

Unsupervised Segmentation of Ultrasound Images by Fusion of Spatio-Frequential Textural Features

S. Benameur^a, M. Mignotte^b and F. Lavoie^a

^aEiffel Medtech Inc., Montréal, Québec, Canada;

^bDépartement d'Informatique et de Recherche Opérationnelle,
Université de Montréal, Québec, Canada

ABSTRACT

Image segmentation plays an important role in both qualitative and quantitative analysis of medical ultrasound images. However, due to their poor resolution and strong speckle noise, segmenting objects from this imaging modality remains a challenging task and may not be satisfactory with traditional image segmentation methods. To this end, this paper presents a simple, reliable, and conceptually different segmentation technique to locate and extract bone contours from ultrasound images. Instead of considering a new elaborate (texture) segmentation model specifically adapted for the ultrasound images, our technique proposes to fuse (i.e. efficiently combine) several segmentation maps associated with simpler segmentation models in order to get a final reliable and accurate segmentation result. More precisely, our segmentation model aims at fusing several K -means clustering results, each one exploiting, as simple cues, a set of complementary textural features, either spatial or frequential. Eligible models include the gray-level co-occurrence matrix, the re-quantized histogram, the Gabor filter bank, and local DCT coefficients. The experiments reported in this paper demonstrate the efficiency and illustrate all the potential of this segmentation approach.

Keywords: Segmentation, ultrasound images, fusion of spatio-frequential, textural features

1. INTRODUCTION

Nowadays, ultrasound imaging is considered to be a safe, non-invasive, portable, and non-expensive visualization medical tool compared to other medical imaging techniques. Automatic ultrasonic detection and segmentation of anatomical structures would allow various clinical applications but is a real challenge due to speckle noise and image artifacts. More precisely, the poor image quality (i.e. low signal-to-noise ratio, low resolution and contrast), and artifacts caused by blood and muscles make contours of anatomical structures, such as bone contours, sometimes imperceptible in certain zones.

Many segmentation models, specifically adapted for ultrasound imaging systems, have already been proposed in the literature (see¹ for a good review of existing segmentation methods). Most of these techniques are generally similar to those commonly used for the segmentation of natural textured images; frequently, a first step consists of extracting the textural features computed on (possibly) overlapping small windows centered around the pixel to be classified. Then, the feature samples, handled as vectors, are grouped together, as much as possible, in compact and well-separated clusters corresponding to each class of the image. Following this general strategy, and in order to offer a better texture description capacity, the fusion of different textural cues, generated from different texture models, has been recently proposed to improve the segmentation results. To this end, some authors have proposed to fuse different textural feature sets by simple concatenation, causing in this way the well-known "curse of dimensionality" problem. This concept recognizes that there is a limit to the feature space dimension due to the finite sample size (i.e. conceptually, if n samples are required to estimate a parameter or a region class, n^N features would be required for the increased N -dimensional feature space²).

To overcome this problem, some authors have investigated various feature selection methods to find the optimal (i.e. complementary and non-redundant) combination (or subset) of texture features. Amongst these

Further author information: (Send correspondence to S. Benameur)

S. Benameur: E-mail: sbenameur@eiffelmedtech.com, Telephone: 1 514 380 8666

F. Lavoie: E-mail: fredericlavoie.md@gmail.com, Telephone: 1 514 803 9784

Medical Imaging 2011: Image Processing, edited by Benoit M. Dawant, David R. Haynor,
Proc. of SPIE Vol. 7962, 796239 · © 2011 SPIE · CCC code: 1605-7422/11/\$18 · doi: 10.1117/12.871172

techniques, we can briefly cite those including dimensionality (i.e. feature) reduction techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) - more generally, those exploiting the mutual information criterion - or those using the Fisher's method which seeks the optimal combination of the initial textural features to achieve compact and well-separated clusters, hereby ensuring both minimum scattering between samples of a same class (also called *within-class inertia*) and maximum separability between classes (*between-class inertia*) in the feature space.³

Very few papers have investigated a methodology to combine or fuse different textural features that both do not explicitly use a dimensionality reduction model and are not affected by the above-mentioned "curse of dimensionality" problem. Amongst these rare existing textural feature fusion approaches, we can however cite the methods that simply compute spatial and local statistics on a frequential (Fourier, Gabor, or wavelet) decomposition of the image. Instead of considering spatial- statistics-based textural features concatenated with the magnitude responses of frequential filter outputs, Cheng *et al.* have proposed to use the co-occurrence matrix features computed from a multi-scale directional filter bank.⁴ In the same way, Woumer *et al.*⁵ have shown that first and second order spatial statistics computed on the wavelet details coefficients could improve texture characterization. Finally, Clausi *et al.*⁶ have proposed to replace the high-frequency Gabor filter features with the co-occurrence probability features which seem to be more suitable and appropriate features for detecting higher frequency components of the texture. These above-mentioned methods allow to mix in an ad-hoc way two different textural feature sets generated by two different texture models. However, they are not general enough to be applied to other textural feature sets.

The fusion approach proposed in this paper is conceptually different and explores a new strategy introduced in;⁷ in fact, instead of fusing different textural features generated from different texture models, our technique rather explores the possible alternative of fusing (i.e. efficiently combining) several segmentation results from simpler segmentation models, with each segmentation to be fused using a single set of textural features derived from a single texture model. Our fusion strategy aims at combining these segmentation maps with a simple final clustering procedure using, as input features for each pixel to be classified, the set of local histogram values of the class (or *texton*) labels of each segmentation to be fused. By using this strategy, our fusion method is neither affected by the "curse of dimensionality" problem nor concerned by the feature space normalization problem*.

Conceptually, our fusion strategy is in the framework of the so-called *decision fusion* approaches recently proposed in pattern recognition or machine learning.⁸⁻¹² With these methods, a series of energy functions are first minimized before their outputs (i.e. their decisions) are merged. More generally, the concept of combining classifiers for the improvement of the performance of individual classifiers is known, in machine learning field, as a committee machine or mixture of experts.^{13,14} In this context, Dietterich¹⁴ have provided an accessible and informal reasoning, from statistical, computational, and representational viewpoints, of why ensembles can improve results. In this recent field of research, two major categories of committee machines are generally found in the literature. Our fusion decision approach is in the category that utilizes an ensemble of classifiers with a static structure type. In this class of committee machines, the responses of several classifiers are combined by means of a mechanism that does not involve the input data[†].

2. SEGMENTATIONS TO BE FUSED

The four initial segmentation maps which will be fused together are simply given, in our application, by a K -means clustering technique (with the Euclidean similarity distance), using four sets of (complementary) spatial and frequential (local or semi-global) textural feature sets as simple textural cues. The feature sets are computed on overlapping small square N_w -fixed-size windows centered around the pixel to be classified. The textural cues are the gray level co-occurrence matrix (GLCM),¹⁵ the local gray level quantified histogram, the magnitude responses of Gabor filter outputs,¹⁶ and a subset of DCT coefficients. Figure ?? depicts the flow diagram of the proposed method.

*when different features with different units are blended together, the normalization step prevents that the similarity measure, used to evaluate the distance between feature vectors, will wrongly give an overwhelming importance to features having a larger unit range.

[†]contrary to the dynamic structure type-based mixture of experts

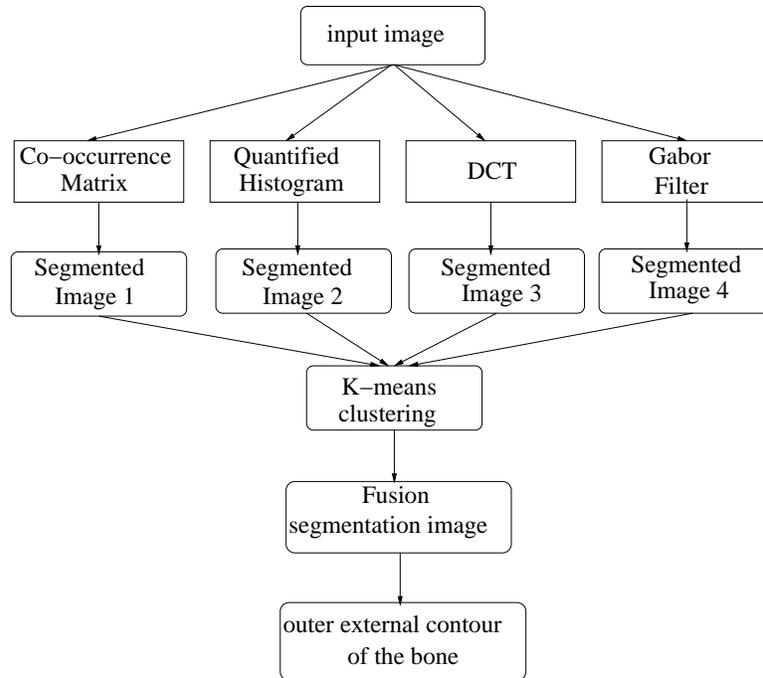


Figure 1. Flow diagram of proposed method.

2.1 Spatial Textural Features

Co-occurrence Matrix: The co-occurrence matrix proposed by Haralick *et al.* contains the relative frequencies of all pairwise combinations of grey levels (i, j) at a certain distance d_x and d_y and orientation θ . From this matrix, 14 statistics (angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy and other two information measures of correlation) are computed allowing identification of some qualitative characteristics of the local image texture (e.g., smoothness, directivity, energy, etc.). Prior to the computation of this co-occurrence matrix, the number of grey levels N_g in the input image needs to be reduced to a small number to get reliable estimates.

'Angular second moment' is a measure of homogeneity of an image. The higher value of this feature indicates that the intensity varies less in an image. 'Contrast' measures local variation in an image. A high contrast value indicates a high degree of local variation. 'Correlation' is a measure of linear dependency of intensity values in an image. For an image with large areas of similar intensities, correlation is much higher than for an image with noisier, uncorrelated intensities. 'Variance' indicates the variation of image intensity values. For an image with identical intensity for all images, the variance would be zero. 'Inverse difference moment' is another feature to represent image contrast. 'Sum average' and 'sum variance' are the average and variance of normalized greytone image in the spatial domain, respectively. The 'sum entropy' is a measure of randomness within an image and 'entropy' is an indication of the complexity within an image. A complex image produces a high entropy value. The 'difference variance' is an image variation in a normalized. The 'difference entropy' is also an indication of the amount of randomness in an image. More detailed information of algorithms to calculate texture features was given by Park *et al.*¹⁷

Quantified Histogram: The textural features are simply the set of values of the re-quantized gray-level histogram with equidistant binning N_b . Contrary to the co-occurrence matrix, these textural features are able to characterize the grey level mixture of the local image texture semi-globally because they are estimated on a small square window centered around the pixel to be classified.

Formally, for each pixel i of a gray-scale input image, the bins descriptor denoted $h(i) = \{h(n, i)\}_{n \in \{0, \dots, N_b - 1\}}$

is computed by the following procedure

$$h(n, i) = \kappa \sum_{j \in N(i)} \delta(b(j) - n)$$

where δ is the Kronecker delta function, $\kappa = \frac{1}{N_w^2}$ is a normalization constant ensuring $\sum_{n=0}^{N_b-1} h(n, i) = 1$, $N(i)$ is a set of pixel within the $N_w \times N_w$ neighborhood region centered at pixel i

2.2 Frequential Textural Features

Discrete Cosine Transform: Due to its nice decorrelation, energy compaction properties, and robustness to noise, the first N_{DCT} Discrete Cosine Transform (DCT) coefficients can efficiently be used to characterize a local image texture.¹⁸ These DCT coefficients are also estimated on overlapping small square N_w -fixed-size windows centered around the pixel to be characterized.

The equation used for the DCT calculation of a gray-scale input image is given as follows:

$$C(u, v) = \alpha(u)\alpha(v) \frac{2}{N_w} \sum_{j=0}^{N_w-1} \sum_{i=0}^{N_w-1} I(i, j) \cos\left(\frac{(2i+1)u\pi}{2N_w}\right) \cos\left(\frac{(2j+1)v\pi}{2N_w}\right)$$

where $I(i, j)$ is the gray scale value at the (i, j) coordinate position in the image, n is the size of window, $C(u, v)$ is DCT domain representation of $I(i, j)$ image. u, v represent vertical and horizontal frequencies. x, y, u, v have values from 0 to 7, $N_w = 8$ and

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u = 0 \\ 1 & \text{otherwise} \end{cases}$$

Gabor Filter Bank: Due to its appealing ability to model the filter characteristics of human vision and its optimum joint spatial and frequential localization, the Gabor function is attractive to characterize a local image texture.

A two-dimensional Gabor function consists of a sinusoidal plane wave of some frequency and orientation, modulated by a two-dimensional Gaussian. The Gabor filter in the spatial domain is given by

$$G_{\lambda\theta\psi\sigma\gamma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)$$

where

$$\begin{cases} x' = x \cos(\theta) + y \sin(\theta) \\ y' = y \cos(\theta) - x \sin(\theta) \end{cases}$$

In this equation, λ represents the wavelength of the cosine factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function in degrees, ψ is the phase offset in degrees, and γ is the spatial aspect ratio and specifies the ellipticity of the support of the Gabor function, and σ is the standard deviation of the Gaussian determines the (linear) size of the receptive field.

The λ parameter represents the wavelength or the spatial frequency $\frac{1}{\lambda}$. The term $\frac{\sigma}{\lambda}$ determine the bandwidth of the filter. The relationship between b, λ and σ is as follows:

$$b = \log_2 \frac{\frac{\sigma}{\lambda} \pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda} \pi - \sqrt{\frac{\ln 2}{2}}}$$

$$\frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2 \cdot 2^b + 1}{2 \cdot 2^b - 1}}$$

In our application, we use, as textural features, the magnitude responses of Gabor filter outputs and Gabor kernels with 4 orientations, 4 scales σ and 4 wavelengths λ .¹⁹

3. FUSION OF SEGMENTATIONS

Our fusion technique aims at combining these 4 segmentation maps with a simple K -means clustering technique using, as high-level textural cues, the set of local histogram values of the class label of each one of the 4 segmentations to be fused.⁷ For a fusion of 4 segmentations with K_1 classes into a segmentation with K_2 classes, the preliminary feature extraction step thus yields 4 (K_1 -bin) histograms which are then gathered together in order to form a $K_1 \times 4$ -dimensional feature vector or a final ($K_1 \times 4$)-bin histogram. This histogram is then normalized to sum to one, so that it is also a probability distribution function. The proposed fusion procedure is then simply considered as a problem of clustering local histograms of (preliminary estimated) class labels computed around and associated to each site. To this end, we use, once again, a K -means clustering procedure exploiting, for this fusion step, an histogram-based similarity measure derived from the Bhattacharya[‡] similarity coefficient.

It is worth mentioning that the textural cues used by our final fusion procedure can be considered as high-level or multi-model since each class label (of each segmentation to be fused) can be viewed as a *texton*[§] label according to its textural model. Besides, our procedure allows to efficiently combine the different discriminant abilities and characteristics (spatio-frequential and local-global) of each texture model. More precisely, our procedure allows to blend both the abilities of accurately capturing lower frequency textural information of the frequential Gabor and spatial quantified histogram models and the capabilities of capturing higher frequency textural characteristics with the spatial co-occurrence and frequential DCT models.

4. EXPERIMENTAL RESULTS

4.1 Set Up

In all the experiments, we have thus considered the following set-up. The size of the squared window N_w , used to compute all the local textural features was set to 11 except for the DCT coefficient where $N_w = 8$. For the co-occurrence matrix, $d_x = d_y = 1$ and $\theta = 4$ directions, namely (0° , 45° , 90° , 135°). An average was taken on these 4 direction so that these features were rotation invariant. The quantification level N_g was set to 25. For the quantized histogram, the number of bins was set to $N_b = 5$. For the DCT based textural features, we used the first $N_{\text{DCT}} = 15$ lowest frequency coefficients. For the Gabor based textural features, we used the magnitude responses of Gabor filter outputs with 4 scales $\sigma \in \{1, 2, 4, 8\}$ and 4 wavelengths $\lambda \in \{3.7, 7.4, 14.8, 29.6\}$. Finally, for the fusion procedure, we used $K_1 = 4$ and $K_2 = 4$.

We have tested our fusion procedure on ultrasound images of several bones acquired using a portable B-mode ultrasound imaging system (Titan, SonoSite, Bothell, WA, USA).

4.2 Results & Discussion

The echographic appearance of the various tissues ranges from dark (low-echoic) to bright (high-echoic), depending on their acoustic impedance (Figure 2). This physical characteristic is used to identify the bone structures on the echographic images, knowing that the outer surface of bones is highly echoic and will appear as a bright structure on echographic images.

Figure 3 shows K -means-based ultrasound image segmentations with the 4 sets of textural features previously mentioned. Each initial segmentation to be fused succeeded more or less to identify these two important regions,

[‡]The Bhattacharya distance between a normalized histogram $h(n; i)$ and a reference histogram $h'(n)$ is defined as

$$Dist_{Bhatt}(h', h(i)) = \sqrt{1 - \sum_{n=0}^{N_b-1} \sqrt{h'(n)h(n, i)}}$$

[§]the repetitive character or element of a textured image, also called a texture primitive

without ambiguity, even with a K -means-based over-segmentation with $K = 4$ (Figure 2). Our fusion approach allows to find several regions where one corresponds to the outer surface of the bone. The outer contour of the bone is then easily localized by searching the contour which delineates two regions (namely bone and soft tissues) for which the magnitude of the luminance gradient level, computed on this contour, is the highest on average. Fig. 4 shows two examples of segmentation maps fused using our technique. The segmentation and fusion procedures take about 12 seconds on an Intel Centrino 2 Processor, 2.8 GHz and running on Linux.

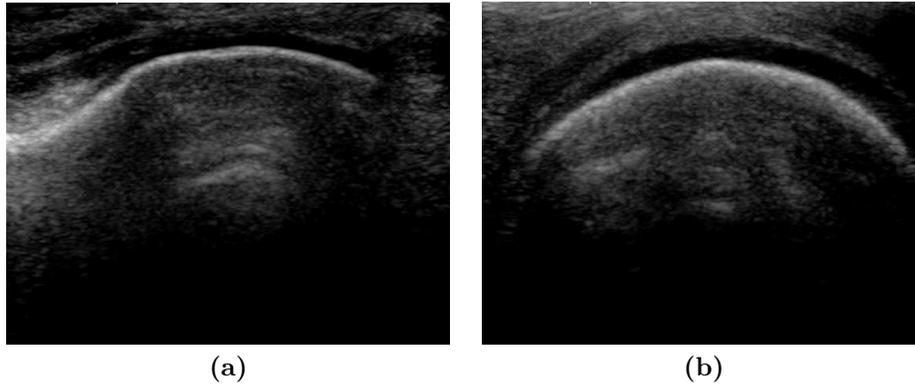


Figure 2. *Ultrasound images of the distal femur; (a) Medial side, coronal plane; (b) Medial posterior condyle, axial plane.*

Figure 5 shows examples of outer region and outer external contour of a same bone under different angles of view. The ability to precisely localize bone contours has several potential therapeutic applications, including radiotherapy guidance and computer-assisted orthopaedic surgery. Indeed, if the location of a tumor relative to the bone it invades is known through magnetic resonance or CT imaging, precisely defining the position of the bone in a non-invasive fashion by using ultrasound imaging can improve the accuracy of the radiation beam and therefore decrease healthy tissue irradiation. In computer-assisted orthopaedic surgery, knowledge of the location and morphology of the bones can increase the precision level of various actions like fracture reduction and bone cuts.

5. CONCLUSION

In this paper, we have presented a segmentation strategy based on a fusion procedure whose goal is to combine several segmentation maps in order to get a more reliable and accurate segmentation result. We have validated this procedure with ultrasound images acquired with a portable ultrasound machine in the case of unsupervised segmentation of the outer contour of the femur. This fusion method remains simple, fast, easily parallelizable, reliable, and is neither affected by the "curse of dimensionality" problem nor concerned by the feature space normalization problem.

REFERENCES

- [1] Noble, J. and Boukerroui, D., "Ultrasound image segmentation: a survey," *IEEE Trans. on Medical Imaging* **25**, 987–1010 (Aug. 2006).
- [2] Jain, A. and Chandrasekaran, B., [*Dimensionality and sample size consideration in pattern recognition practice*].
- [3] Mignotte, M., Meunier, J., Soucy, J.-P., and Janicki, C., "Segmentation and classification of brain spect images using 3d markov random field and density mixture estimations," in [*5th World Multi-Conference on Systemics, Cybernetics and Informatics, SCI'01, Concepts and Applications of Systemics and Informatics*], 239–244 (July 2001).
- [4] Cheng, K., law, N., and Siu, W., "Co-occurrence features of multi scale directional filter bank for texture characterization," in [*Proceeding of IEEE*], 5499–5502 (2006).
- [5] Woumer, G., Scheunders, P., and Dyck, D., "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. on Image Processing* **8**, 592–598 (Apr. 1999).

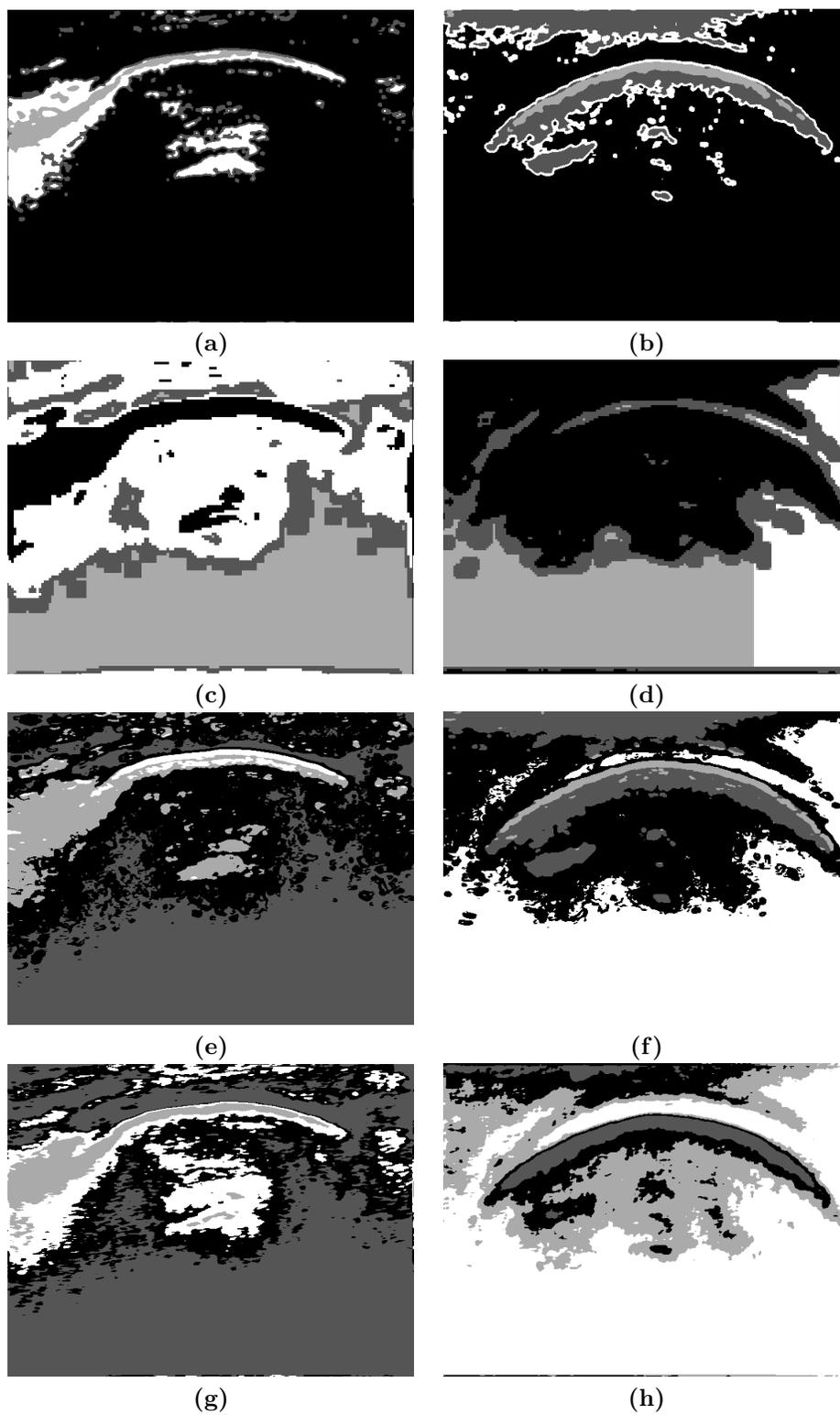


Figure 3. *K-means-based Segmentation maps achieved on original images of Figure 2 by the following texture model: (a), (b) Quantified histogram; (c), (d) Co-occurrence matrix; (e), (f) Gabor filter; (g), (h) Discrete cosine transform.*

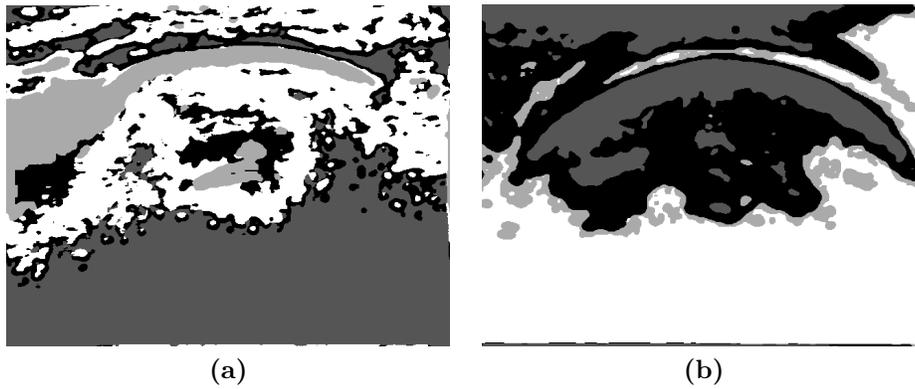


Figure 4. Fusion segmentation maps: (a) a, c, e and g of Figure 3.(a). (b) b, d, f and h of Figure 3.(b).

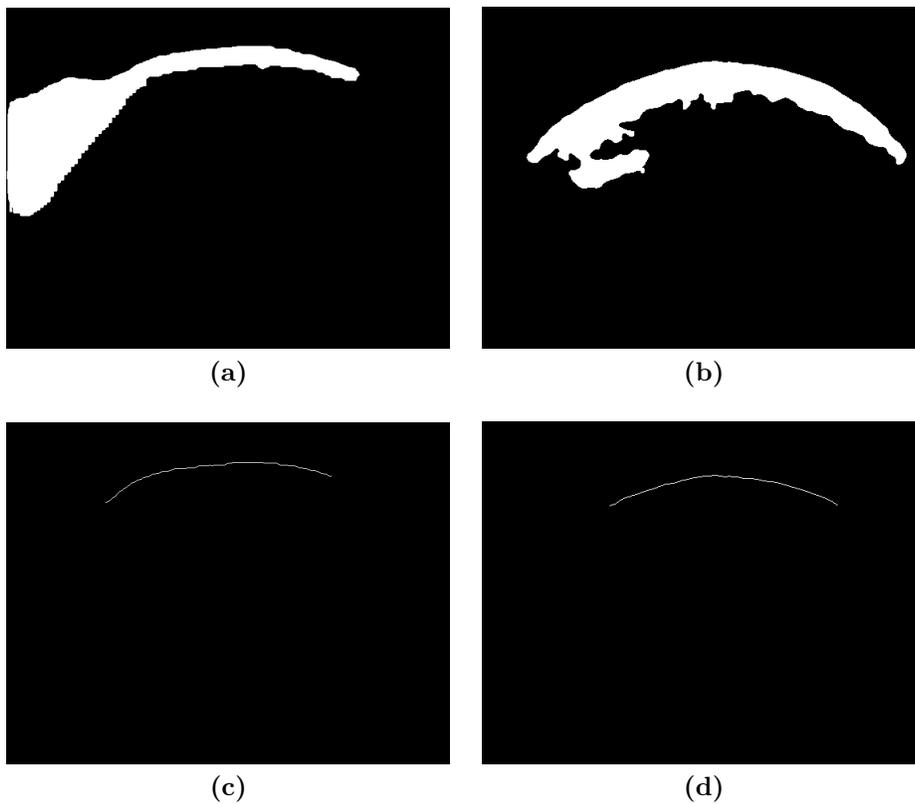


Figure 5. Segmentation results : (a) (b) outer region of the bone. (c) (d) outer external contour of the bone.

- [6] Clausi, D. and Deng, H., "Design based texture feature fusion using gabor filters and co-occurrence probabilities," *IEEE Trans. on Image Processing* **14**, 925–936 (July 2005).
- [7] Mignotte, M., "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Trans. on Image Processing* **17**(5), 780–787 (May 2008).
- [8] Fred, A. L. N. and Jain, A. K., "Combining multiple clusterings using evidence accumulation," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**, 835–850 (June 2005).
- [9] Jabri, S., Duric, Z., Wechsler, H., and Rosenfeld, A., "Detection and localization of people in video images using adaptive fusion of color and edge information," in [*Proc. of the International Conference on Pattern recognition (ICPR)*], 4627–4631 (September 2000).
- [10] Kittler, J., Hatef, M., Duin, R., and Matas, J., "On combining classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**, 226–239 (March 1998).
- [11] Kushki, A., Androustos, P., Plataniotis, K., and Venetsanopoulos, A., "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. on Image Processing* **13**, 277–292 (March 2004).
- [12] Reed, S., Ruiz, I. R., Capus, C., and Petillot, Y., "The fusion of large scale classified side-scan sonar image mosaics," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **15**, 2049–2060 (July 2006).
- [13] Haykin, S., [*Neural Networks: A comprehensive Foundation*], Prentice Hall, Inc. (1999).
- [14] Dietterich, T., "Ensemble methods in machine learning," in [*Proceedings of the First International Workshop on Multiple Classifier Systems, LNCS, Multiple Classifier Systems*], Science, L. N. I. C., ed., **1857**, 1–15, Springer (2000).
- [15] Haralick, R., Shanmugam, K., and Dinstein, I., "Texture features for image classification," *IEEE Trans. on Systems, Man, and Cybernetics, SMC-* **3**(6), 610–621 (1973).
- [16] Alezn-Flores, M., Alezn-Flores, P., lvarez Len, L., Esteban-Snchez, M., Fuentes-Pavn, R., and Santana-Montesdeoca, J., [*Lecture Notes in Computer Science*], ch. Texture-Based Filtering and Front-Propagation Techniques for the Segmentation of Ultrasound Images, 960–967, no. 4739, Springer Verlag Berlin Heidelberg (2007).
- [17] Park, B. and Chen, Y., "Multispectral image co-occurrence matrix analysis for poultry carcasses inspection," *Transactions of the ASAE* **4**(39), 1485–1491 (1996).
- [18] Feng, G. and Jiang, J., "JPEG compressed image retrieval via statistical features," *Pattern Recognition* **36**, 977–985 (Apr. 2003).
- [19] Jain, A. and Farrokhnia, F., "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition* **24**(12), 1167–1186 (1991).