 * 230	

Abbreviations: MRI = magnetic resonance imaging; T1W = T1 weighted; T2W = T2 weighted.

tocol evaluation by studying doses in brain structures. However, because of the difficulty of such a task, a very careful validation study has to be conducted. This is outside the scope of the present work, which intends to demonstrate, on a small number of subjects, the feasibility of automatic delineation with an atlas-based approach.

METHODS AND MATERIALS

The automatic delineation (AD) was performed by an atlas-based segmentation technique whose results on brainstem were compared with the experts' delineation. Atlas-based segmentation consisted of a series of image alignments (or matching) that allowed the propagation of the labeled structures onto the MRI. This MRI was used to elaborate the treatment planning. It helped obtain a rigorous anatomical model and offered an alternative to the usual manual method. Thus the pivotal point was the computation of an elastic transformation between the atlas and the patient MR image.

Atlas

Practically, our atlas was made of two images, an MRI and a series of structures labeled in this image. In our first experiments, the atlas was built from the T1-W MRIs of healthy subjects. However, those images showed that the individual variation of our healthy subject introduced a systematic error reported on every patient MRI. To overcome this issue, a fully artificial MRI was chosen. This artificial MRI dataset was generated with BrainWeb (12) software. However, asymmetries still introduced errors; therefore, we then mirrored the right side of the head to generate a perfectly symmetrical atlas MRI (Fig. 1). Artificial MRI characteristics are shown in Table 1. To minimize partial volume effects during the registration process, a high-definition MRI with smaller voxels than on the patient MRI was used. An expert delineated the contours of each cerebral structure of interest in this artificial MRI. Because our primary interest was radiotherapy, we focused on structures of interest for this application: that is, the organs at risk and prominent normal structures. The main organs at risk (OAR) are eyes, optic nerve, optic chiasm and optic tract, pituitary gland, brainstem, and basal ganglia (thalamus, caudate ganglia, putamen, and pallidum). Prominent normal structures are skull, ventricular system, brain (gray matter and white matter), brainstem, and cerebellum. Fat and skin were added in the atlas for a more realistic radiotherapy representation. Basilar artery was added in the artificial MRI to avoid matching of the brainstem with this structure. The delineation of all these structures was done manually according to anatomical data of the atlas of Talairach. Although it was a tedious task, we are convinced that it was necessary given the potential benefits of this automatic segmentation approach.

Table 2. Image sequences

MRI	T1W, T1W with contrast	T2W
TE (ms)	1.7	102
TR (ms)	7.9	5000
Bandwidth	15.153	31.25
FOV (cm)	26	26
Excitation	2	2

Abbreviations: MRI = magnetic resonance imaging; T1W = T1 weighted; T2W = T2 weighted; FOV = field of view.

Computation of the warping transformation

Given the atlas and an input patient MRI, we needed to compute an elastic transformation that would enable warping the labeled structures of the atlas onto the patient images (13). We performed a transformation search as a process of computing a deformation field; that is, a 3D vector per each 3D point, which yielded to a large number of unknown parameters.

To facilitate this computation, we first roughly aligned the MRI of the atlas with the patient's image using a block-matching algorithm. In this case, a 3D version of the approach is detailed in references (14, 15). This yielded a 3D rigid transformation (three parameters for the translation, three parameters for the rotation) that allowed the resampling of the original atlas MRIs and yielded to an individualized low-resolution atlas MRI roughly aligned with the patient's image.

We then computed the deformation field that consisted of the residual deformations between the low-resolution atlas MRIs and the patient's MRI. This was done by minimizing the sum of the square differences of the image intensities.

The elastic transformation was done by minimizing the sum of the squares and was derived from optical flow principles (16). This algorithm provided a method to find the transformation that maps the intensities of one image to those of another (Fig. 2) to provide

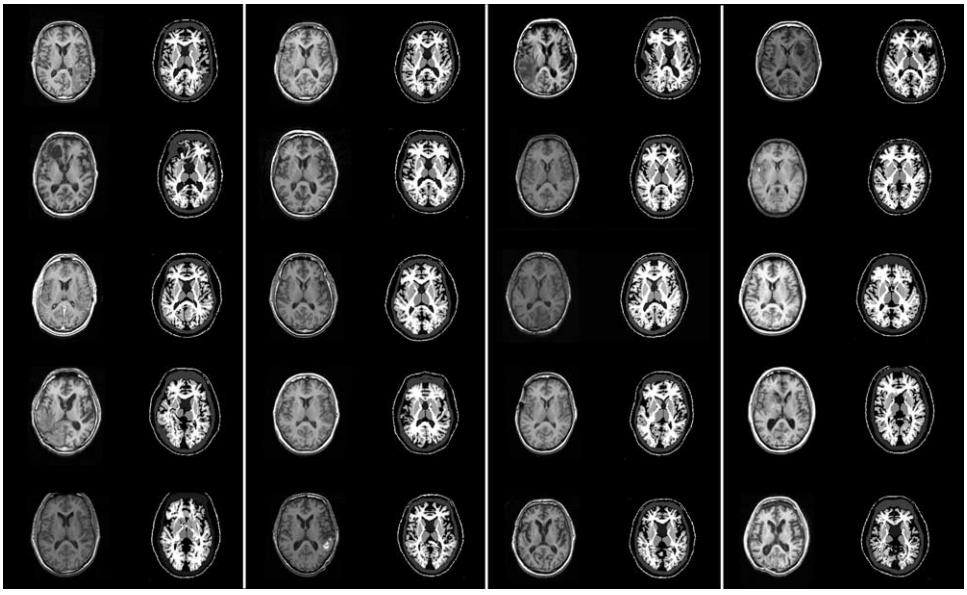

Fig. 3. Results of the automatic delineation (AD) for a set of 20 patients' magnetic resonance images made during the calibration of the AD with the same parameters set for the 20 patients.

an AD for image segmentation. In our database, all MRIs had the same intensity scale; if images could have quite different intensities, one would have also to correct for this by intensity scaling.

Parameter estimation

The ability to determine a reasonable estimate of the critical parameters was essential to a successful matching. To this end, we selected the best parameters set of the rigid and nonrigid automatic matching methods by visually inspecting the results of the AD for 20 patients' MRIs (Fig. 3) with different pathologies (astrocytoma Grade 3, glioblastoma, meningioma, and metastasis). This was helpful to the calibration of the AD.

Images

Standard MRI protocols for brain tumor radiotherapy were used for this study, which was more challenging but of higher interest. MRI acquisition was performed with a head coil. The MRIs were later matched via rigid matching techniques with the dosimetric computed tomography image for treatment planning purposes. Three sequences—T1, T2, and T1—with gadolinium contrast injection were acquired, with strictly the same imaging sequence for all patients (see details of the imaging sequences in Table 2). For this work, we used only T1-weighted images exported in Dicom-3 format; image characteristics are shown in Table 1. Because our purpose was to validate the atlas-based segmentation in MRI, we will not describe the rigid matching between patient MRI and computed tomography scan image, which can be done by various methods.

Expert segmentation

A variety of brain specialists formed our panel of experts: neurosurgeons, neuroradiologists, and radiotherapists. Moreover, each radiotherapist had a particular competence in proton therapy, radiosurgery, conformal radiotherapy, or gamma-knife treatment. We chose the brainstem to estimate the quality of our atlas-based segmentation because of its moderate size (a smaller structure would have induced too much variability between experts), its central localization, and the frequency with which our experts delineate it.

Experts were asked to manually delineate the brainstem in 6 patients. These 6 patients were randomly chosen out of the 20 patients. The experts did the delineation with a mouse by drawing a curve outside of the delineated structure. There was no lesion in the brainstem, but patients did have brain lesions elsewhere. The images used for the manual segmentations were the T1-weighted MRIs with gadolinium contrast injection that were rigidly registered beforehand against the T1-weighted MR images.

Validation study of atlas-based segmentation in clinical context

For each patient, the experts' delineation of the brainstem was fused with the approach described by Warfield (17). This expectation-maximization algorithm allows the estimation of a "ground truth" segmentation from a group of experts' segmentation and a measure of the quality (in terms of sensitivity and specificity) of each expert. This "ground truth" was a 3D map of probabilities (W_i) in which W_i expressed the probability of voxel i belonging to the brainstem. The sensitivity can be described as the ratio of the probability for a voxel to be correctly classified by the automatic delineation as a brainstem voxel over the probability of a voxel to belong to the brainstem. The specificity was the probability for a voxel to have been incorrectly classified as part of the brainstem by the automatic delineation with respect to the probability of a voxel not to belong to the brainstem. Given a probabilistic segmentation, the expressions of the sensitivity, P, and the specificity, Q, become:

$$P = \frac{\sum_{D_i=1} W_i}{\sum_{D_i=1} W_i + \sum_{D_i=0} W_i} \text{ and } Q = \frac{\sum_{D_i=0} (1 - W_i)}{\sum_{D_i=1} (1 - W_i) + \sum_{D_i=0} (1 - W_i)} \quad (1)$$

where $D_i = 1$ represents the voxels labeled as brainstem and $D_i = 0$ the other voxels. Classical formulas with numbers of true/false

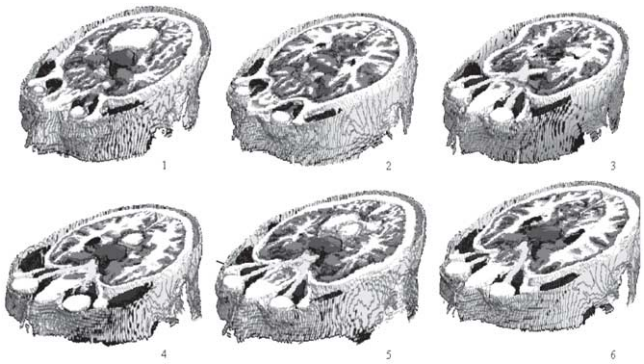

Fig. 4. Result of the automatic delineation (AD) segmentation on 6 patients; axial slice of a three-dimensional view. Segmented structures are the same as those of the atlas (Fig. 1). AD parameters have been optimized on a set of 20 patients with brain lesions.

positive/negative points can be retrieved with binary probabilities (i.e., $W_i \in \{0,1\}$). After the "ground truth" was computed from all experts' segmentations, this method enabled the assessment of our AD method by computing the same measure of quality and direct comparison with the expert's performance.

RESULTS

Atlas deformation

The validation study of atlas-based segmentation in clinical context was conducted on 6 patients' MRIs of 20 patients' MRIs, fully labeled by the AD, to compare segmentation of the brainstem between experts and AD. Results of patient MRI fully labeled by the AD are shown (Figs. 4 and 5). In the vicinity of the tumor and, because there was no volume for tumor in the atlas, there was a normal structure with an intensity similar to the tumor that deformed and took the place of the tumor in the AD-labeled MRI (Fig. 5, Patient 5). The total duration of the AD to obtain a fully labeled MRI was 20 min on a 1-GHz dual processor computer.

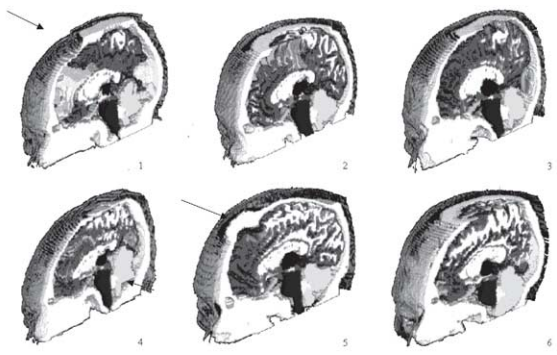

Fig. 5. Result of the automatic delineation segmentation on 6 patients, sagittal view. In Patient 1, the skull is deformed by an edge effect. The (wrong) deformation of the skull made by the tumor (arrow) can be seen in Patient 5. Tumor deforms the skull and the cerebellum in Patients 1 and 4 (arrows).

Volume of the brainstem

For each patient, the maximal, minimal, and median volumes of the expert segmented brainstem were computed. The interobserver variability was important: brainstem volume varied between 16.70 cm^3 and 41.26 cm^3 across our patients. For the same patient, interobserver volume ranged from 48% to 70%. Median volume varied from 19.47 cm^3 to 27.66 cm^3 (Table 3). Volume of the brainstem calculated by AD using the matching methods varied from 17.75 cm^3 to 24.54 cm^3 (Fig. 6).

Position of the brainstem

To characterize the performance of image segmentation was a persistent challenge, and the position of the brainstem was as important a descriptor as its volume. To investigate position variation, we computed a probability image of the brainstem outline from expert segmentations. In this image, the voxel intensity of each brainstem voxel represents the probability for this voxel to belong to the brainstem. The lowest probability is in regions in which the expert segmentation was most difficult. From this image, we computed the maximal and minimal volume. The structure of maximal volume drawn by the experts was defined by the set of all voxels belonging to the structure for one or more experts: it represents the set of voxels that at least one expert considered belonging to the brainstem. The structure of minimal volume drawn by experts was defined by the set of all voxels belonging to the structure for all experts; it represents voxels for which all experts consider that they belong to the brainstem.

These volumes can be used to reflect the position of the AD segmentation of the brainstem (see Fig. 7 for a sagittal view and Fig. 8 for a rear view). Quantitative measures of quality (i.e., sensitivity and specificity as described previously) can also be calculated and compared with those of experts. Warfield proposed an expectation-maximization (18) algorithm to compute a probabilistic estimate of the "ground truth" segmentation from a group of expert segmentations and a simultaneous measure of the quality of each expert. It provided a direct comparison of expert and algorithm performances by estimating the sensitivity and specificity of the AD segmentation with respect to the "ground truth" derived from experts' segmentations.

Sensitivity and specificity

The sensitivity and specificity of experts and AD computed with respect to the "ground truth" derived from the experts' segmentation were computed. For sensitivity and specificity, respectively, the medians of the experts range from 0.75 to 0.98 and from 0.85 to 0.99. Medians of AD are, respectively, 0.77 and 0.97 (Tables 4 and 5). Mean of the experts range, respectively, from 0.78 to 0.97 and from 0.86 to 0.99. Mean of AD are, respectively, 0.76 and 0.97. To be more comprehensive, these performances have been represented for each patient; we plot the sensitivity and specificity of each expert and AD (Fig. 9). Because the relevant statistical concept is similar to receiver operating character-

Table 3. Volume of brainstem in square centimeters by seven experts (exp_n) of 6 patients and results with automatic delineation

	PAT_1	PAT_2	PAT_3	PAT_4	PAT_5	PAT_6
exp_1	19.47	20.03	25.29	25.65	21.99	28.92
exp_2	19.36	20.83	23.59	22.25	22.25	22.09
exp_3	20.15	28.27	26.47	22.47	22.56	22.77
exp_4	26.76	34.66	33.64	41.26	38.93	39.94
exp_5	24.89	25.58	29.05	29.22	27.92	27.66
exp_6	16.72	19.02	27.30	19.97	18.91	27.97
exp_7	16.70	19.52	24.54	22.57	22.79	24.35
AD	17.75	22.40	23.74	22.69	24.54	19.28
Mean	20.58	23.99	27.13	26.20	25.05	27.67
Median	19.47	20.83	26.47	22.57	22.56	27.66
Minimum	16.70	19.02	23.59	19.97	18.91	22.09
Maximum	26.76	34.66	33.64	41.26	38.93	39.94
Standard deviation	1.8610	3.6776	2.3222	5.2654	4.7786	4.3217

istic, we calculate best-fit A_z curves for our experts and for the AD (Fig. 10).

DISCUSSION

A variety of automatic techniques for the segmentation of the OAR has been proposed in the literature (19–28). Some approaches require an observer to guide the segmentation of structures.

It should be emphasized that all the image processing steps of the AD are fully automatic to ensure reproducibility of the results. This useful property yields an operator-independent, and thus perfectly reproducible, structure (as OAR) segmentation system. Rigid and nonrigid matching use an iconic method (29) that does not require fiducial markers. Furthermore, the user is not required to identify common structures on the two image sets. The final model carries out a segmentation of different brain structures that are clearly identified by different gray levels.

Expert and AD brainstem volume

Delineation of the OAR is a critical step during the CRT procedure. In addition, the OAR must be interpreted cautiously in light of the observed topologic differences, because delineation of structures of interest (gross tumor volume and high-risk organs) varies considerably from one physician to another (30–32). In this study, standard deviation of brainstem volume varies from 1.86 to 5.26. The interobserver variability is important, which illustrates that small surface variation may produce high volume variation. Another source of volume variation is the limits of a structure in the image—for example, the superior limit for the brainstem is difficult to clearly identify in the MRI. The problems of limits cannot even be solved with anatomical

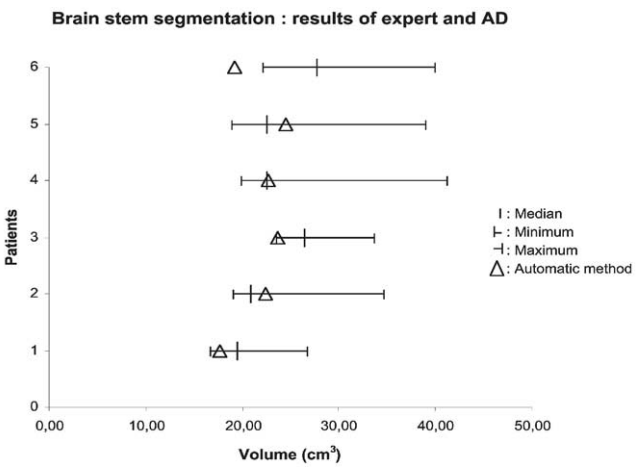

Fig. 6. Brainstem segmentation: results of experts and automatic delineation (AD). For the 6 patients, the maximum, minimum, and median volumes of the brainstem according to experts segmentation are plotted. The volume of the brainstem given by the AD is also shown. In all cases but one, the volume of the AD is within the range of experts' segmentations.

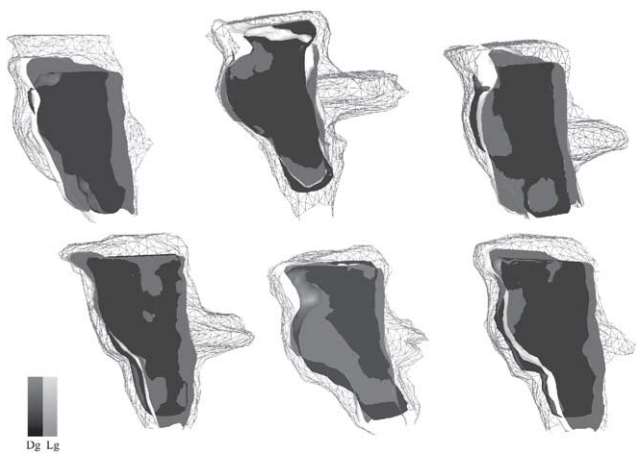

Fig. 7. Surface-rendered sagittal view of brainstem for the 6 patients in the study. The maximal volume segmented by experts is wireframe, automatic delineation (AD) segmentation is in light gray, and minimal volume segmented by experts is in dark gray. In most of the cases, the volume of the AD is between the minimal and the maximal volume segmented by experts. White color is produced by lighting effects in surface rendering.

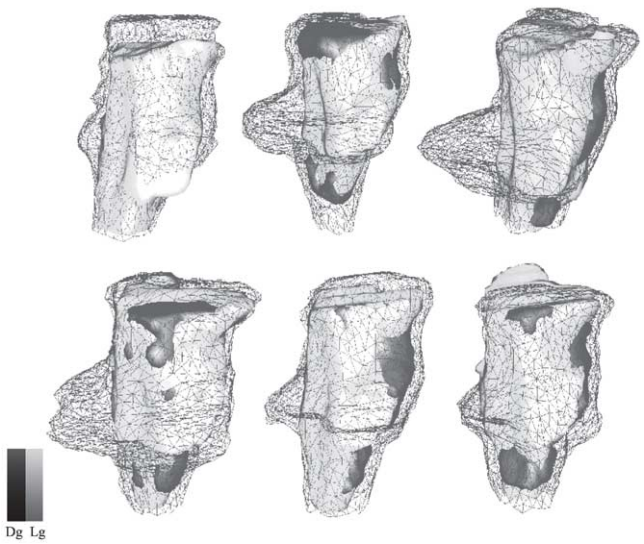

Fig. 8. Surface-rendered right rear-superior view of brainstem for the 6 patients. The maximal volume segmented by experts is wireframe, automatic delineation (AD) segmentation is in light gray, and minimal volume segmented by experts is in dark gray. The brainstem segmented by AD is inside the maximal volume segmented by experts. White color is produced by lighting effects in surface rendering.

definition of a structure, because the posterior limits of the brainstem in front of the fourth ventricle are made with the inferior, middle, and superior cerebellum peduncle. Anatomically, the middle cerebellum peduncle forms part of the brainstem and part of the cerebellum, so the boundary between brainstem and cerebellum is in the middle cerebellum peduncle. Because the middle cerebral peduncle is a continuous structure that goes from the brainstem to the cerebellum, the limit can be very difficult to define, even for experts. This is illustrated in Figs. 6 and 7, in which the middle cerebellum peduncle appears in wireframe reflecting the variations of expert segmentations.

The comparison between experts and AD results shows that the brainstem volume lies in five cases between the variations of the experts. In one case (Patient 6) the AD results are less important than the minimal volume of experts. In this case, the AD stops the reconstruction in the pyramid, which is the lower part of the brainstem (Fig. 6: arrow, Patient 6), whereas minimal volume drawn by the expert includes the entire pyramid. The vicinity of the tumor near the brainstem can explain this problem. The AD segmentation is reproducible (no intraexpert variability) and it depends on image resolution (33).

Brainstem position

This technique was useful in the brain representation because the data processing reconstruction of brain allows an accurate localization of the structures. This was particularly important in CRT in so far as one tries to reduce the distal safety margins to protect healthy brain structures. This reduction of margin is very important for pediatric oncology (34, 35). The fully automatic segmentation and labeling of patient brains is useful to save time for drawing a more accurate gross tumor volume. It allows automatic dose–volume histograms distribution. Figure 7 shows that the brainstem segmented with the AD most often remains within the maximal volume segmented by experts. The main variation between the two volumes was in front of the middle cerebellum peduncle, which can be explained by the expert variation as, in the atlas, the segmentation was a straight line across this structure (Fig. 1, right axial view). In Fig. 6, the AD volume appears closer to the minimal volume, and respects the deformation in regard to the pons.

The sensitivity, which is defined as the true positive (TP) divided by the sum of TP and false negative (FN), is maximal if FN is zero. This is the case if a delineation larger than the "ground truth" is made. The AD sensitivity, which represents the probability for a voxel segmented as belonging to the brainstem by the AD to be a TP, is not very high, showing delineation very close to the "ground truth." The AD specificity, which represents the probability for a voxel to belong to the brainstem, is high. The specificity, which is defined as the true negative (TN) divided by the sum of TN and false positive (FP), is maximal if FP is zero. FP is the number of voxels segmented as belonging to the brainstem by the AD and not classified as brainstem by the experts. A mask was used to minimize the high number of TN in the image; the mask was made by a dilation of the maximal brainstem segmentation issued from all expert segmentations.

Figure 5, Patient 5, illustrates the wrong estimation of the brain structure near the lesion in respect of the size and the

Table 4. Sensitivity of experts and automatic delineation segmentation of the brainstem

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Automatic
pat_1	0.85	0.83	0.86	0.99	0.92	0.73	0.74	0.75
pat_2	0.82	0.84	0.95	0.98	0.91	0.76	0.80	0.81
pat_3	0.84	0.86	0.84	0.98	0.97	0.91	0.86	0.76
pat_4	0.92	0.83	0.76	0.96	0.98	0.73	0.85	0.78
pat_5	0.87	0.54	0.81	1.00	0.98	0.73	0.86	0.83
pat_6	0.91	0.75	0.77	0.93	0.87	0.89	0.82	0.64
Mean	0.87	0.78	0.83	0.97	0.94	0.79	0.82	0.76
Median	0.86	0.83	0.82	0.98	0.94	0.75	0.83	0.77

Table 5. Specificity of experts and automatic delineation of the brainstem

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Automatic
pat_1	0.99	0.98	0.98	0.91	0.92	0.99	0.99	0.98
pat_2	1.00	0.99	0.93	0.85	0.95	0.99	1.00	0.96
pat_3	0.96	1.00	0.94	0.89	0.96	0.96	0.98	0.95
pat_4	0.98	0.99	0.97	0.81	0.96	0.99	1.00	0.97
pat_5	1.00	0.90	0.97	0.83	0.96	0.99	0.99	0.95
pat_6	0.97	1.00	0.99	0.84	0.97	0.97	0.99	0.99
Mean	0.98	0.98	0.96	0.86	0.95	0.98	0.99	0.97
Median	0.98	0.99	0.97	0.85	0.96	0.99	0.99	0.97

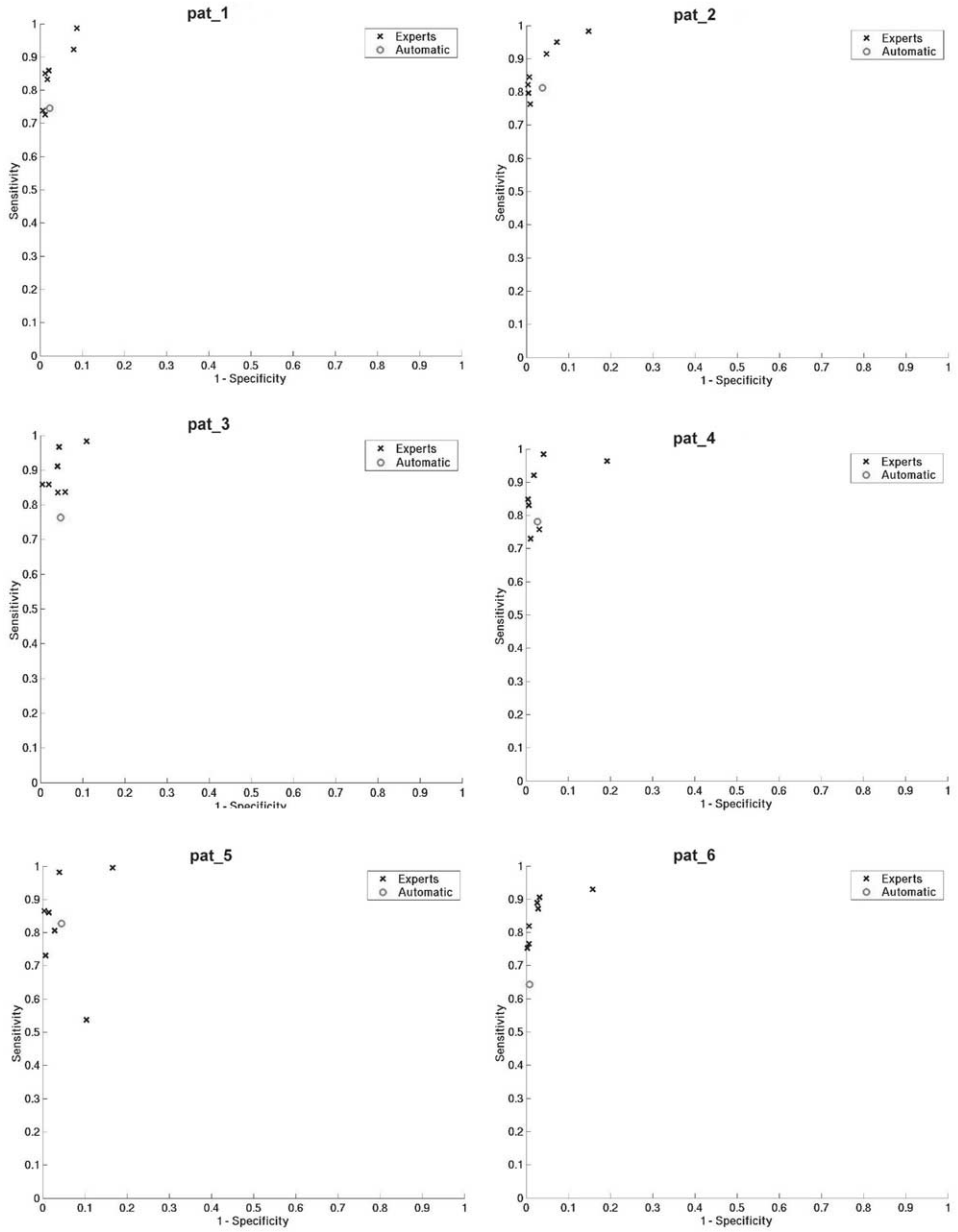

Fig. 9. Results of expert and automatic delineation (AD). Sensitivity and specificity have been plotted for each patient and for each expert: crosses represent experts' sensitivity and specificity; circles represent AD sensitivity and specificity. In only two cases, AD has a lower sensitivity than expert segmentation; in the other four cases, AD sensitivity is within expert variation. In all cases, AD specificity is between expert variation.

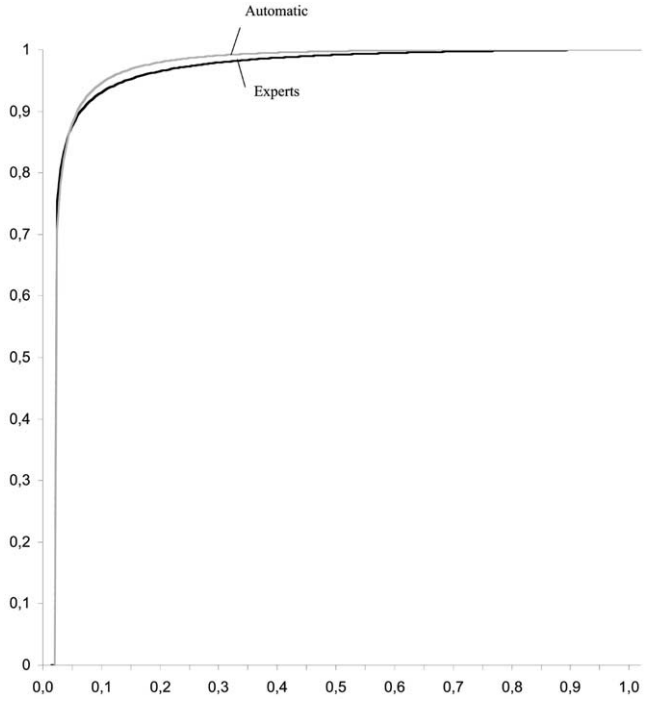

Fig. 10. Receiver operating characteristic–like curves of experts and automatic delineation for all experts and all patients.

position of the tumor. The problem was that the AD tried to deform normal tissues, which had the same intensities as the tumor to fill this volume. A rough segmentation or a manual segmentation of the tumor could be automatically included in the atlas to avoid this deformation of normal tissues (36). Another way to improve the AD could be to use different laws of deformation for the different brain structures.

The results for the sensitivity and the specificity for AD and experts seem to be acceptable for 4 patients (Patients 1, 2, 4, 5) and poor in Patients 3 and 6. In these two cases, the tumor is near the brainstem; this vicinity could explain these poor results. This could be corrected by introducing the tumor volume in the atlas.

Conclusion

This validation study of atlas-based segmentation in clinical context shows that fully automated software is useful in radiotherapy. Results demonstrate that the method is repeatable and leads to reproducible segmentation and labeling. The AD used for the automatic segmentation and labeling suggest that this method provides a good trade-off between accuracy and robustness. A comprehensive validation of the AD has been conducted on a series of 6 patients' MRIs to study the performance of atlas-based segmentation in a clinical context. Volume, position, sensitivity, and specificity can be compared with experts' segmentations.

For sensitivity, the medians of the experts ranged from 0.75 to 0.98 and means ranged from 0.78 to 0.97; the AD median is 0.77 and mean is 0.76.

For specificity, the medians of the experts ranged from 0.85 to 0.99 and means range from 0.86 to 0.99; for the AD, median is 0.97 and mean is 0.97.

Further work is needed to avoid normal tissue deformation near tumors. These results can be improved by enriching the atlas with additional information about the tumor or by using different laws of deformation for the different structures.

These qualitative results suggest that this approach can be used for automatic segmentation of other organs, such as neck, thorax, abdomen, pelvis, and limbs.

REFERENCES

1. Legler JM, Ries LA, Smith MA, *et al*. Brain and other central nervous system cancers: Recent trends in incidence and mortality. *J Natl Cancer Inst* 1999;91:1382–1390.
2. Miralbell R, Mornex F, Greiner R, *et al.* European Organization for Research and Treatment of Cancer Trial 22933. *J Clin Oncol* 1999;17:3143–3149.
3. Wolff JE, Trilling T, Molenkamp G, *et al.* Chemosensitivity of glioma cells in vitro: A meta analysis. *J Cancer Res Clin Oncol* 1999;125:481–486.
4. Frenay M, Lebrun C, Lonjon M, *et al.* Up-front chemotherapy with fotemustine (F)/cisplatin (CDDP)/etoposide (VP16) regimen in the treatment of 33 non-removable glioblastomas. *Eur J Cancer* 2000;36:1026–1031.
5. Fitzek MM, Thornton AF, Rabinov JD, *et al.* Accelerated fractionated proton/photon irradiation to 90 Cobalt gray equivalent for glioblastoma multiforme: Results of a phase II prospective trial. *J Neurosurg* 1999;91:251–260.
6. Szeifert GT, Massager N, Brotchi J, *et al.* Morphological redifferentiation in a malignant astrocytic tumor after gamma knife radiosurgery. *J Neurosurg* 2002;97:627–630.
7. Nakagawa K, Aoki Y, Fujimaki T, *et al.* High-dose conformal radiotherapy influenced the pattern of failure but did not improve survival in glioblastoma multiforme. *Int J Radiat Oncol Biol Phys* 1998;40:1141–1149.
8. Fischer JJ, Moulder JE. The steepness of the dose-response curve in radiation therapy. *Radiology* 1975;117:179–184.
9. Moore JV, Hendrey JH, Hunter RD. Dose incidence curves for tumor control and normal tissue injury in relation to the response of clonogenic cells. *Radiother Oncol* 1983;1:143–157.
10. Grimaud J, Lai M, Thorpe J, *et al.* Quantification of MRI lesion load in multiple sclerosis: A comparison of three computer-assisted techniques. *Magn Reson Imaging* 1996;14:495–505.
11. Yamamoto M, Nagata Y, Okajima K, *et al.* Differences in target outline delineation from CT scans of brain tumours using different methods and different observers. *Radiother Oncol* 1999;50:151–156.
12. Cocosco CA, Kollokian V, Kwan RHS, *et al.* BrainWeb: Online interface to a 3D MRI simulated brain database [Abstract]. *Neuroimage* 1997;5:S425.
13. Woods RP. Characterizing volume and surface deformations in an atlas framework : Theory, applications, and implementation. *Neuroimage* 2003;18:769–788.
14. Ourselin S, Roche A, Subsol G, *et al.* Reconstructing a 3D structure from serial histological sections. *Image Vision Comp* 2001;19:25–31.
15. Guimond A, Roche A, Ayache N, *et al.* Three-dimensional multimodal brain warping using the demons algorithm and

adaptive intensity corrections. *IEEE Trans Med Imaging* 2001;20:58–69.
16. Thirion JP. Image matching as a diffusion process: An analogy with Maxwell's demons. *Med Image Anal* 1998;2:243–260.
17. Warfield SK, Zou KH, Wells WM. Validation of image segmentation and expert quality with an expectation-maximization algorithm. In: Dohi T, Kikinis R, editors. Berlin: Springer-Verlag; 2002. p. 298–206.
18. Dempster A, Laird N, Rubin D. Maximum-likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc Ser B* 1977;39:34–37.
19. Friedlinger M, Schroder J, Schad LR. Ultra-fast automated brain volumetry based on bispectral MR imaging data. *Comput Med Imaging Graph* 1999;23:331–337.
20. Saeed N, Hajnal JV, Oatridge A. Automated brain segmentation from single slice, multislice, or whole-volume MR scans using prior knowledge. *J Comput Assist Tomogr* 1997;21:192–201.
21. Wang D, Galloway GJ, de Zubicaray GI, *et al.* A reproducible method for automated extraction of brain volumes from 3D human head MR images. *J Magn Reson Imaging* 1998;8:480–486.
22. Lemieux L, Hagemann G, Krakow K, *et al.* Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magn Reson Med* 1999;42:127–135.
23. Ong HT, Tieman J, Albert M, *et al.* Semi-automated extraction of brain contours from MRI. *Neuroradiology* 1997;39:797–803.
24. Ruan S, Jaggi C, Xue J, *et al.* Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Trans Med Imaging* 2000;19:1179–1187.
25. Treves ST, Mitchell KD, Habboush IH. Three dimensional image alignment, registration and fusion. *Q J Nucl Med* 1998;42:83–92.
26. Dawant BM, Hartmann SL, Thirion JP, *et al.* Automatic 3D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part I, methodology and validation on normal subjects. *IEEE Trans Med Imaging* 1999;18:909–916.
27. Barra V, Boire JY. Automatic segmentation of subcortical brain structures in MR images using information fusion. *IEEE Trans Med Imaging* 2001;20:549–558.
28. Shan ZY, Yue GH, Liu JZ. Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. *Neuroimage* 2002;17:1587–1598.
29. Cachier P, Bardinet E, Dormont D, *et al.* Iconic feature based nonrigid registration: The PASHA algorithm. *Computer Vision and Image Understanding*. 2003;89:272–298.
30. Weltens C, Menten J, Feron M, *et al.* Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother Oncol* 2001;60:49–59.
31. Giraud P, Elles S, Helfre S, *et al.* Conformal radiotherapy for lung cancer: Different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002;62:27–36.
32. Aoyama H, Shirato H, Nishioka T, *et al.* Magnetic resonance imaging system for three-dimensional conformal radiotherapy and its impact on gross tumor volume delineation of central nervous system tumors. *Int J Radiat Oncol Biol Phys* 2001;1;50:821–827.
33. Roche A, Pennec X, Malandain G, *et al.* Rigid registration of 3D ultrasound with MR images: A new approach combining intensity and gradient information. *IEEE Trans Med Imaging* 2001;20:1038–1049.
34. Habrand JL, Mammar H, Ferrand R, *et al.* Proton beam therapy (PT) in the management of CNS tumors in childhood. *Strahlenther Onkol* 1999;175:91–94.
35. Habrand JL, Bondiau PY, Dupuis O, *et al.* Late effects of radiotherapy in children. *Cancer Radiother* 1997;1:810–816.
36. Dawant BM, Hartmann SL, Pan S, *et al.* Brain atlas deformation in the presence of small and large space-occupying tumors. *Comput Aided Surg* 2002;7:1–10.