Evaluation of Atlas Construction Strategies in the Context of Radiotherapy Planning

Olivier Commowick[1],[2], Grégoire Malandain[1]

[1]INRIA - Epidaure/Asclepios Project, 2004 Rte des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
[2]DOSISoft S.A., 45-47 Avenue Carnot,
94 230 Cachan, France

***Abstract*— Radiotherapy planning requires accurate delineations of the critical structures. Atlas-based segmentation has been shown to be very efficient to delineate brain structures. On other parts of the body, using an atlas built from one single image as for the brain does not seem adequate, since the structures to be delineated are not clearly defined. Using only one image may then introduce undesirable bias. Building an atlas from a set of segmented images address this issue, but it will then depend on the choice of the registration method used to fuse the images. This point is generally not addressed in the literature, and is the aim of this article. Since the atlas is designed to delineate structures, we will evaluate together both the registration method used to build the atlas, and the one used to deform the built atlas on an individual image. We illustrate our framework on the construction of an atlas of the head and neck area. Using atlas-based segmentation to delineate critical structures in this area seems indeed very interesting, as a large part of the cancers (7 %) arise there. We compare the results obtained using three different methods on a real dataset of manually segmented images.**

I. Introduction

Radiotherapy requires to delineate accurately the critical structures as well as the target volume, in order to define precisely the irradiation beams during the treatment. This task is very tedious to do manually and not reproducible. The use of an anatomical atlas to automatically and simultaneously delineate the critical structures can then be very useful to have both accurate and reproducible segmentations in radiotherapy planning systems such as Isogray from DOSIsoft company. The creation [1] and the use [2] of a brain anatomical atlas to segment brain critical structures has been recently studied. In these articles, an individual atlas is used: a simulated MRI has been segmented manually. The atlas-based segmentation is then performed by a MRI to MRI registration with the simulated image.

However, brain cancers represent only a small part of the cancers. A major localization is the head and neck (7 % of all cancers). It would then be of great interest to use an anatomical atlas of this area to help the clinicians with the therapy planning. However, contrary to the brain case, the critical structures (mainly lymph nodes) are not visible structures and their limits are not well defined in the images. Moreover, imaging protocols do not generally include MRI as for the brain. Only CT images are then available, implying a poor definition of soft tissues. Using an individual atlas as for the brain seems then not to be the good solution since the structures to be delineated are not clearly defined in the images. Using only one manually delineated CT image or another may introduce an undesirable bias. Some papers have been published, like [3], reaching a consensus on the way to delineate the head and neck lymph nodes for radiotherapy planning. We will then build an atlas from a group of patients manually delineated following these guidelines.

Methodologies have been devised recently to create a mean image from a set of patients images. [4] introduced a framework to create an unbiased mean image from a database of patients. [5] improved this framework to cope with transformations including large deformations. Finally, [6] proposed a coupled estimation of the mean segmentations and the mean image. However, none of them dealt with the choice and the evaluation of the algorithm used to register the database images in order to obtain the best achievable atlas.

Many registration techniques can indeed be used to register an atlas on a patient. These algorithms can use a dense transformation either using fluid regularization [7] or using a inhomogeneous visco-elastic regularization [8]. Other algorithms prefer to parameterize the transformation using less degrees of freedom, like using B-Splines [9] or locally affine transformations [2]. The range of registration methods, and thus of parameters, for building and registering an atlas is therefore very important. To deal with this wide range of methods, it is mandatory to have a framework to evaluate quantitatively the best registration parameters and method to build an atlas.

In this article, we present, in a first part, a framework to build an atlas from a database of images which have been manually delineated. This framework is then associated, in a second part, with a methodology to evaluate quantitatively both the best registration method to build the atlas and to register the patient on the atlas. We finally present results of head and neck atlas construction and evaluation we conducted using our framework on a real dataset of eight manually segmented CT images.

II. Atlas Building

At this point, we first need to choose a methodology to build an atlas from a dataset of segmented images. We chose not

Corresponding author: O. Commowick, email: ocommowi@sophia.inria.fr, Web site : http://olivier.commowick.org/

to use the method proposed by [6] because the segmentations are user dependent and can vary from one patient to another depending on his morphology. Using these segmentations in the registration process may introduce a bias in the resulting atlas. We thus choose to use a more classical approach, which is decoupled in two steps: the construction of a mean image as unbiased as possible with respect to the image dataset, and afterwards computing mean segmentations from the individual segmentations.

A. Mean Image Construction

We choose to use the method proposed by [4] to build the atlas. This method has the advantage to be faster and simpler than the method proposed by [5]. It can perform well as our images are all acquired using the same patient position, avoiding very large deformations between the images. This methodology simply amounts to choosing a reference image among the dataset and do a first registration of the other images on this reference. We are then able to produce a mean image from the obtained transformations. To ensure that we build an unbiased atlas, we then iterate this process by taking each time the resulting mean image of the current iteration as the reference image for the registration.

B. Registration Methods

Each non-rigid registration is preceded by a global affine matching in order to bring the images in global correspondence. We choose to build our atlas using three different non-rigid registration methods, in order to evaluate the differences induced by the registration method used to build the atlas. The first two methods are presented below. The third method is a combination of the two first ones: we use a locally affine registration followed by a dense registration with a larger smoothing as we are closer to the solution.

1) Locally Affine Registration: The first method we used to register the images is the locally affine registration, which has been proposed by [2]. For this method, only user specified areas have to be registered. Each area is then registered using a local affine transformation. The global transformation is interpolated between the areas using weight functions defined from the areas. The fast polyaffine framework then insures a smooth transformation between the areas. This alorithm is able to give very smooth transformations with an accuracy comparable to dense transformation algorithms.

2) Dense Registration: The second registration method is an extension to dense transformation of Block-Matching based rigid registration [10]. We evaluate a dense transformation iteratively using a multi-resolution scheme. At each iteration i, we use the pairings between the images to compute a sparse vector field U_i. This field is associated to a sparse confidence field k_i, which is made from the actual values of the correlation coefficient for each pairing. We then interpolate a correction ΔT for the current transformation T: $\Delta T = (G_\sigma * (k_i U_i)) / (G_\sigma * k_i)$. This ensures a smooth transformation, which will be close to the pairings we have a good confidence in, and will be more interpolated anywhere else. Finally, we do an outlier rejection by comparing ΔT and U_i. If the norm $\|U_i(x) - \Delta T(x)\|$ on the pairing points is greater than an automatically defined threshold, then the pairing is removed from the dense field calculation at iteration i. At the end of the iteration, T is updated by composing it with ΔT. We have chosen to use this method for the dense registration because it is able to produce smooth deformation fields and can cope with large deformations.

C. Mean Segmentations Computation

At this point, we have a manual segmentation for each image and a transformation between the mean image and each of the individual images. This gives us a way to put all the segmentations in the same geometry. We will now focus on obtaining the mean segmentations from the manual segmentations of the database. A classical way to obtain these segmentations is to do a classical mean of the manual delineations. However, using this method has several drawbacks for our application. First, manual delineations are not reproducible. We thus can have some delineations which are not the same, independently of the quality of the registration algorithm. Using a simple mean will not discard this type of problem. An other main drawback is for multi-label segmentations as in our case. Using a simple mean for labels that are close to each other can produce overlapping mean segmentations. This is not satisfactory as we want to have separated areas or a probability for each voxel to belong to one label or another.

To overcome these drawbacks, we then choose to use the method proposed by [11]. This method uses a set of multi-label segmentations to produce a multi-label "ground truth", and the sensitivity and specificity parameters for each manual segmentation with respect to the ground truth. This is done using an Expectation Maximization algorithm: the Expectation step estimates the hidden parameter (ground truth) while the "performance" parameters (sensitivity and specificity generalized to the multi-label case) for each segmentation are estimated during the Maximization step. This approach gives a resulting probabilistic segmentation for each label and then ensures to have well separated mean segmentations.

III. Atlas Construction Method Evaluation

We have seen so far how to obtain an atlas from a dataset of images. We present now a simple method to evaluate the goodness of the atlas with respect to the registration method used to build it and to the method used to register the patient image on the atlas. This method relies on a Leave-One-Out process.

In this process, we select one of the images of the database and its segmentation and put it aside. Then, we build the atlas using the rest of the images. This produces a mean image associated with a mean segmentation. The remaining patient image can then be registered on the atlas and delineated from it. Finally, we can compute any quantitative measure between the automatic segmentation obtained by atlas registration and the manual segmentation done on the patient. We choose here to use the couple sensitivity/specificity [11] to measure the quality of the algorithm rather than a simple overlap measure. This couple of measures gives indeed more information on the way the automatic segmentation went wrong (more false negatives or more false positives). We also compute the distance from this couple of measures to the best achievable measure (Sensitivity = 1, Specificity = 1). This is defined as the norm of $(1 - Sens., 1 - Spec.)$. We compute this last measure to give a simplified idea of the quality of the result.

Using this framework, we can compute performance parameters on several patients, by picking them alternatively from the database. To evaluate the overall goodness of one registration method, we then compute the means of sensitivity and specificity on all the patients we left out of the construction. The overall distance to the best achievable measure is obtained as the mean of the distance to $(1,1)$ for each patient.

Using this method allows us to evaluate the quality of the atlas with respect to the registration method used to build it. It is also useful to evaluate the best registration method to register a patient on the atlas, given an atlas. To achieve this, we can register the discarded patient on the atlas using different registration methods and then see the best adapted to our problem.

IV. Results

We have used our framework to build an atlas using eight patients. All these patients have been imaged in a CT scanner and structures of interest for the radiotherapy planning have been manually delineated. These structures are mainly a simplified set of those proposed in [3]: the spinal cord, the mandible, the two sub-mandible glands, the two parotid glands and the node levels on both sides of the neck.

Fig. 1. **Atlas Examples obtained using our framework.** From left to right: Sagittal slices of the atlas obtained using the first, second and third method (see section II-B) with the mean structures contours superimposed.

We first test our method to build the atlas using all the eight images. We present in Fig. 1 the mean images with the mean structures contours superimposed. The mean images were obtained using the three methods presented in section II-A. We can see on these images that we managed to obtain qualitatively good atlases. We have indeed a good match between the mean structures obtained and the mean images. We can see also that the contours are not intersecting, showing that the mean structures are well separated thanks to the Staple algorithm. The contours are a little smoother on the atlas using the first method, but seem a little less accurate than for the other atlases. We can also see (arrows on Fig. 1), that there are effectively changes in the mean structures depending on the algorithm used to build the atlas.

We have then used the method presented in section III to evaluate which method of the three we presented is the best to build the atlas and register the patients on it. The Leave-One-Out test was then repeated by picking seven different patients alternatively from the database. The mean sensitivity and specificity have been evaluated over all these tests. The results are shown in Table I, where the columns stand for the registration method used to register the patient on the atlas and the lines stand for the registration method used to build the atlas. We also reported in this table the mean distance to the best achievable measure (Sensitivity = 1, Specificity = 1), as defined in section III.

Atlas Const. \ Reg. Meth.	DT			LAF			Both		
	Sens.	Spec.	Dist.	Sens.	Spec.	Dist.	Sens.	Spec.	Dist.
Dense Transformation (DT)	0.797	0.891	0.253	0.837	0.856	0.236	0.852	0.857	0.224
Locally Affine (LAF)	0.870	0.846	0.219	0.831	0.851	0.241	0.850	0.851	0.227
Both	0.868	0.854	**0.216**	0.836	0.858	**0.236**	0.853	0.857	**0.223**

TABLE I

Mean quantitative measures using the three different methods for building and registering the atlas. Columns: registration method used for registering the patient on the atlas. Sens. stands for sensitivity, Spec. stands for specificity and Dist. stands for the distance to the best achievable measure (1,1). The lines stand for the method used to build the atlas (see text).

We can see in this table that the results are quite good and similar, which means that none of the methods failed completely. However, we can notice (in bold) that the third method (locally affine followed by dense transformation) performs always better than the other methods for building the atlas. The best result was obtained here by building the atlas using the third method and registering it using the dense transformation algorithm. However, other results are very close to this one, like LAF atlas followed by DT registration or third method atlas followed by third method registration. The LAF method performs a little less good than the others. This is partly due to the soft tissues we have segmented. The deformations of these soft tissues can be indeed hard to model using local affine transformations. This can result in some misregistrations on these areas giving less good sensitivity and specificity measures.

Finally, we want to see qualitatively the results of the registration of a patient on the atlas. As the images used for building and testing the atlas are acquired with nearly the same patient position, the deformations and displacements are not as important as in a real case, where the patient position (mainly for the spine) can be very different from the atlas position. In this case, we want the algorithm used to register the atlas to be robust with respect to the differences in the position of the patient. To get a first idea, we test the registration of one of the patients on the atlas given by the others.

The results are shown in Fig. 2. We can see on this figure that the difference in positions is not really important. We indeed only have a small deformation of the spine. However,

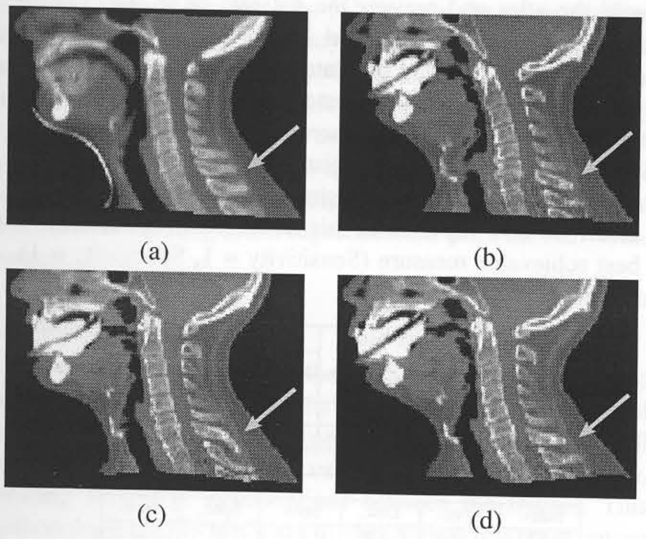

Fig. 2. **Registration of a patient on an atlas.** (a) : Atlas sagittal slice, (b) : patient registered with a global affine registration, (c) : patient registered with our dense transformation algorithm and (d) : patient registered with a locally affine transformation.

our dense transformation algorithm may encounter problems on specific areas, like at the base of the spine. This can become a problem if we encounter a more important deformation for the neck region of patients coming from other hospitals. On the other side, the locally affine algorithm seems more robust for these deformations. The spine is indeed deformed with local affine transformations, as the vertebrae are rigid structures. The locally affine algorithm is thus well designed for this type of registration and hence more robust, as we can see on Fig. 2. Using the second or the third method to register a patient not in the database on the atlas seems then a better solution given this result and the relatively close results in Table I.

V. Conclusion

We have presented in this article a method to assess the registration methods for atlas construction in the context of atlas-based segmentation. This framework allows to evaluate the results of both the atlas construction and the registration of a patient on the atlas. This is done through a Leave-One-Out process on the database of images. We used this framework to evaluate the construction of a head and neck atlas from a group of patients. The obtained anatomical atlas has been implemented in the Isogray software for radiotherapy planning from DOSIsoft company. This atlas will allow to segment automatically the lymph nodes and the other critical structures in the neck for radiotherapy planning.

We have shown thanks to this validation method that our atlas performs well on the patients used in the database. This is true for all the methods used to build and register the atlas. However, we have shown in our experiments some desirable qualities of certain methods to register the patient on the atlas and a little superiority of the third method presented. This experiments should however be extended, to be completely valid, to more registration methods and even more atlas construction frameworks. This can be done in a simple manner as we already have the comparison framework.

We intend first to extend the atlas construction on a larger database. Then, we will focus in a near future to conduct experiments on real cases which do not come from the database and evaluate quantitative results for it. This seems an important task to validate the robustness of the registration method with respect to different neck positions in the images. The goal of our atlas is also to segment automatically more structures to get closer to the guidelines. We would like mainly to separate the lymph nodes that are only one structure for the moment in the different areas defined in [3].

Finally, we have seen, even on our limited database, a large amount of change between the morphologies of the patients. It seems then very relevant to search for morphology groups in the patient database. We could then use several atlases instead of just one mean atlas. A method to build several atlases from an image database could therefore be very interesting to build a more patient-adapted atlas.

Acknowledgments

This work was partially founded by ECIP project MAESTRO (IP CE503564) and ANRT. The authors are grateful to Dr. V. Grégoire for providing its expertise, the image database and their manual delineations.

References

[1] PY Bondiau, G Malandain, O Commowick, PY Marcy, S Chanalet, and N Ayache, "Atlas-based automatic segmentation of MR images: Validation study on the brainstem in radiotherapy context," in *RSNA*, Chicago, 2004.

[2] Olivier Commowick, Vincent Arsigny, Jimena Costa, Nicholas Ayache, and Grégoire Malandain, "An efficient locally affine framework for the registration of anatomical structures," in *Proceedings of ISBI 2006*, April 2006, pp. 478–481.

[3] Vincent Gregoire, Peter Levendag, Kian K Ang, Jacques Bernier, Marijel Braaksma, Volker Budach, Cliff Chao, Emmanuel Coche, Jay S Cooper, Guy Cosnard, Avraham Eisbruch, Samy El-Sayed, Bahman Emami, Cai Grau, Marc Hamoir, Nancy Lee, Philippe Maingon, Karin Muller, and Herve Reychler, "CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines," *Radiother Oncol*, vol. 69, no. 3, pp. 227–36, December 2003.

[4] A. Guimond, J. Meunier, and J.-P. Thirion, "Average brain models: A convergence study," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 192–210, 2000.

[5] Peter Lorenzen, Brad Davis, and Sarang C. Joshi, "Unbiased atlas formation via large deformations metric mapping.," in *MICCAI (II)*, 2005, pp. 411–418.

[6] M. De Craene, A. du Bois d'Aische, B. Macq, and S.K. Warfield, "Multi-subject registration for unbiased statistical atlas construction," in *MICCAI*, September 2004, pp. 655–662.

[7] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE TIP*, vol. 5, no. 10, pp. 1435–1447, 1996.

[8] R. Stefanescu, X. Pennec, and N. Ayache, "Grid powered nonlinear image registration with locally adaptive regularization," *MedIA*, vol. 8, no. 3, pp. 325–342, September 2004.

[9] D. Rueckert, L. L. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE TMI*, vol. 18, no. 8, pp. 712–721, 1999.

[10] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *MICCAI*, 2000, pp. 557–566.

[11] Simon K Warfield, Kelly H Zou, and William M Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE TMI*, vol. 23, no. 7, pp. 903–21, July 2004.