Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation

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

[1] INRIA Sophia Antipolis - Asclepios Team, 2004 Rte des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
Olivier.Commowick@sophia.inria.fr,
[2] DOSIsoft S.A, FRANCE, 45-47 Avenue Carnot,
94 230 Cachan, France

Abstract. Radiotherapy planning needs accurate delineations of the critical structures. Atlas-based segmentation has been shown to be very efficient to delineate brain structures [1]. However, the construction of an atlas from a dataset of images [2], particularly for the head and neck region, is very difficult due to the high variability of the images and can generate over-segmented structures in the atlas. To overcome this drawback, we present in this paper an alternative method to select as a template the image in a database that is the most similar to the patient to be segmented. This similarity is based on a distance between transformations. A major contribution is that we do not compute every patient-to-sample registration to find the most similar template, but only the registration of the patient towards an average image. This method has therefore the advantage of being computationally very efficient. We present a qualitative and quantitative comparison between the proposed method and a classical atlas-based segmentation method. This evaluation is performed on a subset of 45 patients using a Leave-One-Out method and shows a great improvement of the specificity of the results.

1   Introduction

Conformal radiotherapy requires a very accurate planning in order to optimize the irradiation dose on the tumor while controlling the doses on critical structures. This task requires their precise delineation. However, it is yet performed manually and is therefore time consuming and tedious. Recently, the creation and use of an anatomical atlas (an image of the anatomy, typically MRI or CT, associated to its segmentation) has been explored to delineate brain critical structures [1]. This atlas-based segmentation is decomposed in two steps: first, the patient image is registered on the atlas image ; then, the transformation is applied to the atlas structures to obtain the segmentation. This approach has shown good results on the brain. It would therefore be of great interest to use a similar approach on the head and neck region where 7% of the cancers arise.

Two main approaches can be used to build an anatomical atlas. The first one (i) consists in using a single image delineated by an expert. For example,

it is possible to use a symmetric atlas built from the BrainWEB[3] and manually delineated, as presented in [1]. The second method (ii) is to build an average image and its segmentation from a dataset of manually delineated images.

We have chosen this last solution (ii) in [2] to build an atlas of the head and neck region from images delineated following the guidelines provided in [3]. However, anatomical variability is very high in this region, particularly on the spinal cord or the lymph nodes areas. The generated mean contours can thus be too large in the atlas yielding over-segmentations. The atlas can also be very different from the patient to segment. Registration discrepancies can therefore appear leading to segmentation errors.

To overcome these drawbacks, a previous work in the literature [4] presented an interesting approach that clusters the database of delineated images to build several atlases representing homogeneous sub-populations. However, the selection of the most adequate atlas for a given patient is not addressed. Moreover, a large number of samples is required in the database to build meaningful clusters.

In this paper, we aim at segmenting the critical structures in a patient image P using a database of manually delineated images. Since our database has too few samples to build meaningful clusters, we propose to find the most similar database image to the patient image to be segmented, and to register it non linearly on this image. The key point is then to efficiently select among a database the most similar image to a given patient image, up to an affine transformation. This most similar sample is defined as the one that needs the smallest local deformations to be registered on the patient image to be segmented.

The remainder of the article is organized as follows. We will first present our approach to select the image that is the most similar to the patient to delineate. This method requires to build a mean atlas from a database of delineated images. We will then focus on the method that is used to build an average image and its average segmentations from the database. Finally, we will show some qualitative and quantitative results on a database of 45 Head and Neck CT images, using a Leave-One-Out evaluation methodology.

2   Method

2.1   Selection of the Most Similar Image

In this paper, we aim at selecting the most similar image among the database and use it as the atlas. As mentioned above, the most similar image is defined as the one that will be the "less" deformed to be non linearly registered on the patient. The selection is then based on a comparison of the non linear transformations $T_{P \to I_k}$ to bring P on each image I_k, i.e the most similar image is defined as:

$$\tilde{I} = \arg\min_{I_k} d(I_k, P) = \arg\min_{I_k} \|T_{P \to I_k} - Id\|. \tag{1}$$

---

[3] http://www.bic.mni.mcgill.ca/brainweb/

However, this comparison is computationally very expensive as it requires to perform all the registrations between P and the images I_k. To reduce this computation time, we suppose that we are using an average image M built from the images of the database (see section 2.2). From the atlas construction, we indeed obtain for each image I_k a non linear transformation $T_{M \rightarrow I_k}$ bringing it on the average image. When registering P on M, we also obtain a non linear transformation $T_{M \rightarrow P}$. The key hypothesis of our work is then to assume that $T_{P \rightarrow I_k}$ can be approximated by $T_{P \rightarrow I_k} \approx T_{M \rightarrow I_k} \circ T_{M \rightarrow P}^{-1}$. Using our hypothesis, the similarity between P and I_k can then be evaluated, up to an affine transformation, using the following equation:

$$d(I_k, P) = \|T_{M \rightarrow I_k} \circ T_{M \rightarrow P}^{-1} - Id\| = \sum_i \|T_{M \rightarrow I_k} \circ T_{M \rightarrow P}^{-1}(i) - Id\| \quad (2)$$

where i corresponds to the voxels of the dense transformation. As $T_{M \rightarrow I_k} \circ T_{M \rightarrow P}^{-1}$ is close to the identity, the use of Euclidean norm is reasonable. This measure could be also computed using a Log-Euclidean distance [5].

From the average image construction, the non linear transformations $T_{M \rightarrow I_k}$ have already been computed. Using our hypothesis, we now only need to perform one non linear registration between M and P to select the most similar image to the patient to delineate. Finally, the patient P is delineated by registering it on $\tilde{I}$ and applying the transformation to the structures of $\tilde{I}$.

2.2   Construction of an Average Image

To select the most similar image, we need to build an atlas from the database of images. Many methods have been explored to perform this task. Guimond et al. [6] introduced a framework to create an unbiased mean image from a database of patients. Lorenzen et al. [7] improved this framework to cope with large deformations. De Craene et al. [8] proposed a coupled estimation of the average segmentations and the average image. Finally, Grabner et al. [9] presented an interesting method to compute directly an average symmetric image.

We use in this paper a method that was already presented in [2]. As in that paper, we did not use the method of [8] because of the variability of the manual segmentations in the dataset. The construction is composed of three main steps. First, we build an average image using the method developed in [6]. This method has the advantage of being faster and simpler than the method proposed in [7]. This first step produces for each image I_k a transformation $\tilde{T}_{M \rightarrow I_k}$ bringing it on the average image. Then, the atlas image is symmetrized in two steps. First, the method of Prima et al. [10] is used to find the inter-hemispheric plane in the average image. Then, it is symmetrized with respect to this plane.

In addition, we compute in [2] average structures from the manual delineations. These are first brought on the average image using the $\tilde{T}_{M \rightarrow I_k}$ transformations. Then, we use the STAPLE algorithm [11] in its multi-label implementation to compute non-overlapping most probable average segmentations. These delineations are then symmetrized in the same way as for the average image.

3   Evaluation Methodology

To evaluate our method, we have chosen to use a Leave-One-Out method. It consists in picking out one patient from the dataset of delineated images. The atlas is then built from the remaining images using the method presented in section 2.2. Two methods were compared for the automatic segmentation. First, the classical atlas-based segmentation method was used to segment automatically the remaining image. This image is therefore registered on the average atlas and the transformation is applied to the structures to get the segmentation. Then, we used our most similar image method, i.e. the most similar image among the remaining samples of the database is registered on the left out patient, using the same registration algorithms and parameters, to get its segmentation.

The results obtained by each of those methods can then be compared to the manual delineations of the remaining patient using quantitative measures. In this paper, we have chosen to compute the sensitivity and specificity measures to evaluate the quality of the automatic segmentation methods.

The evaluation has been performed by following this Leave-One-Out methodology on a database of 45 CT images of patients delineated for head and neck radiotherapy following the guidelines provided in [3]. As they were delineated for radiotherapy, some structures were not delineated or fused with other structures. Some structures are therefore not available for some patients. We have therefore used for the Leave-One-Out evaluation a subset of only 12 patients which were completely delineated manually.

4   Results

In this section, we first focus on qualitative results of the most similar image selection and of segmentation of one patient left out of the dataset. Then, we present quantitative results using the evaluation method presented above.

4.1   Qualitative Results

We first show in Fig. 1 a qualitative view of the image selected as the most similar compared to the atlas and to the patient to segment. Our method is dedicated to select the most similar image up to a global affine transformation and we therefore present in this figure all the images registered on the patient to segment using a global affine transformation.

First, the patient and the atlas are very different. There is indeed an important difference between their morphologies. These differences often lead to registration discrepancies, that can cause segmentation errors. On the contrary, the most similar image is much closer to the patient than the atlas and should therefore be much easier to register.

The second part of our qualitative validation was then to compare the results of classical atlas-based segmentation and our method of most similar image
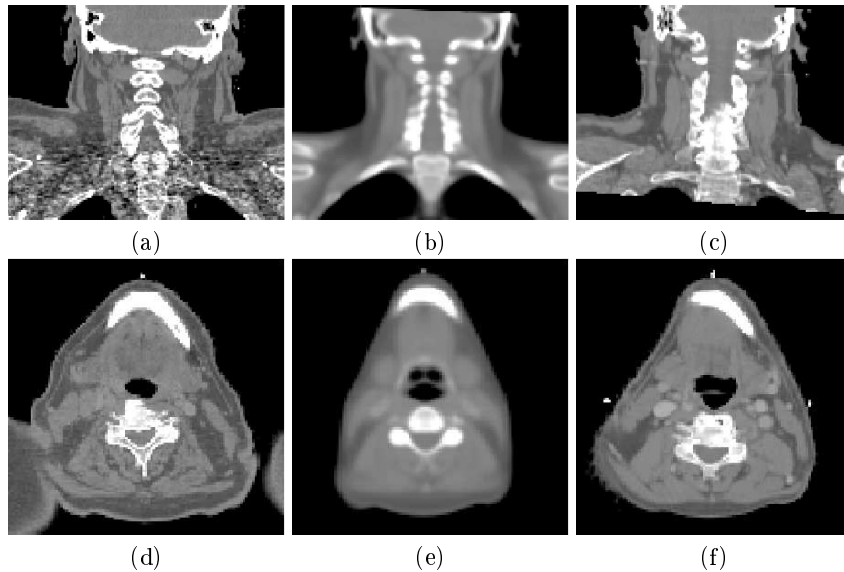
Fig. 1. Example of selected most similar image. Comparison between the atlas, most similar image and the patient. These images are affinely registered on the patient (see text). (a), (d): Slices of the patient. (b), (e): Corresponding atlas slices. (c), (f): Most similar image slices. Upper line: Coronal slices. Bottom line: Axial slices.

based delineation. We therefore show in Fig. 2 coronal and axial views of the segmentation of an other patient left out of the dataset.

We first see in this figure (arrows on (b)) that the results obtained from the atlas are over-segmented when compared to the manual segmentations. This is due, as mentioned in the introduction, to the variabilities between contours when creating the atlas structures. This variability indeed results in over-segmented mean structures in the atlas itself. In the most similar image based segmentation, the results are not over-segmented anymore, as the atlas is now made from one single image, and are much closer to the real structures, particularly on lymph nodes and parotids. However, some regions such as the lymph nodes areas IV (arrows at the bottom of the images) and the right parotid (arrows at the upper left of the images) still have some errors. These are mainly due to the intra-expert, inter-patient variability in the manual delineation of the structures. We show in Fig. 3 an example illustrating these segmentation differences on the images after a global affine registration.

This figure clearly shows that the manual delineations of the most similar image used for Fig. 2 are different from those of the patient to segment. First, the right parotid (arrows on axial and coronal views) incorporate a small region near the spinal column in the most similar image and not in the patient to segment. Moreover, the lymph node areas (arrows on the coronal views) also incorporate extra regions at the bottom that we can find back om image (c) and (f) in Fig. 2. Finally, there is also a smaller variability on the spinal cord delineation.
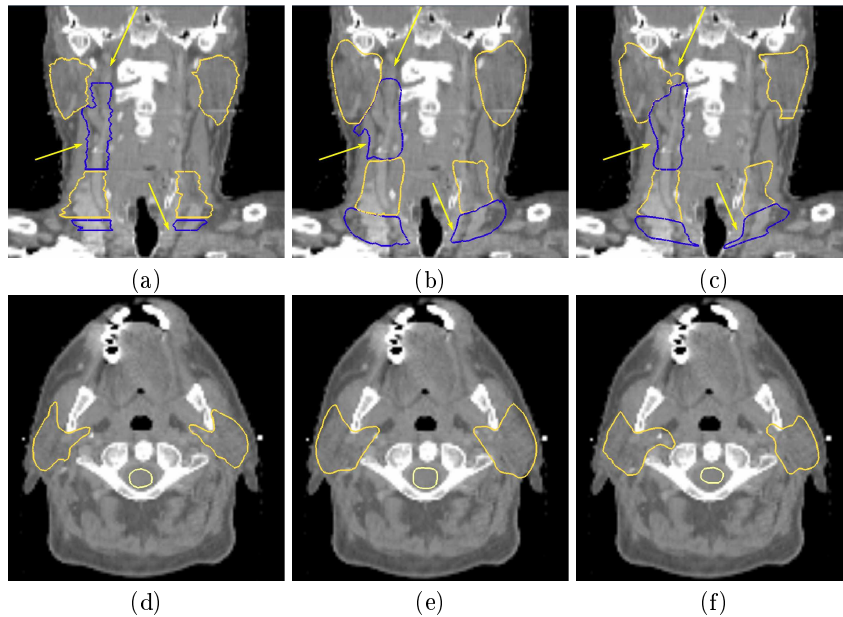
Fig. 2. Qualitative atlas and most similar image based segmentation results.
(a), (d): Manual segmentation of the patient. (b), (e): Atlas-based segmentation of the
patient. (c), (f): Segmentation using the most similar image in the database. Upper
line: Coronal slices. Bottom line: Axial slices.
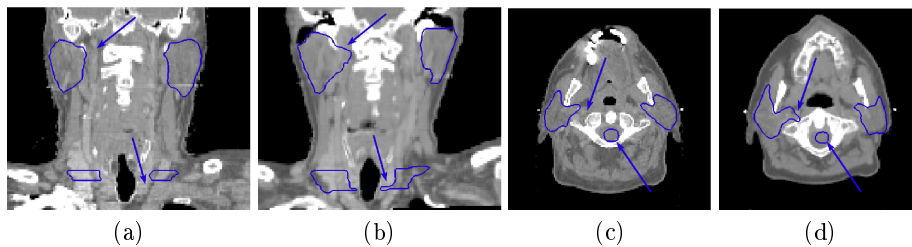
Fig. 3. Contours variability between the dataset images. Example showing
views of the manual delineations on the patient presented in Fig. 2 (a), (c) compared
to the manual contours on the most similar sample (b), (d). These images are affinely
registered on the patient image (see text).

4.2   Quantitative Evaluation

We have seen above that the qualitative results were good, even if there were
some errors due to contours variabilities in the database. We now present the
quantitative evaluation described in section 3 over a subset of 12 patients fully
delineated. We show in Table 1 the mean quantitative measures computed over
the structures of the 12 patients. We have added in this table a third measure
called distance, which corresponds to the distance to the best sensitivity and

specificity achievable, i.e. $d = \|(1 - Sens., 1 - Spec.)\|$. This last measure is computed to give a simplified idea of the quality of the result.

	Sensitivity	Specificity	Distance
Atlas-based segmentation	0.827	0.684	**0.389**
Most similar image	0.675	0.849	**0.380**

Table 1. Mean quantitative measures using the two different segmentation methods. Columns stand for the quality measures. Distance corresponds to the distance to the best achievable measure ($Sensitivity = 1, Specificity = 1$). The lines stand for the method used to delineate the images (see text).

This table shows first a great improvement in specificity when using our method, which confirms our first qualitative observations in the preceding results. However, the sensitivity is also lower than the one of the atlas-based segmentation. This is again due to the high variability of the contours in the dataset, as we have seen in Fig. 3. Nevertheless, the results are still very promising and the global measure for the most similar image based segmentation is a little better than the one obtained by the atlas.

5 Conclusion

We have presented in this paper a method to select up to a global affine transformation the most similar image to the patient to be delineated. This method is based on the use of a distance between the transformations used to build the atlas and the transformation computed to register the patient on the atlas. As the atlas is computed once and for all patients to be delineated, this method is therefore very easy to implement and very efficient. It indeed only requires one more registration when compared to classical atlas-based segmentation.

This method has been validated on a subset of 12 patients in a database of 45 patients using a Leave-One-Out method. The obtained segmentations are not over-segmented anymore. However, we have seen through our experiments that the results are corrupted by an important intra-expert, inter-patient variability of the manual segmentations. One way to cope with this problem would be to use a database with repeated segmentations made by several experts. This would on one side help to quantify this delineation variability (inter and intra-expert variability) and also to obtain better results.

This variability also explains why we obtain too large contours when building the atlas, as mentioned in the introduction. The coupling could therefore also be done when creating the atlas. The contours and their variability could then be introduced in the building process to better constrain the atlas creation. Previous work in this direction include the work of [8], where the binary structures were included in the atlas formation process.

To handle the variability of contours, the use of several atlases representing sub-populations as in [4] could also be very interesting. An important perspective

to this work is then to use an extension of our approach to find efficiently sub-populations in the database. In addition to this efficiency, the selection of the most similar average atlas would then be solved using our selection method.

Finally, more validation is to be added to this work. The quantitative results shown here are indeed computed on the structures that are both present in the most similar image and in the left out patient. As the structures of the dataset were delineated for radiotherapy, some structures are missing and therefore the validation is not done on all structures. More validation on more structures and on images acquired in clinical conditions in different centers would then be a major step towards clinical validation and use.

Acknowledgments

References

1. Bondiau, P.Y., Malandain, G., *et al.*: Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. Int J Radiat Oncol Biol Phys **61**(1) (2005) 289–98
2. Commowick, O., Malandain, G.: Evaluation of atlas construction strategies in the context of radiotherapy planning. In: SA2PM Workshop (From Statistical Atlases to Personalized Models). (2006) Held in conjunction with MICCAI 2006.
3. Grégoire, V., Levendag, P., *et al.*: CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. Radiotherapy Oncology **69**(3) (2003) 227–36
4. Blezek, D.J., Miller, J.V.: Atlas stratification. In: MICCAI (I). Volume 4190 of LNCS. (2006) 712–719
5. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A Log-Euclidean framework for statistics on diffeomorphisms. In: MICCAI (I). Volume 4190 of LNCS. (2006)
6. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. Computer Vision and Image Understanding **77**(2) (2000) 192–210
7. Lorenzen, P., Davis, B., Joshi, S.C.: Unbiased atlas formation via large deforma-tions metric mapping. In: MICCAI (II). Volume 3750 of LNCS. (2005) 411–418
8. De Craene, M., du Bois d'Aische, A., Macq, B., Warfield, S.: Multi-subject regis-tration for unbiased statistical atlas construction. In: MICCAI'04. Volume 3216 of LNCS. (2004) 655–662
9. Grabner, G., Janke, A.L., Budge, M.M., Smith, D., Pruessner, J., Collins, D.L.: Symmetric atlasing and model based segmentation: An application to the hip-pocampus in older adults. In: MICCAI (II). Volume 4191 of LNCS. (2006) 58–66
10. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3D brain images. IEEE Transaction on Medical Imaging **21**(2) (2002) 122–138
11. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging **23**(7) (2004) 903–921