

ACCURATE AND ROBUST SHAPE DESCRIPTORS FOR THE IDENTIFICATION OF RIB CAGE STRUCTURES IN CT-IMAGES WITH RANDOM FORESTS

Mariem Gargouri^{*†} Julien Tierny^{*} Erwan Jolivet[‡] Philippe Petit[†] Elsa D. Angelini^{*}

^{*} Institut Mines Telecom - CNRS LTCI - Telecom ParisTech, Paris, FRANCE

[†] LAB-PSA Peugeot Citroen-Renault, FRANCE

[‡] CEESAR - European centre of studies and risk analysis, FRANCE

ABSTRACT

This paper presents a new automatic technique for the segmentation of the rib cage on CT images. Motivated by a usage scenario in the context of large, heterogeneous databases of CT-images, we introduce two shape descriptors to be used in conjunction with a Random Forests (RF) classifier. These descriptors were specifically designed to address the challenges of rib identification under various acquisition conditions affecting subject's orientation and image quality. Extensive experiments demonstrate the superiority of our proposed shape descriptors in nominal configurations. Robustness with respect to subject's orientation variation and additive noise is also demonstrated, with an improvement of classification performance of up to 25%, comparing to intensity-based descriptors, without neither pre-registration nor pre-smoothing.

Index Terms— CT images, bone segmentation, thorax, random forests, shape descriptors.

1. INTRODUCTION

There is an increasing interest in image-based anatomical models, exploiting diverse 3D volumes of medical images from several modalities. In this work, we are interested in human skeleton modeling based on CT images. In the context of our work, numerical models of human skeletons are used by car manufacturers who dedicate important research efforts in developing numerical simulations of crash tests to improve the safety of their new vehicles. These numerical simulations enable to study the mechanical reactions of specific bones during diverse types of car crashes. Until now, numerical models of human skeletons have been built based on CT images of individual subjects, which had to be manually segmented.

However, a large database of numerical models derived from a diverse population of subjects (categorized by gender, age and morphology) is required to build up a sufficient and reliable knowledge of fracture risks at the level of a population. In this context, it is planned to use in the near future a large, heterogeneous, database of thousands of CT-images of diverse subjects. Constructing such a database requires to use

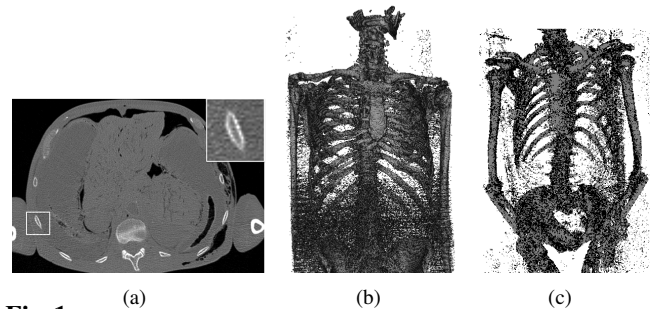


Fig. 1. Illustrations of the challenges in rib identification on a heterogeneous database of CT images: (a) Axial slice of the torso showing very thin ribs' cortical cross-sections (see the inset zoom) with a typical thickness of 1.5 mm; (b) High noise level in some acquisitions; (c) Variable subject's posture.

automated segmentation tools to extract detailed bone structures. In this context, we focus in this paper on the rib cage segmentation, required to study car crash-induced thoracic injuries, which are life threatening and remain among occupant safety priorities for car manufacturers.

Regarding bone segmentation on CT images, a variety of approaches have been derived to extract large bones such as the tibia or the femur [1, 2, 3] and small bones such as vertebrae [4, 5]. However, ribs turn out to be particularly challenging for existing automatic segmentation techniques. Indeed, the thickness of the cortical bone is of the same order of magnitude as the voxel resolution used for patient imaging. Therefore, the intensity of cortical bone voxels can drastically decrease and become difficult to distinguish from the surrounding soft tissues (see Fig. 1). Based on this observation, we describe in this paper an approach for the identification of rib cage structures that leverages the expertise of users through a learning procedure. In particular, Random-Forests (RF) [6] has shown to be a very successful learning-based approach [7], especially with rather simple shape descriptors, for the segmentation of several types of bone structures on CT images [8, 9]. In addition, in the context of a large heterogeneous database of CT images, the classification procedure should be robust to variability in acquisition settings (and therefore noise level) and to the possible changes in subject postures (see Fig. 1). In this paper, we propose to use two new shape descriptors in a RF classifier that enable a reliable

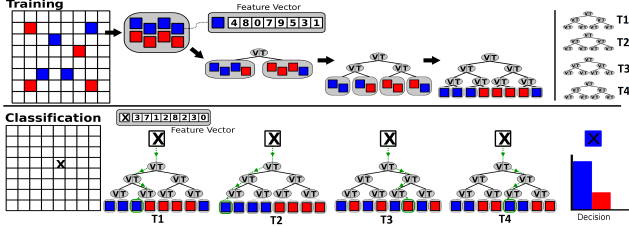


Fig. 2. Illustration of the Random Forests classifier structure. During the training (top), a random subset of labeled samples is used to build a binary decision tree. For each node, random variables of the sample’s feature vector (top) are evaluated and the pair Variable-Threshold which best splits the subset is associated to the node. The process is repeated for a large number of trees (top right). During the classification (bottom), the sample to classify is sent to each decision tree and the class of the sample is determined by the combination of the predictions of the trees of the forest.

and robust identification of the rib cage structures. Extensive experiments demonstrate the superior accuracy in classification performance of our proposed descriptors (in comparison to intensity-based descriptors used in [8]) as well as significant classification performance improvements (up to 25%) on data-sets subject to posture changes and random noise addition. The remainder of the paper is organized as follows: Sec. 2 describes our overall segmentation framework, Sec. 3 our proposed shape descriptors, and Sec. 4 presents our experimental results.

2. RIB CAGE LABELING WITH RANDOM FORESTS

Random Forests (RF) [6] is a classifier that combines multiple randomized decision trees. Based on a simple algorithm, it yields very fast and accurate supervised classifications especially for large-dimension feature spaces [7]. RF classifiers can also be generalized easily to multi-class classification tasks.

In our work, each input CT volume, representing a human torso (from the pelvis to the neck), was encoded on a voxel grid \mathcal{G} of size $\sim 1500 \times 512 \times 512$ (voxel resolution: $0.75 \times 0.97 \times 0.97$ mm) where each voxel $v = (x, y, z)$ is associated with an intensity value $i(v)$. To speed up both the training and the classification procedures, the set of *bone voxels* \mathcal{B} to label is pre-identified as the set of voxels whose intensity is higher than a given threshold $i_{\mathcal{B}}$ (typically, $i_{\mathcal{B}} = 400HU$):

$$\mathcal{B} = \{v \in \mathcal{G} / i(v) \geq i_{\mathcal{B}}\} \quad (1)$$

Fig. 1 and 4 illustrate the ensemble of bone structures to label on three different CT scans.

2.1. RF training procedure

The input to the RF training procedure is a set of bone voxels manually pre-labelled by an expert. Each bone voxel $v \in \mathcal{B}$ is associated with a feature vector $f(v) \in \mathbb{R}^n$ (in our case, a set of n real values) and a class $c(v) \in \{0, 1\}$ (respectively, *non-rib voxels* and *rib voxels*). The feature vector can be composed of various measurements and we further discuss some examples in Sec. 3. As illustrated in Fig. 2, the training

procedure builds a set of binary decision trees, called a forest $\mathcal{F} = \{\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_{N_{\mathcal{T}}-1}\}$ (where $N_{\mathcal{T}}$ is the number of trees in the forest). Each tree $\mathcal{T} \in \mathcal{F}$ is built independently in the following manner, given a random subset \mathcal{R} of \mathcal{B} (of size $N_{\mathcal{R}}$). Each node of \mathcal{T} is first associated with a set of $N_{\mathcal{V}}$ pairs of Variable-Threshold. The variables are picked randomly among the components of the feature space (between 0 and n) and their corresponding thresholds are also picked randomly within the range of values of each feature. Each pair of Variable-Threshold yields a binary classification of \mathcal{R} into \mathcal{R}_0 and \mathcal{R}_1 that is evaluated by its gain G on the Gini index:

$$G = \mathcal{G}(\mathcal{R}) - \mathcal{G}(\mathcal{R}_0) \times \frac{|\mathcal{R}_0|}{|\mathcal{R}|} - \mathcal{G}(\mathcal{R}_1) \times \frac{|\mathcal{R}_1|}{|\mathcal{R}|} \quad (2)$$

and where the Gini index \mathcal{G} (a standard purity measure in classification) is given by:

$$\mathcal{G}(\mathcal{R}) = 2 \times \frac{|\{v \in \mathcal{R} / c(v) = 0\}|}{|\mathcal{R}|} \times \frac{|\{v \in \mathcal{R} / c(v) = 1\}|}{|\mathcal{R}|} \quad (3)$$

The pair of Variable-Threshold which maximizes G is then stored in the node and the sets \mathcal{R}_0 and \mathcal{R}_1 are provided as inputs to the creation of two children-nodes. The rest of \mathcal{T} is constructed recursively with this algorithm until the tree reaches a pre-set maximal depth \mathcal{D}_{max} or until $\mathcal{G}(\mathcal{R}) = 0$. Finally, each leaf of \mathcal{T} is labelled with the class represented in majority by its associated voxels (see Fig. 2).

2.2. RF classification procedure

As illustrated in Fig. 2, the classification of a new voxel v is straightforward and fast. Each decision tree \mathcal{T} is traversed by v from its root by routing v to the appropriate child-node if the node’s variable in $f(v)$ is lower or greater than the node’s threshold. The leaf node in which v stops its traversal will then yield a prediction on $c(v)$. Finally, the predictions of all the trees are combined to assign the final value for $c(v)$, via majority voting.

3. ROBUST SHAPE DESCRIPTORS FOR RF CLASSIFICATION

In this section, we present three shape descriptors that we propose to use in the RF classifier for the automated segmentation of the rib cage. In particular, we introduce two new shape descriptors which aim at providing more accurate and robust classifications than intensity-based features.

3.1. Intensity-based features

For our experiments, we evaluated an intensity-based descriptor, called baseline descriptor, which was introduced in previous work [8] and is computed as follows: For each voxel $v \in \mathcal{B}$ to train or classify, a set of 60 boxes of random size (with a length between 3 and 21 voxels along each dimension), and centered at random locations within the neighborhood of v (within 20 voxels) is generated. A feature vector $f_{baseline} \in \mathbb{R}^{60}$ is then computed with the average intensity values within each box (cf Fig. 3). Note that, for consistency,

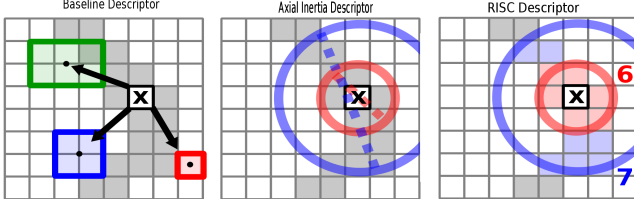


Fig. 3. Illustration of the intensity and shape descriptors around a pixel p marked with a cross. From left to right: baseline (intensity-based) descriptor with box neighborhoods, axial inertia and rotation invariant shape context descriptors, both with spherical concentric neighborhoods.

the randomization regarding the position and size of the boxes only occurs once, during the initialization of the RF training. The same box parameters are employed for the classification task.

This baseline descriptor, due to its simplicity, is well suited for the labeling of homogeneous structures. However, in the context of rib identification, as illustrated in Fig. 1, the average intensity of voxels exhibits some heterogeneity along the rib, and the intensity of the surrounding soft tissues is quite homogeneous, which injects some ambiguity in the classifier. In addition, although, in theory, average intensities should be robust to noise, this baseline descriptor is not rotation invariant, which prevents it from yielding accurate classification in case of posture change. We propose to overcome these difficulties with our new shape features introduced in the following sub-sections.

3.2. Axial-inertia descriptors

In order to circumvent the ambiguities related to homogeneity measures within and around the ribs, we focus our new features on the description of the local geometry of the bone voxels \mathcal{B} . In particular, we propose to evaluate the local curvature of the set of bone voxels in small neighborhoods around the voxels of interest. Given a voxel p , a neighborhood region \mathcal{S}_p is defined within a sphere of radius r and centered in p . We then compute on the set of neighborhood bone voxels $\mathcal{B}_p = \{v \in \mathcal{B} \cap \mathcal{S}_p\}$, its barycenter b_p and its inertia matrix I_p , defined as:

$$I_p = \begin{bmatrix} A & -F & -E \\ -F & B & -D \\ -E & -D & C \end{bmatrix} \quad (4)$$

with:

$$\begin{aligned} A &= \sum_{v \in \mathcal{B}_p} (y(v) - y(b))^2 + (z(v) - z(b))^2 \\ B &= \sum_{v \in \mathcal{B}_p} (x(v) - x(b))^2 + (z(v) - z(b))^2 \\ C &= \sum_{v \in \mathcal{B}_p} (x(v) - x(b))^2 + (y(v) - y(b))^2 \\ D &= \sum_{v \in \mathcal{B}_p} (y(v) - y(b)) \times (z(v) - z(b)) \\ E &= \sum_{v \in \mathcal{B}_p} (x(v) - x(b)) \times (z(v) - z(b)) \\ F &= \sum_{v \in \mathcal{B}_p} (y(v) - y(b)) \times (x(v) - x(b)) \end{aligned}$$

where $(x(v), y(v), z(v))$ are the XYZ coordinates of a voxel v .

The eigen vectors of I_p correspond to the principal directions of the bone voxels \mathcal{S}_p , and the corresponding eigen values quantify the moments of inertia of the bone voxels \mathcal{B}_p along these principal directions (i.e. their deviation from the direction of the corresponding eigen vector). Let λ_p^{max} and λ_p^{min} be the maximal and minimal eigen values of I_p . The proposed feature vector, which characterizes the axial inertia, is denoted $f_{AI}(p) \in \mathbb{R}^{42}$, and composed of the pairs $(\lambda_p^{max}, \frac{\lambda_p^{max}}{\lambda_p^{min}})$ for 21 spheres with a radius r progressively increasing from 10 to 30 voxels. The greater λ_p^{max} , the more aligned the bone voxels are. As illustrated in Fig. 3, this descriptor measures the deviation of the bone voxels from their local principal direction, within neighborhood of increasing scales. Note that, by design, this descriptor is rotation invariant.

3.3. Rotation invariant shape context (RISC) descriptors

The second proposed descriptor is inspired from the shape-context descriptor [10], which is popular in computer vision. To obtain a rotation-invariant version of this descriptor (and avoid registration ambiguities), As illustrated in Fig. 3, we compute, for a given voxel p , a set of 56 bands between spheres centered in p and of radius values randomly sampled between 2 and 35 voxels. Then for each band, the number of bone voxels is calculated, yielding a feature vector $f_{RISC}(p) \in \mathbb{R}^{56}$. This descriptor is rotation-invariant and depicts the local volume occupied by bone voxels in spherical bands of increasing radius around the voxel of interest.

4. EXPERIMENTS AND RESULTS

We experimented RF classification exploiting each of the three descriptors, for the task of rib labeling on thoracic CT scans. For each descriptor, we trained the RF classifier on a set of 437,000 randomly chosen bone samples (45% rib bones, 55% non-rib bones) (taken from one thoracic CT scan of an average size old male (above 60 years old)) and tested it with 3,938 randomly chosen bone samples (48% rib bones, 52% non-rib bones) from a different subject with similar size and age. We set the parameters of the RF to the following values: $N_V = 7$, $N_T = 500$, $D_{max} = 10$ and N_R is a random value (between 10 and $|\mathcal{B}|$). The training procedure, for each descriptor, on a desktop computer with an i7 CPU (2.93GHz) and 8Gb of RAM, with an implementation combining MatLab and C++ codes (using OpenCV), took about 3 hours while the entire classification task took on average 0.5 seconds. A visual illustration of the labeling quality is provided in Fig. 4 for one thorax.

Table 1 reports the classification performance with each descriptor. The best results were obtained for our RISC descriptor, improving by 4% the baseline performance.

Table 1. Performance of RF classification for rib bone extraction using three different descriptors: baseline, axial inertia (AI) and the rotation-invariant shape-context (RISC).

	Baseline	AI	RISC
Correct classification	92.12%	90.32%	96.19%
False positives	0.45%	7.54%	0.83%

Table 2. Performance of RF classification for rib bone extraction on the same CT scan as in Table 1, but rotated by 20°.

	Baseline	AI	RISC
Correct classification	81.28%	87.22%	97.33%
False positives	3.02%	11.7%	1.01%

Table 3. Performance of RF classification for rib bone extraction on the same CT scan as in Table 1, corrupted with additive random noise.

	Baseline	AI	RISC
<i>1% noise</i>			
Correct classification	91.92%	87.51%	96.92%
False positives	2.18%	10.99%	0.81%
<i>2% noise</i>			
Correct classification	72.09%	76.38%	97.48%
False positives	24.42%	22.65%	0.68%

As discussed in the introduction, homogenous acquisition settings cannot be assumed when processing large databases of CT-images. In particular, subjects can have different postures and CT scanners can generate images with varying noise characteristics that can be hard to predict beforehand and not represented during training. Addressing such issues with pre-smoothing of the images might be highly unreliable, based on our experience. Classification performance was further tested on the same CT scan thorax and points to label as in Table 1 but after some modification: (1) rotation of the volume by 20° around the direction orthogonal to the coronal plane to simulate posture change, with results in Table 2. We observed that both AI and RISC descriptors provided significant classification improvements (up to 16%) over the baseline descriptor performance; (2) addition of a white Gaussian random noise (which amplitude was set as a fraction of the whole intensity range, at 1% and 2%), with results in Table 3. In that case, the RISC descriptor outperformed the baseline and AI descriptors by up to 25%. Note that various combinations of these three descriptors were also experimented but with no significant improvement over the RISC descriptor alone. Labeling of a full thoracic CT scan using such combination is illustrated in Fig. 4

5. DISCUSSION AND CONCLUSION

In this paper, we presented a new approach for the automatic segmentation of rib bones in CT-images which leverages the expertise of users through a learning procedure. We introduced two shape descriptors to be used in conjunction with a Random Forests classifier. Experiments demonstrated the superiority of rotation-invariant shape context descriptors over

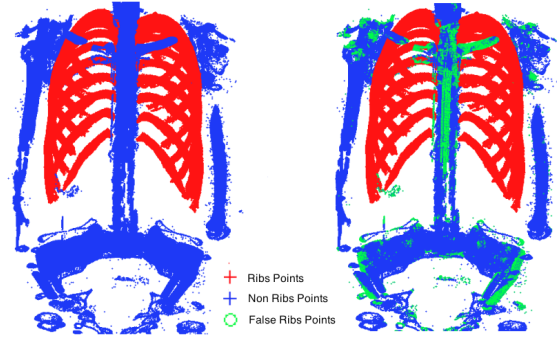


Fig. 4. Rib bone labeling on a full thoracic CT scan: (left) manual expert segmentation, (right) RF classification combining baseline and RISC descriptors. (Blue: true non-rib points, Red: true rib points, Green: false rib points).

nominal intensity-based descriptors, especially under variations of subject’s posture and noise level (with improvements over existing descriptors of 4%, 16% and 25% respectively). Such descriptors seem therefore well suited for the automated classification of rib bones on heterogeneous databases of CT images, without requiring pre-registration nor smoothing.

In future work, we will extend our descriptors for multi-class RF classification to individually segment each rib bone, and to distinguish between cortical and spongy bone components.

6. REFERENCES

- [1] X.M. Pardo, M.J. Carreira, A. Mosquera, and D. Cabello, “A snake for CT image segmentation integrating region and edge information,” *Image and Vision Computing*, vol. 19, pp. 461–475, 2001.
- [2] Y. Kang, K. Engelke, and W.A. Kalender, “A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data,” *IEEE Transactions on Medical Imaging*, vol. 22, pp. 586–598, 2003.
- [3] M. Krcah, G. Szekely, and R. Blanc, “Fully automatic and fast segmentation of the femur bone from 3D-CT images with no shape prior,” in *IEEE ISBI*, 2011, pp. 2087–2090.
- [4] T. Klinder, M.E. Ostermann, A. Franz, R. Kneser, and C. Lorenz, “Automated model-based vertebra detection, identification and segmentation in CT images,” *Medical Image Analysis*, vol. 13, pp. 471–482, 2009.
- [5] J. Ma, L. Lu, Y. Zhan, X. Zhou, M. Salganicoff, and A. Krishnan, “Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model,” in *MICCAI*, 2010, pp. 19–27.
- [6] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001, 10.1023/A:1010933404324.
- [7] Rich. Caruana, Nikos. Karampatziakis, and Ainur. Yessenalina, “An empirical evaluation of supervised learning in high dimensions,” *Proceedings of the 25th International Conference on Machine Learning (2008)*, pp. 96–103, 2008.
- [8] K. Rajpoot, A. Noble, V. Grau, C. Szmigielski, and H. Becher, “Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography,” in *Functional Imaging and Modeling of the Heart*, 2009, pp. 447–456.
- [9] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, “Regression forests for efficient anatomy detection and localization in CT studies,” in *MICCAI Medical Computer Vision Workshop*, 2010, pp. 106–117.
- [10] G. Mori, S. Belongie, and J. Malik, “Shape contexts enable efficient retrieval of similar shapes,” in *IEEE CVPR*, 2001, pp. 723–730.