

Cortical Bone Classification

by

Local Context Analysis

S. Battiato¹, G. M. Farinella¹, G. Impoco¹, O. Garretto², and C. Privitera²

¹ Computer Science Department, University of Catania, Catania, ITALY
{battiato, gfarinella, impoco}@dmi.unict.it,
<http://www.dmi.unict.it/~iplab/>

² Radiology Department, Vittorio Emanuele Hospital, Catania, ITALY

Abstract. Digital 3D models of patients' organs or tissues are often needed for surgical planning and outcome evaluation, or to select prostheses adapted to patients' anatomy. Tissue classification is one of the hardest problems in automatic model generation from raw data. The existing solutions do not give reliable estimates of the accuracy of the resulting model. We propose a simple generative model using Gaussian Mixture Models (GMMs) to describe the likelihood functions involved in the computation of posterior probabilities. Multiscale feature descriptors are used to exploit the surrounding context of each element to be classified. Supervised learning is carried out using datasets manually annotated by expert radiologists. 3D models are generated from the binary volumetric models, obtained by labelling cortical bone pixels according to maximal likelihoods.

1 Introduction and Motivation

The rapid technological advances in digital medical imaging devices have brought the attention of various medical communities to computer-assisted diagnosis.

One of the trends of modern surgery is computer-aided pre-operative planning and post-operative outcome evaluation. Digital models of the interested organs or tissues can be of great value both to predict and to evaluate the outcome of an operation. In particular, planning and evaluation often involve the creation of a model from patient-specific data, acquired by means of CT, MRI, and so on. 3D models of human body parts can be also useful for selecting perfectly-fitting prostheses e.g., for knee replacement [1] (using FEM stress analysis) and plastic surgery [2] (by means of ad-hoc geometric measurements).

Unfortunately, due to the limitations of current imaging devices the acquired data is often noisy. Moreover, in most technologies the ranges of data values of different tissues often overlap. This is especially true for soft tissues and trabecular bone in aged patients, where osteoporosis degenerates the bone density and thus the intensity of the bone is decreased (see Figure 1). Thus, data values cannot be uniquely associated with specific tissues i.e., the data cannot be

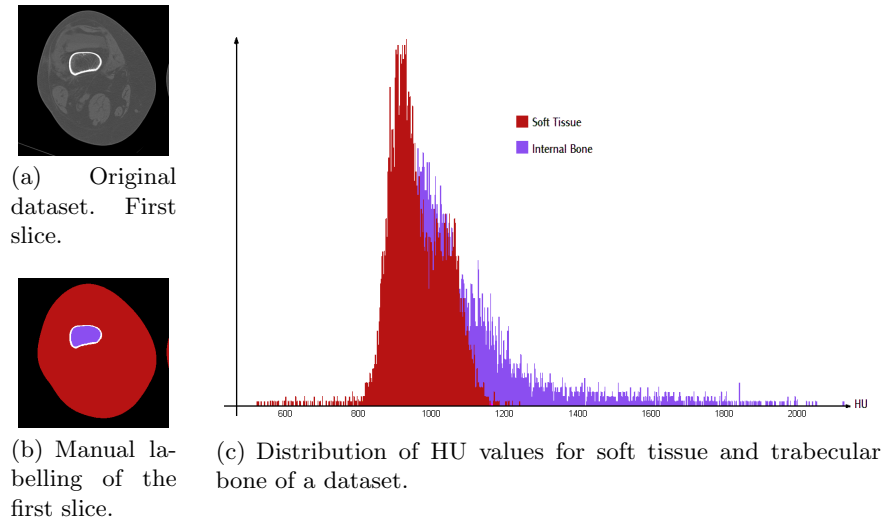


Fig. 1. A CT scan of a 80-years-old patient, affected by osteoporosis and arthrosis. A manual labelling is also shown together with the relative distribution of *Hounsfield Units* (HU) values of two tissue classes: *soft tissue* and *trabecular bone* (not used in our experiments).

partitioned using *Hounsfield Units* (HU) values alone. This rules out global as well as local thresholding techniques. Moreover, defining a similarity function between neighbouring pixels is hard, since the same tissues often have uneven values in different positions. Hence, region-growing, or edge detection algorithms are unable to effectively cope with this data.

Although a large variety of segmentation methods have been developed for medical image processing, ad-hoc solution are often preferred especially to properly detect complex structures, such as vessel, brain, or skeletal structures. In [3] a multi-step approach for 3D bone segmentation of CT data is presented. It combines a simple thresholding strategy with a 3D region growing followed by a refining step taking into account anatomically oriented boundary adjustment. The involved factors are in general of high complexity and require extreme accuracy. Moreover, the discriminative thresholds for tissues are chosen by experience, without any statistical justification. A computerised system, called *MedEdit* is presented in [4] to help surgeons in operation planning by making use of a Finite Element Analysis (FEA) program to measure and compare the effects of the modifications involved in skeletal injury operations. A seeded region growing techniques enriched with a suitable fuzzy heuristics is used for segmentation purposes. In our case, the CT data are acquired from old patients, which means that the bone density is lower than average. Since the intensity values in CT data are directly related to the tissue density, the low bone density is reflected in low intensity values for the affected parts. This makes the classification task more challenging. In addition to this, it is difficult to automatically separate

adjoining bones and regard them as different objects due to the limited spatial resolution of CT datasets. This is a requirement in the segmentation process, though, where the goal is to generate patient-specific models. Model-based approaches are needed as suggested in [5] and [6] even if there is no availability of reliable anatomical models.

Almost all previous works produce in some cases effective 3D reconstructions but without providing a reliable measure of uncertainty of the output. This is a serious drawback in applications that require an error measure, such as mechanical simulations [1]. Finally, and most important, segmentation algorithm group pixels into regions but do not label regions as being part of a semantic region or another. Hence, they lack a real classification mechanism.

In [7] a statistically-founded approach is used to infer the percentage of different tissues within each voxel due to the *partial volume* effect. Each voxel is modelled as a mixture of contributions from different tissues. They search for the tissue mixture that maximises the posterior probability. However, rather than labelling real data to compute HU distributions, they assume that the HU values of each tissue class are normally distributed. Our collected data show that this assumption does not hold for the knee region (Figure 1(c)). A learning procedure is clearly needed to build a reliable classification system.

Another statistical approach to bone segmentation from X-ray data is presented in [5]. However, as stated by the authors themselves, it is a largely immature work.

In this paper, we address the problem of classifying CT data in order to extract cortical bone structures. A simple generative model is used to classify the pixels belonging to a CT dataset into two classes: cortical bone, other. The likelihood functions involved in the computation of posterior probabilities are modelled by Gaussian Mixture Models (GMM), and are learned in a supervised way from manually-labelled data using the Expectation-Maximisation (EM) algorithm. Our approach shares some similarities with Lu et al.'s [7]. Anyway, while they model voxels as mixtures of contributions from different tissues, we use a single label for each pixel. This may look like a simplification, since our model cannot capture the intrinsic partial volume (PV) effects of the data. However, this choice is justified by the fact that there is no straightforward manual labelling procedure for supervised learning models. Basically, one can ask radiologists to give a single classification for each pixel (crisp classification), not to guess the percentage of a pixel occupied by a certain tissue (fuzzy classification).

The learned model is used to assign a probability to each pixel to belong to a certain class. A *Maximum A-posteriori Probability* (MAP) rule is used to get a crisp classification. The user can enforce a threshold on these probabilities to bound the classification uncertainty. A binary volumetric model is generated by labelling cortical bone voxels with low uncertainty. This model, in turn, can be used to build a mesh model of cortical bones using the well-known Marching Cubes algorithm [8]. The learning model can be easily extended to include a wider range of tissue classes by gathering and labelling more data. Results are presented for knee joint data.

2 Classification Model

Since our model is intended to be used for surgical planning, one of our main objectives is bounding the classification uncertainty. This is a good property because it is intuitive and is easily related to background knowledge of medical experts. Basically, one would say that she would accept the output of segmentation if the algorithm is p -percent sure about the result. Hence, we would classify only the pixels whose probability of belonging to a certain class is above a user-defined threshold, and leave the others unclassified.

2.1 Model

Suppose we want to partition our data into N different classes. We denote with c_k the k -th class. Let us define a number of feature functions $F_1(\cdot), \dots, F_M(\cdot)$ on the pixel domain. We use a simple generative approach [9] to learn the model for the posterior probability $P(c_k|F_1(z), \dots, F_M(z))$ for each pixel z . We model the likelihoods $P(F_1(z), \dots, F_M(z)|c_k)$ using GMMs (one for each class). Manually-annotated training datasets are used to learn the likelihoods, using the EM algorithm. We assume that the priors $P(c_k)$ are uniformly distributed. One might argue that these probabilities are different for each class and that we can compute them from our training sets by counting the number of pixels in each class. Anyway, too many uncontrolled elements can affect the percentage of bone pixels over the total volume, such as biological factors (e.g., age, sex, osteoporosis), and machine setup parameters (e.g., percentage of patient bone contained into the imaged volume). Hence, equiprobability of the priors $P(c_k)$ is a reasonable choice. Assuming equiprobable priors, by the Bayes' theorem we get

$$P(c_k|F_1(z), \dots, F_M(z)) = \frac{P(F_1(z), \dots, F_M(z)|c_k)}{\sum_k P(F_1(z), \dots, F_M(z)|c_k)} \quad (1)$$

where the evidence can be expressed in terms of likelihood functions alone.

Once we get the posterior probabilities $\hat{p}_k(z) = P(c_k|F_1(z), \dots, F_M(z))$, we partition the CT data using a MAP classification rule. In particular, for each pixel z the most probable labelling is

$$C(z) = \arg \max_{c_k} [\hat{p}_k(z)] \quad (2)$$

with associated probability $\hat{p}_C(z) = \max[\hat{p}_k(z)]$. In our application [1], it is preferable to be conservative employing a strict rejection rule to retain only very probable classifications. Thus, we accept this classification if $\hat{p}_C(z) \geq \varepsilon$, where ε is a user-defined threshold which bounds the classification uncertainty. If we restrict to two classes, c_1 and c_2 , and to a single feature $F_1(z)$ for the sake of visualisation, our simple rejection option can be depicted using the diagram in Figure 2.

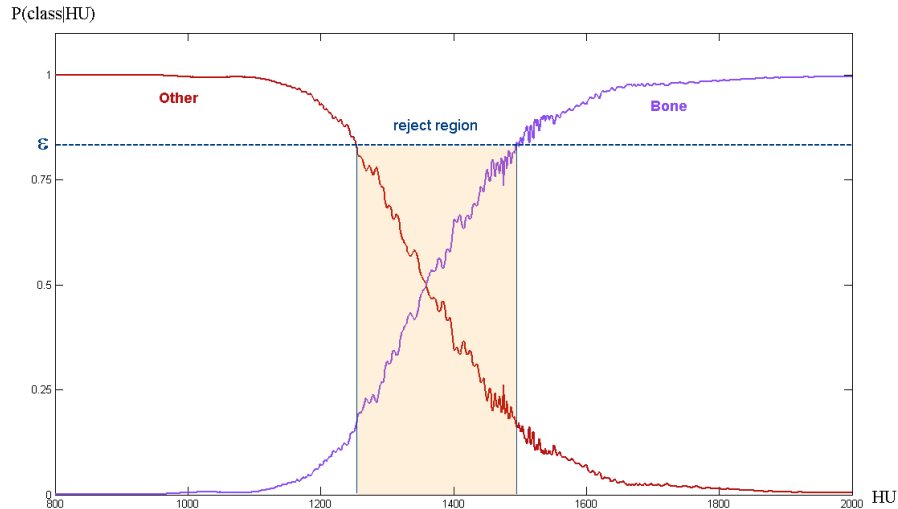


Fig. 2. Rejection rule. Probability distributions for two classes (*cortical bone* and *other*), conditioned by the HU value (our simplest feature). The threshold ε is used for rejection of low-probability classifications (see text). The shadowed region depicts the rejection interval.

2.2 Local Context Features

As pointed out in Section 1, the HU values of pixels do not suffice to get a robust classification. Pixel values can be affected by noise which rules out simple pixel-wise partitioning methods. According to [10], there are four sources of noise in CT images: physics-based artifacts, patient-based artifacts, scanner-based artifacts, and helical and multisection artifacts. Although the effect of some of these artefacts can be reduced or compensated, it can still strongly affect the acquired data. Moreover, the distributions HU values of different classes can largely overlap (Figure 1) especially for aged patients or for patients affected by osteoporosis or arthrosis. For this reason, we employ a more robust pixel analysis by looking at the neighbourhood, which can give useful information to estimate the probability of pixels to belong to a certain class, given their surrounding context. Hence, we employ a set of features at different scales to capture the variability of HU values in the surrounding context of a pixel.

For each scale s , we employ a $s \times s$ window centred around the interest pixel. We define $v_s = (F_{s,N}(\cdot), F_{s,E}(\cdot), F_{s,S}(\cdot), F_{s,W}(\cdot))$ as the mean of the HU values for four regions in the surrounding window: North (N), East (E), South (S), and West (W) (see Figure 3). In our implementation, we use the HU value of the interest pixel together with the feature vectors v_s .

We choose these features to selectively evaluate the surrounding context of a pixel in four directions. Being independent from orientation, circular features

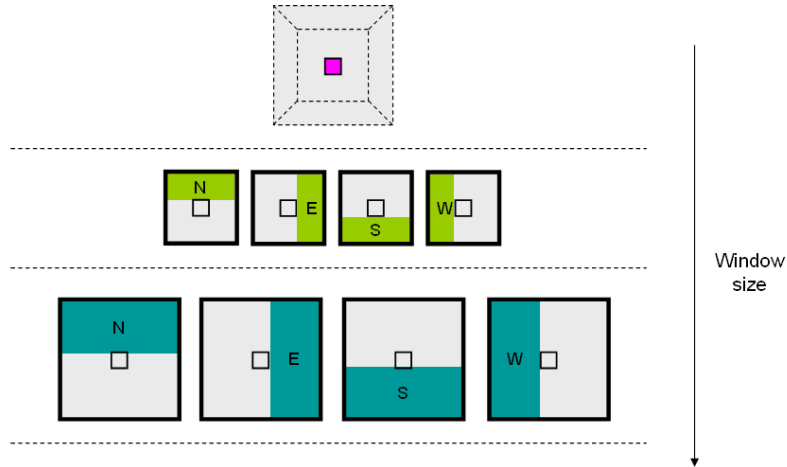


Fig. 3. Multiscale feature descriptors employed.

Parameter	Value	
Exposure	200	Sv
kVp	140	kiloVolt
Slice thickness	1.3	mm
Slice spacing	0.6	mm
Slice resolution	512 × 512 pixels	
Number of slices	70–140	

Table 1. Parameter settings used for the acquisition of the knee CT data employed in our experiments.

would bring less useful information. We also use simple mean instead of more complex distance weighings since distance is accounted for using multiple scales.

Due to the dimension of feature windows, especially in larger scales, a special treatment should be reserved to border pixels. Anyway, we do not bother about image borders since in our application interest pixels always lie in the central area of CT scans.

3 Experimental Procedure

The simple model presented in Section 2 has been tested on knee CT data. Nine knee datasets were imaged at the Radiology Department of the Vittorio Emanuele Hospital in Catania. The data were captured by means of a Multidetector CT Scanning (MDCT) device in spiral mode, using the acquisition parameters reported in Table 1. The age of the patients ranged from 70 to 80

years and all of them suffer from osteoporosis and arthrosis. This choice is motivated by the fact that this is one of the most difficult cases (due both to age and to severe osteoporosis and arthrosis) and the most widespread in the clinical cases of knee replacement surgery, which is our main interest [1]. The acquired data was then manually labelled by expert radiologists of the Vittorio Emanuele Hospital. 75% of the labelled data were used for learning, and the remaining were employed for testing.

3.1 Labelling Procedure

In order to simplify the labelling phase, we segmented the CT data using the SRM algorithm [11, 12]. Namely, labelling *superpixels* (i.e., pixel patches with homogeneous properties [13]) is much easier and faster than working with single pixels. We extended SRM to cope with 3D datasets. Basically, we insert in the couple pixel table couple of corresponding pixels of neighbouring slices, besides those in the same slice. Since inter-slice distance is usually much greater than pixel size, we weighted the priority of pixel couples by their physical distance. In principle, couple statistics should be collected for the whole dataset since regions might span the whole dataset along the scanning direction. However, this is quite computationally demanding, due to its memory requirements. Hence we segment each slice using only a small window of neighbouring slices. Although any window dimension can be used, we found that in practice using three slices is the best tradeoff between quality (the result is very similar to those obtained with larger windows) and computational burden. The rationale behind the use of SRM is that while the HU values of soft tissues and trabecular bone might largely overlap as in Figure 1(c), they are spatially separated by cortical bone. Hence, grouping pixels into superpixels using a similarity criterion based on HU value is safe and enriches the pixel value information with spatial context.

For labelling purposes, the user can select only superpixels as a whole, not single pixels. Hence, during learning the pixels in each superpixel share the same label. Classification is run on a pixel basis.

The labelling proceeds in two steps. First, the radiologists set three thresholds on HU values: background/soft tissue, soft tissue/trabecular bone, trabecular bone/cortical bone. These threshold are selected by looking at a single slice and then are propagated to the whole datasets. They are used as a basis for further manual refinement. Then, misclassified superpixels can be selected separately and singularly re-classified. The labelling software also provided a HU tolerance when selecting superpixels. However, the users found it of little use and its concept tricky to understand. Hence, we excluded it. In order to make the labelling environment more familiar to the radiologists, we show three panels containing the original slice, its segmented version (used to select superpixels), and the current annotation.

For the time being, only two classes are discriminated: *cortical bone* and *other*. However, the presented model can be easily extended to include a wider range of tissue classes by gathering and labelling more data. However, due to the

lack of fast labelling methods and label propagation procedures, collecting a sufficient amount of labelled data can be prohibitive. We are currently investigating labelling propagation procedures to alleviate the burden of manual annotation.

3.2 Model Parameters and Learning Procedure

In our experiments we use the features $v_s = (F_{s,N}(\cdot), F_{s,E}(\cdot), F_{s,S}(\cdot), F_{s,W}(\cdot))$ described in Section 2 with $s \in \{5, 15, 45, 135, 151\}$, and the HU value of the interest pixel.

We model the likelihoods involved in our classification framework using GMMs

$$P(F_1(z), \dots, F_M(z)|c_k) = \sum_{i=1}^M \pi_i G(F_1(z), \dots, F_M(z), \mu_i, \Sigma_i) \quad (3)$$

Manually-annotated training datasets are used to learn the likelihoods using the Expectation Maximisation algorithm [9] (20 iterations). In particular, we learn $P(F_1(z), \dots, F_M(z)|bone)$ using 100 Gaussians ($M = 100$) and $M = 150$ for $c_k \neq bone$. The k -means algorithm was used to determine the initial centres μ_i . The initial priors π_i were computed from the proportion of examples belonging to each cluster. The initial covariance matrices Σ_i were calculated as the sample covariance of the points associated with the corresponding centres.

3.3 Results

Figure 4 shows the result of MAP classification (i.e., $\varepsilon = 1$) on three slices from a test dataset (slice no. 9, 17, 21 over 70 total slices). For the sake of comparison, the same results are compared with simple thresholding in Figure 5. A global threshold has been chosen for the whole dataset, and optimised by hand by an expert operator. Due to the lack of space, only one of the worst cases (Slice no. 21) is shown for comparison. Although different thresholds could be chosen for each slice, it would be too much human-intensive. Moreover, global thresholds on the HU values alone cannot do a good job, as already pointed out. Using smarter thresholding techniques does not lead to significant improvements because they miss important features, lacking a robust classification mechanism. This simple experiment highlights the advantage of using a simple classification rather than thresholding on HU values.

In order to test the influence of the reject rule on the final result, we classified our test datasets using $\varepsilon \in \{0.85, 0.90, 0.95, 1\}$. A slice of the same test dataset is shown in Figure 6. Clearly, there is no noticeable difference between the results. Hence, the result is not strongly affected by the choice of ε .

Figure 7 shows a 3D mesh model of the knee joint generated (by means of the Marching Cubes [8] algorithm) from a CT dataset in which the cortical bone has been classified using the proposed method.

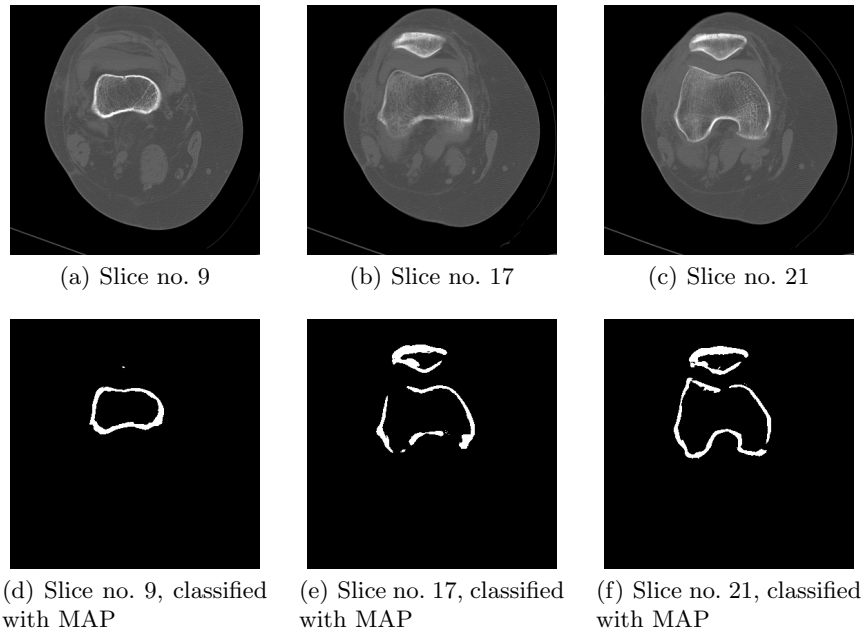


Fig. 4. Examples of three slices classified with MAP.

4 Future Work

We have presented a method to classify cortical bone pixels from CT scans, employing a simple generative model. Manually-labelled data are used to learn the posterior probability of two classes of tissues: *cortical bone* and *other*. Classification is carried out using multiscale features and a MAP criterion. A rejection rule is used to bound classification uncertainty. 3D mesh models are generated from classified datasets.

The proposed method can be easily extended to cope with more classes, provided that more labelled data is available. In order to gather more data, we are currently run a systematic acquisition and labelling campaign on aged patients affected by osteoporosis or arthritis.

The proposed model will be extended by using superpixels rather than pixels for classification. We expect that discrimination will be improved when working with more than two classes. Computation will be speed up as well.

References

1. Battiato, S., Bosco, C., Farinella, G.M., Impoco, G.: 3D CT segmentation for clinical evaluation of knee prosthesis operations. In: Proceedings of the Fourth Eurographics Italian Chapter Conference. (2006)

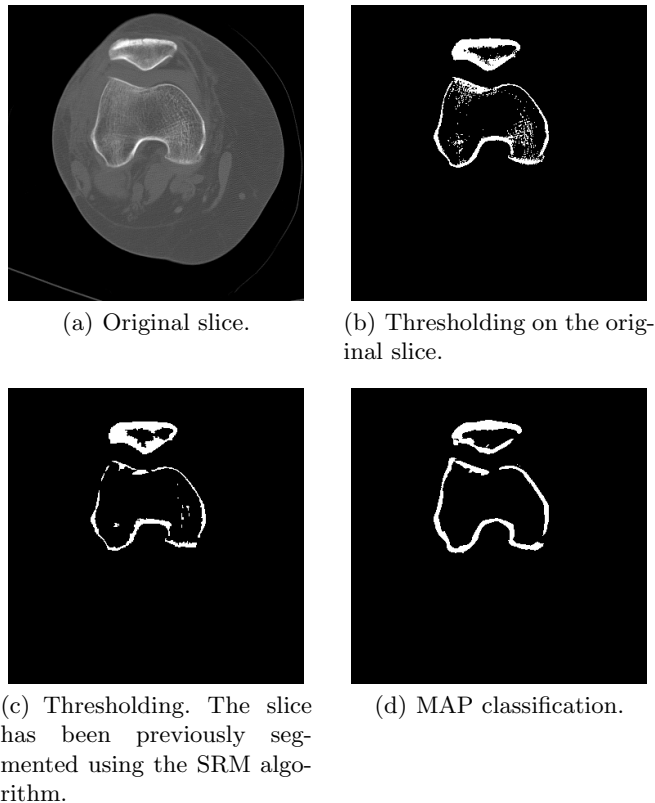


Fig. 5. Comparison between the proposed MAP classification and simple thresholding of Slice no. 21 of the same dataset as Figure 4. A global threshold has been chosen for the whole dataset, and optimised by hand by an expert operator.

2. Farinella, G.M., Impoco, G., Gallo, G., Spoto, S., Catanuto, G., Nava, M.: Objective outcome evaluation of breast surgery. *Lecture Notes in Computer Science* **4190/2006** (2006)
3. Kang, Y., Engelke, K., Kalender, W.A.: A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. *IEEE Transactions on Medical Imaging* **22** (2003) 586–598
4. Ollé, K., Erdőhelyi, B., Halmi, C., Kuba, A.: MedEdit: A computer assisted planning and simulation system for orthopedic-trauma surgery. In: *8th Central European Seminar on Computer Graphics Conference Proceedings*. (2004)
5. Tzacheva, A., El-Kwae, E., Kellam, J.: Model-based bone segmentation from digital X-ray images. In: *Proceedings of the Second Joint EMBS/BMES and IEEE Engineering in Medicine and Biology*. (2002)
6. Petterson, J., Knutsson, H., Borga, M.: Automatic hip bone segmentation using non-rigid registration. In: *Proceedings of the IEEE International Conference on Pattern Recognition*. (2006)
7. Lu, H., Liang, Z., Li, B., Li, X., Meng, J., Liu, X.: Mixture-based bone segmentation and its application in computer aided diagnosis and treatment planning.

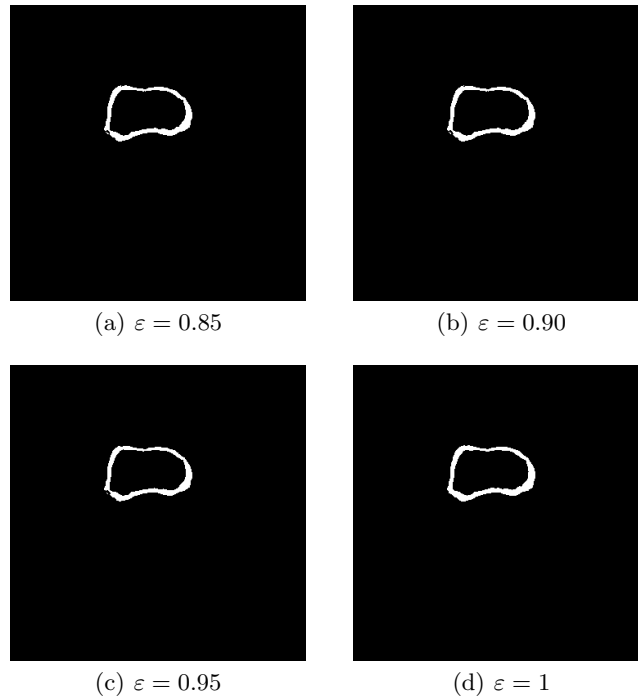


Fig. 6. Comparison between the classification with different values of ε . Slice no. 9 of the same test dataset as Figure 4 is used.

- In: Proceedings of the Third International Conference on Image and Graphics (ICIG'04), Washington, DC, USA, IEEE Computer Society (2004) 507–510
8. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. In: ACM Computer Graphics (SIGGRAPH 87 Proceedings). Volume 21. (1987) 163–170
 9. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
 10. Barrett, J.F., Keat, N.: Artifacts in CT: recognition and avoidance. Radiographics **24** (2004) 16791691
 11. Nock, R., Nielsen, F.: Statistical region merging. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 1452–1458
 12. Nock, R., Nielsen, F.: Semi-supervised statistical region refinement for color image segmentation. Pattern Recognition **38** (2005) 835–846
 13. Ren, X., Malik, J.: Learning a classification model for segmentation. In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. (2003)

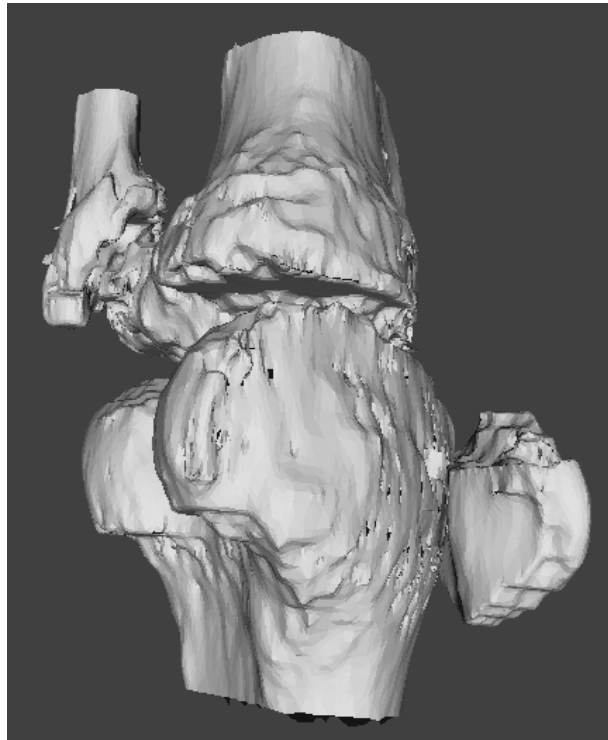


Fig. 7. A 3D model of the knee joint generated from a classified CT dataset.