Segmentation of head bones in 3-D CT images from an example

Sylvain Faisan, Nicolas Passat, Vincent Noblet, Renée Chabrier, Jean-Paul Armspach, Christophe Meyer

SEGMENTATION OF HEAD BONES IN 3-D CT IMAGES FROM AN EXAMPLE

S. Faisan, N. Passat, V. Noblet — LSIIT, UMR ULP-CNRS 7005, Strasbourg, France

R. Chabrier, J.-P. Armspach — LINC, UMR ULP-CNRS 7191, Strasbourg, France

C. Meyer — University hospital of Besançon, France

ABSTRACT

We propose a new generic framework for segmentation of 3D digital data, based on knowledge contained in a segmentation example of similar data. The integration of prior knowledge is made by registering the image to segment on the segmentation example. Since the registration step relies on binary segmented data, segmentation and registration are performed jointly in a coarse-to-fine way using a multiscale parametric representation of a threshold map and of a deformation field. The threshold map is required by the segmentation procedure which is also devoted to recover topological details. The benefit of such an approach is illustrated in the context of head bone segmentation in 3-D computed tomography (CT) images.

Index Terms— Image segmentation, image registration.

1. INTRODUCTION

Devising reliable and fully automatic segmentation methods for head bones in 3-D CT images is still an open issue, and few methods in the literature are addressing this problem [1, 2]. One of the key points making this task difficult is related to the nature of CT images, which are subject to noise and to severe streaking artifacts due to metal objects such as dental fillings. Another issue is to separate "bones of interest" from the other ones. For example, in the context of this work, bones of the spinal column should not be segmented. These two points are usually tackled using pre- or post-processings: in [2], a metal artifact removal procedure is applied on the CT images before segmentation, and in [1], a post-labelling of the structures of interest is performed after segmentation. Tackling all these issues during the segmentation process would require the introduction of strong priors. However, extracting knowledge remains a difficult task. That is why segmenting images based on available segmentation examples is of increasing interest in image processing community. The key point is to use one or several manually labelled examples to improve the segmentation of a similar image [3].

In this paper, we consider the segmentation of a 3-D CT image I of the head, while taking advantage of a reference segmentation B_m of another similar CT image (B_m has been segmented manually by an expert). In the sequel, all binary images are denoted by B (if a voxel $\boldsymbol{s}$ is classified as bone of interest, $B(\boldsymbol{s})$ is set to 1, and to 0 otherwise). To use the knowledge provided by B_m, we follow an idea similar to [3], which consists in finding correspondences between parts of B_m and parts of I. To this end, the transformation that best maps I onto B_m should be estimated. Most registration methods dedicated to the maxillofacial region are feature-based. This implies that features have to be extracted *before* registration. However, as the feature extraction step requires to segment the image, a chicken-and-egg problem arises: taking into account B_m to segment I requires to register I on B_m and the registration of I on B_m requires to segment I. In order to tackle this problem, we propose to jointly perform segmentation and registration. The benefit is to progressively improve, on the one hand, the segmentation quality thanks to the knowledge provided by the registration procedure, and on the other hand, the registration accuracy thanks to a better segmentation of I. The paper is organized as follows. In Section 2, the proposed method is described. Section 3 presents results. Conclusion and perspectives are given in Section 4.

2. PROPOSED FRAMEWORK

In the proposed method, segmentation and registration are performed in a coarse-to-fine way based on two multiresolution parametric models: one for the non-rigid transformation, and one for the threshold map. At scale l, the deformation model is estimated by mapping the segmented version of I obtained at scale l-1, namely B^{l-1}, on B_m. Since the mapping of I on B_m becomes more accurate as the scale increases, it can be assumed that the knowledge extracted from B_m to drive the segmentation of I becomes more and more confident (extracted features enable to estimate the threshold map, which is used during the segmentation procedure). The different steps of the algorithm, summarized in Alg. 1, are described in the following subsections.

2.1. Registration

The purpose of registration at scale l is to estimate the parameters of the deformation model so as to match B^{l-1} on B_m given a cost function. Before describing the cost function, we present the multiresolution deformation model.

Algorithm 1 Joint segmentation and registration (notations are defined in the text)

Initialization of segmentation using Otsu's method: B^0
for $l = 1$ to l_f **do** {/*registration loop*/}
  Registration at scale l (Sec. 2.1): $\boldsymbol{h}^l$
  Knowledge extraction (Sec. 2.2): Z^l_{boi}, Z^l_{ref}, $\{p^{l,r}_{i,j,k}\}$
  Estimation of the threshold map (Sec. 2.3): $I^{l,0}_t$
  Initialization of segmentation (Sec. 2.4): $B^{l,0}$
  for $r = 1$ to $l-1$ **do** {/*segmentation loop*/}
    Estimation of the threshold map (Sec. 2.3): $I^{l,r}_t$
    Refinement of the segmentation (Sec. 2.4): $B^{l,r}$
  end for
  Fine topology details recovering (Sec. 2.4): B^l
end for

Let $\boldsymbol{s} \stackrel{\Delta}{=} [x, y, z]^t$ be a voxel of an image defined on $\Omega = [0, 1]^3$. The mapping $\boldsymbol{h}$ between the source and the target image writes $\boldsymbol{h}(\boldsymbol{s}) = \boldsymbol{s} + \boldsymbol{u}(\boldsymbol{s})$, where $\boldsymbol{u}$ is the displacement vector field. We consider a decomposition of the displacement vector field $\boldsymbol{u}$ over a sequence of nested subspaces $V_1 \subset V_2 \subset \ldots \subset V_{l_f-1} \subset V_{l_f}$, defining a multiresolution approximation of $\boldsymbol{u}$. Space V_1 defines the coarsest scale representation. A basis of V_l may be generated from a scaling function Φ. To handle a 3-D deformation field, three multiresolution decompositions are considered, one for each component of the displacement. First degree polynomial spline scaling function Φ is considered in this work [4] so that the support $\Omega^l_{i,j,k}$ of the spline, $(i, j, k) \in [1 \ldots 2^l - 1]^3$, is a cube of size $2^{1-l} \times 2^{1-l} \times 2^{1-l}$ (the image is defined on $\Omega = [0, 1]^3$).

The parameters of the deformation model are estimated by minimizing a cost function. The cost function quantifies the distance between two binary images, namely $B^{l-1}(\boldsymbol{h}^l)$ and B_m. This problem is often tackled by considering only surfaces of the segmented structures and by using the Iterative Closest Point (ICP) algorithm [5]. The energy function consists in summing for each point of the floating surface its squared distance to the closest point on the target surface representation. However, this criterion, at first proposed for rigid registration, is not suitable for affine or non-rigid transformation [6]. Indeed, if the whole floating surface is matched to a single point on the target surface, the energy function is zero, corresponding to an obviously aberrant solution. Consequently, when registering a surface onto another one with affine or non-rigid transformation, a shrinking of the warped surface may be observed. To circumvent this phenomenon, some methods match points which have similar local shape features [6, 7]. We propose another way to handle this problem based on the following observation: the ICP algorithm converges towards an aberrant solution since it considers only the points of interest and not the other points, which have also to be matched. Consequently, to compute the distance between the floating segmentation $B^{l-1}(\boldsymbol{h}^l)$ and B_m, we consider a distance composed of two terms. The first one, similar to the one used in the ICP algorithm, is the sum for each bone voxel in B_m of its squared distance to the closest bone voxel in the floating image. The second term is computed as the sum for each non-bone voxel of B_m of its squared distance to the closest non-bone voxel in the floating image. To alleviate the computational burden, two chamfer distance maps are used to compute the cost function. Moreover, to enforce the smoothness of the deformation field, we consider an additional regularization term $E_{reg}(\boldsymbol{h}^l)$ in the cost function, corresponding to the elastic membrane energy of the deformation field. The Jacobian $J(\boldsymbol{h}^l)$ of the mapping $\boldsymbol{h}^l$ is constrained to be positive in order to ensure the estimated transformation to be a one-to-one mapping [4]. The optimization problem writes finally as follows:

$$\boldsymbol{h}^l = \underset{J(\boldsymbol{h}^l) > 0}{\arg\min} \sum_{\boldsymbol{s} \in \Omega; B_m(\boldsymbol{s}) = 1} CP_{B^{l-1}}(\boldsymbol{h}^l(\boldsymbol{s})) + \sum_{\boldsymbol{s} \in \Omega; B_m(\boldsymbol{s}) = 0} CB_{B^{l-1}}(\boldsymbol{h}^l(\boldsymbol{s})) + \lambda C E_{reg}(\boldsymbol{h}^l), \quad (1)$$

where $CP_{B^{l-1}}(\boldsymbol{s})$ and $CB_{B^{l-1}}(\boldsymbol{s})$ stand respectively for the squared distance between the voxel $\boldsymbol{s}$ and the closest bone point in B^{l-1}, and the squared distance between $\boldsymbol{s}$ and the closest non-bone voxel of B^{l-1}. λ is the weighting factor of the regularization term (it has been set to 1 in our experiments) and C a scaling factor computed at the beginning of each scale so that the data energy term and E_{reg} are comparable. Estimation of the parameters is made using the blockwise constrained gradient descent algorithm described in [4].

2.2. Knowledge extraction

A way to segment I is to transport the segmentation B_m onto I. However, the quality of such a segmentation depends on the registration accuracy. Suppose that the maximal registration error at scale l along the x-, y- or z-axis, $\epsilon_{max}(\boldsymbol{h}^l)$, is known (estimation of $\epsilon_{max}(\boldsymbol{h}^l)$ is discussed at the end of this section). The first extracted knowledge is an area, denoted Z^l_{boi}, which corresponds to all voxels of I which may possibly correspond to bones of interest with regard to the registration accuracy at scale l. Z^l_{boi} is merely obtained by transporting the segmentation map B_m on I and by dilating this transported segmentation map with a cubic structuring element of size $\epsilon_{max}(\boldsymbol{h}^l)$. Transporting the segmentation map B_m on I requires the computation of the reverse transformation $\boldsymbol{h}^{l-1}$. The inversion of the deformation field is done using a numerical scheme based on interval analysis techniques (not published yet).

The second extracted knowledge is the bone ratio $p^{l,r}_{i,j,k}$ in $Z^l_{boi} \cap \Omega^r_{i,j,k}$ (see 2.1 for the definition of $\Omega^r_{i,j,k}$, and Fig. 1 for a graphical representation of Z^l_{boi} and $p^{l,r}_{i,j,k}$) computed as follows:

$$p^{l,r}_{i,j,k} = \frac{|Z^l_{ref} \cap \Omega^r_{i,j,k}|}{|Z^l_{boi} \cap \Omega^r_{i,j,k}|}, \quad (2)$$

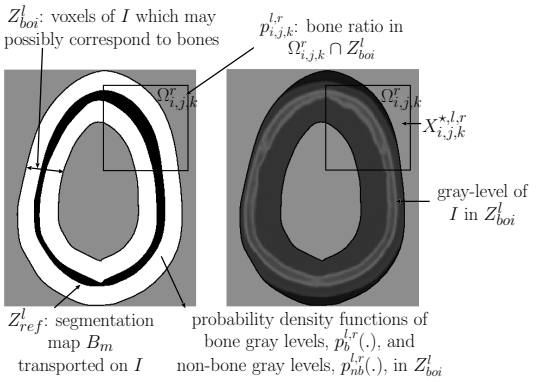

Fig. 1. Extracted features (left) used for the estimation of $X^{\star,l,r}_{i,j,k}$ (right).

Z^l_{ref} corresponding to the segmentation map B_m transported on I (see Fig. 1). The size of $\Omega^r_{i,j,k}$, characterized by scale r, must be large with respect to the accuracy of the registration, in order to guarantee a reliable estimation of $p^{l,r}_{i,j,k}$. Consequently, only r's lower than or equal to l-1 are considered.

Concerning the estimation of the maximal registration error at scale l, $\epsilon_{max}(\boldsymbol{h}^l)$ is set empirically to the quarter of size of a spline support at scale l. This setting has been validated *a posteriori* by the quality of the results. A more confident estimation of $\epsilon_{max}(\boldsymbol{h}^l)$ is under investigation by learning registration errors on some characteristic points defined by an expert on a database of images.

2.3. Bayesian estimation of the threshold map

Features extracted during the previous step are used for estimating the threshold map I_t, which is required by the segmentation procedure. During the segmentation step, a voxel $\boldsymbol{s}$, whose gray-level $I(\boldsymbol{s})$ is greater than or equal to $I_t(\boldsymbol{s})$, will be considered as bone, whereas a voxel $\boldsymbol{s}$, whose gray-level $I(\boldsymbol{s})$ is smaller than $I_t(\boldsymbol{s})$, will be considered as non-bone, unless topological constraints imply the contrary.

The threshold map is estimated in a coarse-to-fine way (from scale r=0 to r=l-1) thanks to the set of features $\{p^{l,r}_{i,j,k}\}$. For a given r, the basic idea consists in estimating for each volume $\Omega^r_{i,j,k} \cap Z^l_{boi}$ a threshold $X^{\star,l,r}_{i,j,k}$ (see Fig. 1). The threshold $X^{\star,l,r}_{i,j,k}$ is computed by minimizing a Bayes risk, which is computed from the prior probability $p^{l,r}_{i,j,k}$ (expected bone ratio in $Z^l_{boi} \cap \Omega^r_{i,j,k}$), and two likelihood terms (the probability density functions (*pdf*s) of the gray-levels in Z^l_{boi} for bone voxels, $p^{l,r}_b(.)$, and for non-bone voxels, $p^{l,r}_{nb}(.)$). Gaussian *pdf*s are considered. They are estimated for a given r using the segmentation $B^{l,r-1}$ if $r >$0, and B^{l-1} otherwise. More precisely, given a threshold X_0, the Bayes risk $C^{l,r}_{i,j,k}(X_0)$ is defined as the sum of two terms: the number of expected voxels in $Z^l_{boi} \cap \Omega^r_{i,j,k}$ corresponding to bone voxels although their gray-level is lower than X_0 and the expected number of voxels in the same area corresponding to non-bone voxels although their gray-level is greater than or equal to X_0.

A continuous model of the threshold map is used in order to prevent block effects in the segmentation. The map is modeled using a hierarchical representation based on a first degree polynomial spline scaling function similar to the one used for modelling the deformation field. Moreover, an additional parameter, denoted $t^{l,0}$, is used to model a constant threshold. The threshold map $I^{l,r}_t$ is derived from the set of thresholds $\{X^{\star,l,r}_{i,j,k}\}$. Moreover, for a voxel $\boldsymbol{s}$, which does not belong to Z^l_{boi}, or which belongs to a volume $Z^l_{boi} \cap \Omega^r_{i,j,k}$ for which the threshold $X^{\star,l,r}_{i,j,k}$ is greater than all gray-levels in $Z^l_{boi} \cap \Omega^r_{i,j,k}$, $I^{l,r}_t(\boldsymbol{s})$ is set to $+\infty$, since no bone is expected.

2.4. Segmentation

The head bones present non-trivial topological properties. They are composed of two connected components (mandible bone *vs.* other bones), and present holes (foramina) and cavities (sinuses) - refer to [8] for classical topological definitions used in this section. Most of these topological properties correspond to fine details in CT images (the diameter of holes is generally small, such as the distance between the connected components). Thus, from a simplified point of view, the head - when considered at a coarse resolution - can be modeled as a simply connected object (*i.e.*, an object composed of one connected component), with no holes and no cavities. At a finer resolution, it presents more specific topological details. Based on these assumptions, it can be justified to propose a topology-controlled segmentation process, composed of a procedure of homotopic deformation from an initial object topologically equivalent to the head from a coarse point of view, and followed by a topology modifying procedure devoted to recover the finer topological details of the head.

Actually, the segmentation process can be split into three distinct steps. The first one is the "initial" head segmentation (*i.e.*, when $r = 0$, at each scale l). A simply connected object is obtained by performing a homotopic reduction on the whole image support. The second step consists in refining the segmentation $B^{l,r}$ ($1 \leq r \leq l-1$) from the information carried out by the segmentation $B^{l,r-1}$, the image I and the threshold map $I^{l,r}_t$. This step is ensured to preserve the topology by removing or adding simple points. The third step consists in recovering the fine topology details of the object obtained from the last segmentation $B^{l,l-1}$. It performs the following operations: recovery of cavities, separation of connected components, and finally, recovery of holes. Note that, in the preliminary version of the proposed method, this step is not yet driven by "a priori" knowledge. However, information on the localization and size of the cavities (corresponding to sinuses), on the number of connected components and on the number of holes can be extracted from the segmented example to drive and constrain the segmentation process in further works.

3. RESULTS

The proposed framework was applied to a dataset of 32 CT images. The final scale, denoted l_f, has been set to 6 in our experiments. As the size of the image is 256^3, the size of final spline supports is 8^3. The proposed approach was compared with Otsu's method. An illustrative example of segmentation results is shown in Fig. 2. The first row represents one CT image of the dataset in coronal, sagittal and axial views. The second and third rows represent the segmentation results obtained with Otsu's method and the proposed approach, respectively. The Otsu's method is very sensitive to noise (spongy bones are sometimes not well segmented) and to artifacts which corrupt CT images (metal objects such as dental fillings lead to severe streaking artifacts). Results obtained by jointly performing registration and segmentation are much satisfactory. These good results are largely explained by the features provided by the reference image and, to a lesser extent, by the topologically-controlled segmentation process (a more intensive use of topological knowledge is under investigation). Moreover, contrary to the Otsu's method, the proposed method allows to segment only the structures of interest (the spinal column does not appear in the third row) thanks to the knowledge provided by B_m. Finally, the proposed approach has shown to be robust to the variability of human head. As an example, subjects with no teeth have also been well-segmented. This illustrates the fact that, even if the extracted features of the example do not exactly correspond to the properties of the image to segment, the data term of the Bayesian rule may drive the segmentation to the good solution.

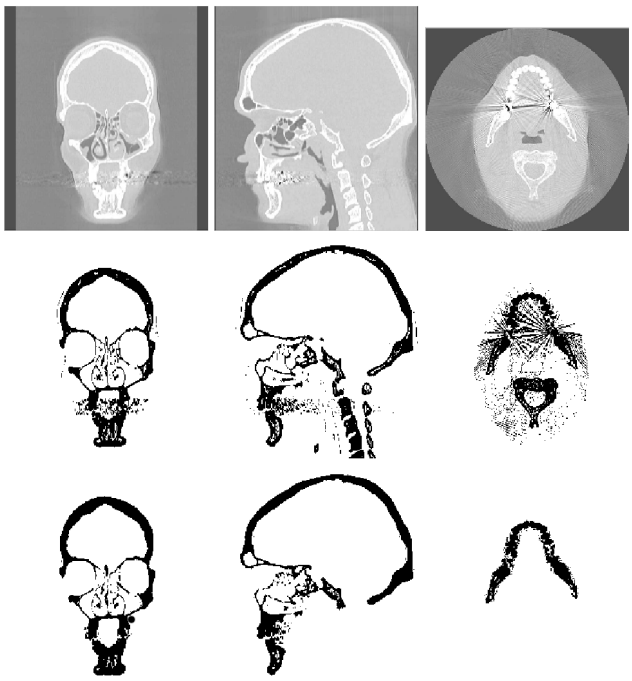

Fig. 2. Segmentation results. First row: 3-D CT image. Second row: result obtained with Otsu's method. Third row: result obtained with the proposed method.

4. CONCLUSION AND PERSPECTIVES

We proposed a scheme to segment complex structures based on an available segmented example. Results obtained in the context of head bone segmentation in 3-D CT images highlight the interest of the approach. Moreover, the proposed framework is versatile and may be extended. For example, further work will consist in using more intensively the topological knowledge of the segmented example to drive the segmentation/registration process.

5. REFERENCES

[1] S. Lončarić and D. Kovačević, "A method for segmentation of CT head images," in *ICIAP (2)*, 1997, pp. 388–395.

[2] D. Vandermeulen, P. Claes, R. Suetens, S. De Greef, and G. Willems, "Volumetric deformable face models for cranio-facial reconstruction," in *ISPA*, 2005, pp. 353–358.

[3] Y. Schnitman, Y. Caspi, D. Cohen-Or, and D. Lischinski, "Inducing semantic segmentation from an example," in *ACCV*, 2006, vol. 3852, pp. 373–384.

[4] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach, "3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization," *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 553–556, 2005.

[5] P.J. Besl and N.D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[6] J. Feldmar and N. Ayache, "Rigid, affine and locally affine registration of free-form surfaces," *International Journal of Computer Vision*, vol. 18, no. 2, pp. 99–119, 1996.

[7] G. Subsol, J.-P. Thirion, and N. Ayache, "A scheme for automatically building 3D morphometric anatomical atlases: application to a skull atlas," *Medical Image Analysis*, vol. 2, no. 1, pp. 37–60, 1998.

[8] G. Bertrand and G. Malandain, "A new characterization of three-dimensional simple points," *Pattern Recognition Letters*, vol. 15, no. 2, pp. 169–175, 1994.