IET Science, Measurement & Technology

Research Article

IET Journals
The Institution of Engineering and Technology

Validation of an automatic segmentation method to detect vertebral interfaces in ultrasound images

ISSN 1751-8822
Received on 18th March 2015
Revised on 29th July 2015
Accepted on 25th August 2015
doi: 10.1049/iet-smt.2015.0042
www.ietdl.org

Matteo Aventaggiato[1], Francesco Conversano[2], Paola Pisani[2], Ernesto Casciaro[2], Roberto Franchini[2], Aimè Lay-Ekuakille[3], Maurizio Muratore[4], Sergio Casciaro[2] ✉

[1]*Echolight Srl, Lecce, Italy*
[2]*National Research Council, Institute of Clinical Physiology, Lecce, Italy*
[3]*Department of Innovation Engineering, University of Salento, Lecce, Italy*
[4]*O.U. of Rheumatology, Galateo Hospital, San Cesario di Lecce, ASL-LE, Lecce, Italy*
✉ *E-mail: sergio.casciaro@cnr.it*

Abstract: Aim of this study was to perform a detailed clinical validation of a new fully automatic algorithm for vertebral interface segmentation in echographic images. Abdominal echographic scans of lumbar vertebrae L1–L4 were carried out on 150 female subjects with variable age and body mass index (BMI). Acquired datasets were automatically processed by the algorithm and the accuracy of the obtained segmentations was then evaluated by three independent experienced operators. Obtained results showed a very good specificity in vertebra detection (93.3%), coupled with a reasonable sensitivity (68.1%), representing a suitable compromise between the detection of a sufficient number of vertebrae for reliable diagnoses and the limitation of the corresponding computation time. Importantly, there was only a minimum presence of 'false vertebrae' detected (2.8%), resulting in a very low influence on subsequent diagnostic analyses. Furthermore, the algorithm was specifically tuned to provide an improved sensitivity (up to 73.1%) with increasing patient BMI, to keep a suitable number of correctly detected vertebrae even when the acquisition was intrinsically more difficult because of the augmented thickness of abdominal soft tissues. The proposed algorithm will represent an essential added value for developing echographic methods for the diagnosis of osteoporosis on lumbar vertebrae.

1 Introduction

Osteoporosis and related fractures are becoming an important worldwide public health problem, associated with a heavy economic burden [1]. For proper management and prevention of osteoporosis, the introduction of cheap and user-friendly methods for population screening is required [2].

Currently, the most adopted method for osteoporosis diagnosis is dual X-ray absorptiometry (DXA), which provides a measure of bone mineral density (BMD). Unfortunately, DXA is not available in primary care settings nor as a screening tool, because of the issues related to ionising radiation employment (i.e. high costs, device bulkiness, need of dedicated structures with certified operators). Nevertheless, DXA is presently considered the gold standard reference for BMD measurement and osteoporosis diagnosis [2–5].

A possible alternative is represented by the employment of dedicated ultrasound (US) techniques, which are radiation-free, portable, less expensive than DXA and available in primary care settings. In fact, bone health assessment is just one out of the several biomedical applications of US systems that have been introduced in the last years because of their mentioned intrinsic advantages over competing technologies [6–16]. In this context, the effectiveness of several quantitative US (QUS) methods for bone health assessment and osteoporosis diagnosis has been recently investigated [17–20]. However, although QUS approaches have the potential to combine mineral density ('bone quantity') with mechanical strength ('bone quality') [18, 20–23], their actual accuracy is still somewhat uncertain since different studies aimed at evaluating clinical effectiveness of QUS systems reported contradictory findings [24–28].

An US device for osteoporosis diagnosis was recently developed in Lecce (Italy) by our research group within the ECHOLIGHT Project through a collaboration between National Research Council, Institute of Clinical Physiology (CNR-IFC) and Echolight srl. The important aspects of this system are [29–36]: (i) US examinations are carried out through a clinically-available echographic transducer that works in reflection mode and is employable on lumbar spine and proximal femur, which are the reference anatomical sites for osteoporosis diagnosis but cannot be investigated by current commercially-available QUS devices; (ii) diagnostic information is obtained by a novel algorithm that performs a series of spectral and statistical analyses combining both echographic images, used to identify and select the region of interest (ROI), and the corresponding 'raw' unprocessed radiofrequency (RF) signals, employed to analyse the bone structure; (iii) diagnostic measurements are always performed on a specific ROI within the target bone, which is automatically identified on the basis of both morphological and spectral characteristics; (iv) data analysis takes into account patient body mass index (BMI); (v) the algorithm is integrated with a database containing reference spectral models of osteoporotic and healthy bones for different combinations of patient age, sex, BMI and anatomical site.

The automatic identification of target bone and ROI is a critical task, whose actual effectiveness has a direct influence on final output accuracy, especially in the case of spinal investigations, which are carried out through abdominal US scans. In this case, in fact, the proper segmentation of vertebral surfaces is made difficult by factors such as the strong variability of abdominal anatomy from subject to subject and the presence in the image of various echogenic interfaces (e.g. abdominal aorta walls) that could be misinterpreted as vertebral surfaces. For these reasons, the typical algorithms adopted for vertebral segmentations in biomedical images obtained through magnetic resonance imaging or X-ray applications are ineffective on US acquisitions [37–45].

In a recent conference paper [46], we reported a preliminary clinical validation of the first version of our method for automatic

segmentation of vertebral interfaces in echographic images. In the present work we describe the implementation of an updated version of our segmentation algorithm, in which two additional parameters were introduced, and a more extended and accurate clinical validation was performed.

2 Materials and methods

2.1 Data acquisition

The US device developed within the ECHOLIGHT Project (see previous section), equipped with a 3.5-MHz convex clinically-available echographic transducer, was used to carry out abdominal US scans of lumbar vertebrae L1–L4 on 150 female subjects with variable age (range: 56–65 y; mean: 60.7 ± 2.7 y) and BMI (range: 18.5–40.5 kg/m^2; mean: 26.7 ± 4.1 kg/m^2). The adopted probe frequency belongs to the range typically employed for abdominal US investigations in clinical contexts, since it ensures a good compromise between the image resolution and the required penetration depth.

US scans were carried out by a researcher who was not used to perform echographic acquisitions but received just a 3-h specific training session. For each subject, the operator kept the probe on subject's abdomen and performed a 20-s sagittal acquisition of the first lumbar vertebra, and then he moved to the next vertebra and so on, until the ventral side of the entire lumbar tract L1–L4 had been acquired. Each scan generated a total of 100 frames of RF data digitised at 40 MS/s that were stored in a computer hard-disk for offline analysis. Scan depth and transducer focus were set for each acquisition in order to keep vertebral interfaces in the US focal region and in the central part of the image, taking into account that the thickness of soft tissues between skin and vertebrae was usually larger for subjects with higher BMI values. The other echographic parameters were the same for all the US scans: power = 45%, gain = 0 dB, mechanical index = 0.4, and linear time gain compensation.

2.2 Segmentation algorithm

The flow chart of the implemented algorithm is reported in Fig. 1. For each acquired frame, RF signals were band-pass filtered in 1.0–2.5 MHz by a finite impulse response filter, in order to reduce the possible noise presence at frequencies lower than 1 MHz and higher than 2.5 MHz, emphasising the contributions of the most penetrating frequencies within the transducer bandwidth.

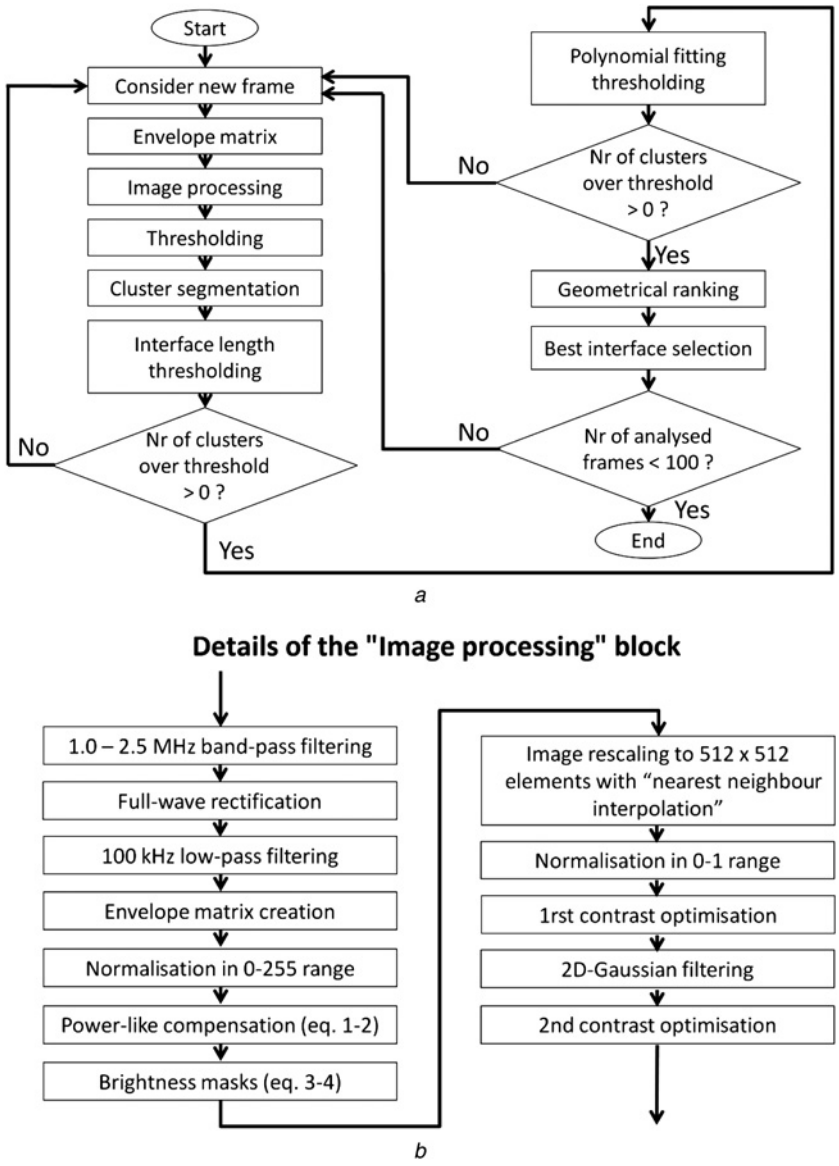

Fig. 1 *Flow chart of*

a Implemented segmentation algorithm

b Zoom of the 'Image processing' block. The corresponding effects of single processing steps on echographic images are shown in Figs. 3 and 4

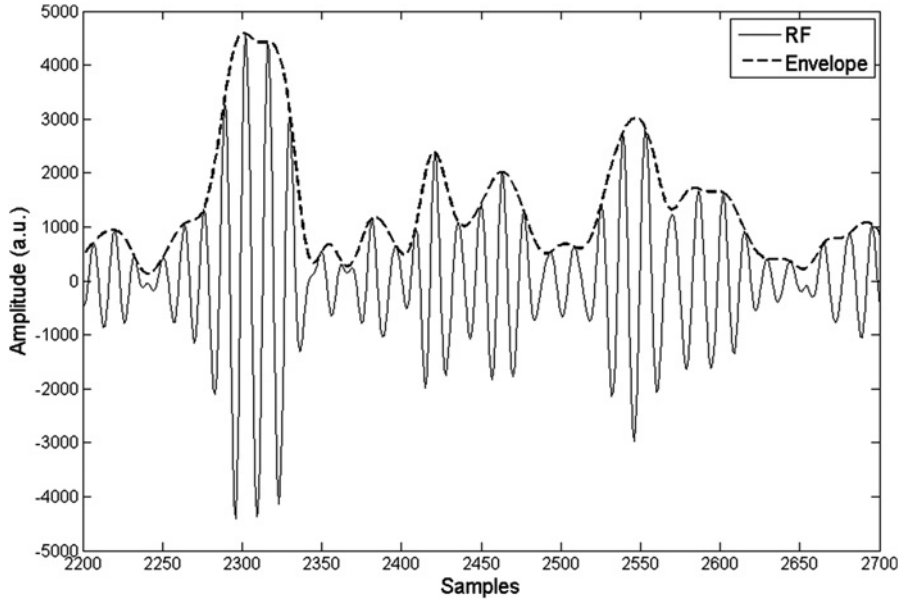

Fig. 2 *Typical RF signal (dashed line) and the corresponding envelope (continuous line)*

Afterwards, in order to obtain a full-wave rectification, the absolute value of each RF signal sample was calculated and the obtained signals were 100-kHz low-pass filtered. The output of this step was the envelope of the original RF signal (see Fig. 2), which was then employed to generate the B-Mode US images used in the next steps of the segmentation algorithm.

The RF signal envelopes of the considered frame were organised in a matrix called *Env_mat* and having size $N_{\text{pix}} \times N_{\text{lines}}$, where N_{pix} is the number of samples per scan line, which is proportional to scan depth, and N_{lines} is the number of scan lines (253 in each frame). *Env_mat* values were then normalised in the range 0–255 and graphically represented as an echographic image that is considered as the starting point of the actual segmentation process (Fig. 3*a*).

Since vertebral surfaces were expected to be in the central part of the image, *Env-mat* underwent a power-like compensation in order to attenuate upper and lower image portions. This was accomplished through a matrix called *Exp_tgc*, which had $N_{\text{pix}} \times N_{\text{lines}}$ elements and was defined as

$$Exp\_tgc[i,j] = \begin{cases} \dfrac{i}{0.3 \cdot N_{\text{pix}}}; & i \le 0.25 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ \dfrac{i + 0.2 \cdot N_{\text{pix}}}{0.5 \cdot N_{\text{pix}}}; & 0.25 \cdot N_{\text{pix}} < i \le 0.55 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ 1.5; & 0.55 \cdot N_{\text{pix}} < i \le 0.75 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ \dfrac{1.4 \cdot N_{\text{pix}} - i}{0.5 \cdot N_{\text{pix}}}; & 0.75 \cdot N_{\text{pix}} < i \le 0.9 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ \dfrac{N_{\text{pix}} - i}{0.1 \cdot N_{\text{pix}}}; & 0.9 \cdot N_{\text{pix}} < i \le N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \end{cases} \tag{1}$$

The values of the numeric coefficients reported in (1) were empirically obtained in order to optimise the segmentation algorithm results by highlighting the central area of the image, where, usually, the vertebral interface is expected to be. Each element $[i, j]$ of *Env_mat* was raised to the power represented by the element $[i, j]$ of *Exp_tgc*, obtaining the matrix *Env_mat2* (corresponding to the image reported in Fig. 3*b*)

$$Env\_mat2[i,j] = Env\_mat[i,j]^{Exp\_tgc[i,j]} \tag{2}$$

The matrix *Env_mat2* was then multiplied element by element by two brightness masks (Mask_1 and Mask_2), which emphasised the central image portion along horizontal and vertical directions, respectively, and were defined as

$$\text{Mask}_1[i,j] = \begin{cases} \dfrac{2j + N_{\text{lines}} - 4}{2 \cdot N_{\text{lines}} - 4}; & 1 \le i \le N_{\text{pix}} \quad 1 \le j \le \dfrac{N_{\text{lines}}}{2} \\ \dfrac{3 \cdot N_{\text{lines}} - 2 - 2j}{2 \cdot N_{\text{lines}} - 4}; & 1 \le i \le N_{\text{pix}} \quad \dfrac{N_{\text{lines}}}{2} + 1 \le j \le N_{\text{lines}} \end{cases} \tag{3}$$

$$\text{Mask}_2[i,j] = \begin{cases} 0.5; & i \le 0.18 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ \dfrac{i + 0.14 \cdot N_{\text{pix}}}{0.64 \cdot N_{\text{pix}}}; & 0.18 \cdot N_{\text{pix}} < i \le 0.5 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ 1; & 0.5 < i \le 0.6 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ \dfrac{1.2 \cdot N_{\text{pix}} - i}{0.6 \cdot N_{\text{pix}}}; & 0.6 \cdot N_{\text{pix}} < i \le 0.9 \cdot N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \\ 0.5; & 0.9 \cdot N_{\text{pix}} < i \le N_{\text{pix}} \quad 1 \le j \le N_{\text{lines}} \end{cases} \tag{4}$$

The image obtained after the application of brightness masks as defined in (3) and (4) is shown in Fig. 3*c*. As for (1), also the numeric coefficients reported in (3) and (4) were empirically obtained to optimise the emphasising effect on the image central area.

The resulting data matrix was rescaled to 512×512 elements by means of the 'nearest neighbour interpolation' and normalised in the range 0–1; then, it underwent the hereinafter described series of image processing steps. Image contrast was optimised through a 'contrast-limited adaptive histogram equalisation': first, the image was divided into 64 identical rectangular regions (called 'tiles'), each tile's histogram was then equalised and the neighbouring tiles were combined using bilinear interpolation to eliminate artificially induced boundaries (the corresponding obtained image is reported in Fig. 3*d*). Afterwards, the image was filtered by using a two-dimensional low-pass Gaussian filter having size = 50×50 and SD = 5 (the resulting image is shown in Fig. 4*a*) and a further 'contrast-limited adaptive histogram equalisation' was applied (providing the image reported in Fig. 4*b*).

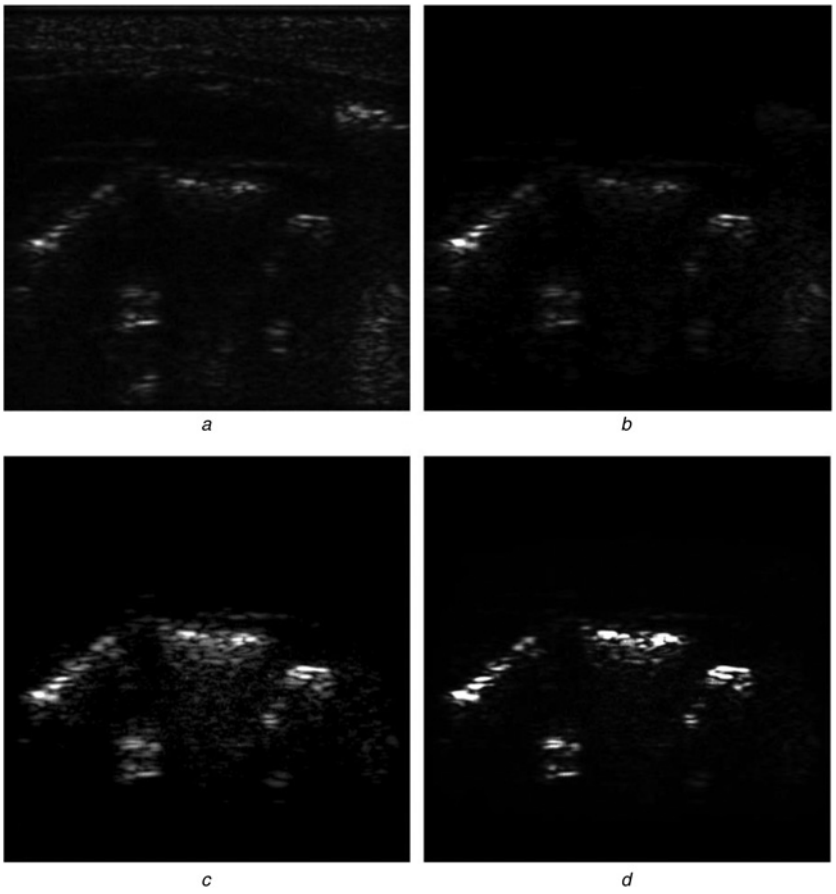

Fig. 3 *Effect of single processing steps on a typical echographic image (first part)*

a Initial echographic image (simply obtained from the RF signal envelope)
b Image obtained after the application of (1)
c Image resulting from the application of brightness masks as defined in (3) and (4)
d Image obtained after the contrast optimisation operated through the contrast-limited adaptive histogram equalisation

After the described image processing steps, the image was thresholded (threshold value = 0.95) to become a binary matrix (corresponding to the black and white image in Fig. 4*c*). At this stage, the image matrix dimensions were restored to $N_{pix} \times N_{lines}$ in order to assure an unambiguous correspondence with RF signal matrix and related envelopes.

All the clusters of white pixels present in the thresholded image were sequentially numbered (Fig. 4*d*) and, in order to take into account the shape of retained clusters and to verify if they actually corresponded to a vertebral interface, for each of them, the local maxima of single scan lines were identified in the corresponding envelope matrix and interpolated by a second order polynomial curve, whose endpoints were located on the outer scan lines. Clusters presenting the absolute value of the first polynomial coefficient higher than 0.3 were discarded, otherwise they underwent the subsequent evaluation. Afterwards, the length of each identified cluster was measured as the distance between the second order polynomial endpoints: if the result was outside the range 20–45 mm, which represents the typical physiological height range of lumbar vertebrae, the cluster was discarded. If no cluster satisfied both the length and shape requirements, the frame was discarded and a new frame was considered, otherwise all the clusters having both length and shape coefficient in the sought ranges were ranked according to the following criteria:

(a) *length:* 2 points were assigned to the longest cluster and 1 point to the second one (motivation: once the clusters whose length was not compatible with a vertebral height had been discarded, this criterion promoted the longest interfaces, since they cross more scan lines and therefore have the potential to provide more diagnostic information);
(b) *average vertical thickness:* 2 points were assigned to the thinnest cluster and 1 point to the second one (motivation: empirical trials documented that the 'optimal' interaction between US beam and vertebral surface produces a sharp thin white line on the echographic image, therefore this criterion promotes the interfaces presenting the minimum average thickness);
(c) *vertical position:* 2 points were assigned to the cluster that was the closest to the image centre along the vertical direction and 1 point to the second one (motivation: this criterion promotes the interfaces that are better positioned with respect to the US beam focal region);
(d) *lateral position:* 2 points were assigned to the cluster that was the closest to the image centre along the horizontal direction and 1 point to the second one (motivation: the closer a cluster is to the image centre along the horizontal direction the less it is influenced by side lobe effects and boundary artefacts, therefore this criterion promotes the interfaces associated to a better signal quality);
(e) *parallelism to probe surface:* 2 points were assigned to the cluster whose inclination angle with respect to the propagation direction of the incident US beam was the closest to 90° and 1 point to the second one (motivation: this criterion promotes the interfaces that are parallel to incident US frontwaves, since they provide the maximum backscatter);
(f) *relative contrast:* 2 points were assigned to the cluster whose coordinates were associated to the maximum average value in the *Env_mat* matrix and 1 point to the second one (motivation: this criterion promotes the interfaces characterised by the highest average brightness, in order to provide the subsequent algorithm for diagnostic calculation with data characterised by minimum noise influence).

At the end of the described ranking process, the cluster that received the highest total score was assumed to be representative

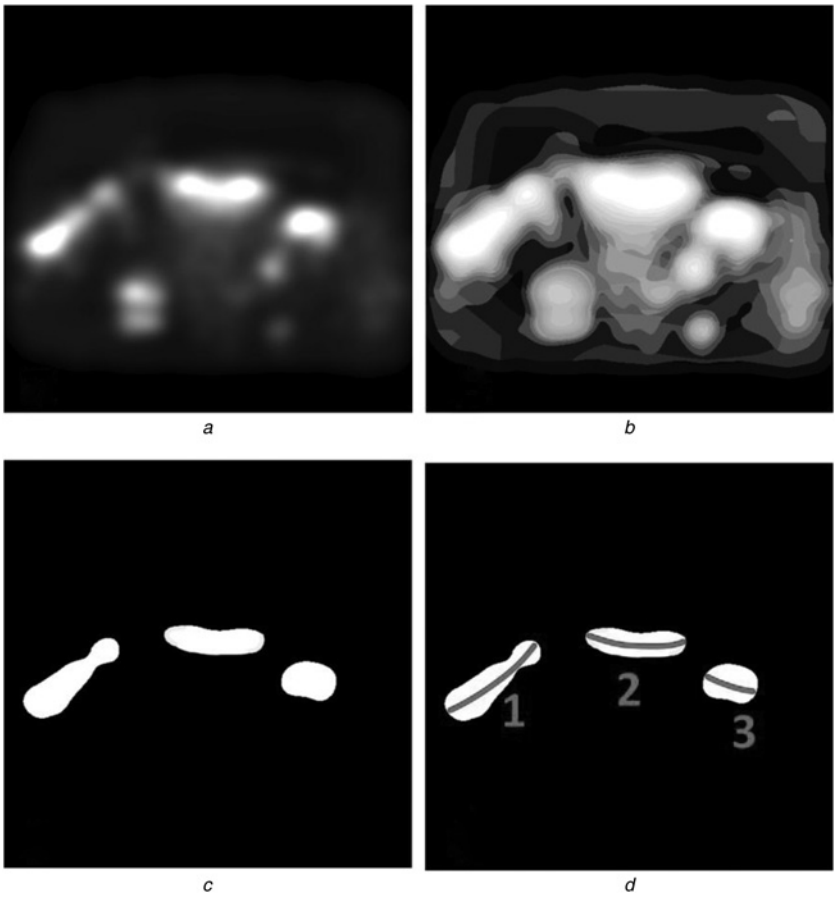

Fig. 4 *Effect of single processing steps on a typical echographic image (second part)*

a Image obtained after the application of the two-dimensional Gaussian filter
b Image resulting from the second application of the contrast-limited adaptive histogram equalisation
c Thresholded image, showing clusters of white pixels corresponding to vertebral interface areas
d Vertebral interfaces (red lines) obtained by interpolating the local maxima of single scan lines with a second order polynomial (after the subsequent 'geometrical ranking' step, the best interface was selected, which, in this case, was the one labelled with '2')

of an actual vertebral interface and the previously calculated second order polynomial was sent to the 'convex image generation' process (see next paragraph), in order to highlight the identified vertebral interface with a red line superimposed on the corresponding sectorial echographic image.

2.3 Convex image generation

For each acquired data frame, the envelope matrix *Env_mat* defined in the previous paragraph underwent also a parallel processing, in order to obtain the conventional sectorial echographic images usually produced by a clinically-available convex transducer and to highlight on them the possible vertebral interfaces identified by the segmentation algorithm. This was accomplished through the following processing steps:

(a) *Env_mat* was multiplied element by element by the brightness masks Mask_1 and Mask_2 defined by (3) and (4);
(b) the resulting matrix was filtered by a two-dimensional low-pass Gaussian filter (size = 10×10; SD = 1), in order to reduce the possible high-frequency noise;
(c) the matrix was transformed into a sectorial image by taking into account radius and aperture angle of the employed convex US transducer;
(d) the obtained image was compressed to an 8-bit image;
(e) if the segmentation algorithm described in the previous paragraph detected a vertebral interface, the corresponding second order polynomial was used to superimpose a red line on the image in correspondence of the identified vertebral surface, otherwise this step was skipped;
(f) the image was saved in Portable Network Graphics (PNG) format.

A typical obtained image is reported in Fig. 5, showing a conventional sectorial echographic image in which the vertebral interface automatically identified by the algorithm was highlighted with a red line. It is important to note that two further vertebral interfaces were visible in Fig. 5, but the algorithm selected the 'optimal' one (best position and best geometric properties).

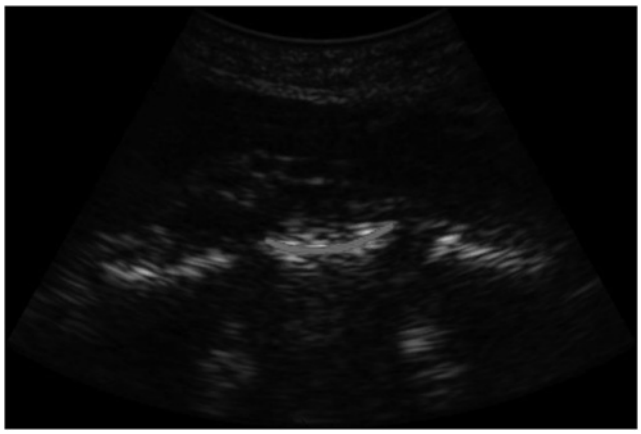

Fig. 5 *Typical echographic image with the automatically identified vertebral interface emphasised by the bold red line*

2.4 Expert operator validation

The accuracy of the implemented algorithm was evaluated by three independent experienced operators, who had previously received a specific training to recognise vertebral interfaces in abdominal echographic images and to identify the most suitable ones for our diagnostic purposes. In our study, the validation procedure of each frame was independently repeated by more than one operator because, in certain conditions, the possible presence of artefacts (such as abdominal air bubbles or abnormal soft tissue attenuation) could affect the frame validation depending on the personal operator experience.

For each investigated subject, all the 100 PNG images corresponding to the 100 acquired RF frames, obtained at the end of the convex image generation process (see previous paragraph) were assessed by each expert operator, who was always blind with respect to other operator findings. Based on operator's judgment, each image frame was classified as 'true positive' (if there was a red line highlighting an 'optimal' vertebral interface, Fig. 6*a*), 'true negative' (if there were no red lines nor suitable vertebral interfaces, Fig. 6*b*), 'false positive' (if there was a red line highlighting something different from a vertebral interface or if the highlighted interface was not the best one, Fig. 6*c*), or 'false negative' (if there were no red lines, although an 'optimal' vertebral interface was present, Fig. 6*d*).

Sensitivity and specificity of the automatic algorithm in vertebral surface identification according to the response of each operator were computed through the following formulas

$$\text{sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \tag{5}$$

$$\text{specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}} \tag{6}$$

Sensitivity and specificity values were considered both as a function of patient BMI and on the whole enrolled population.

To evaluate the consistency of the segmentation algorithm, 100 true positive frames were randomly selected for each expert operator and the Absolute Length Error (ALE) and Relative Length Error (RLE) were calculated as

$$\text{ALE} = |L_\text{A} - L_\text{O}| \tag{7}$$

$$\text{RLE} = \frac{|L_\text{A} - L_\text{O}|}{L_\text{O}} \cdot 100\% \tag{8}$$

where L_A is the length of the vertebral interface calculated by the segmentation algorithm (see par. 2.2) and L_O is the length of the vertebral interface identified by the expert operator.

The actual algorithm accuracy was also evaluated through the standard F_1-score

$$F_1 = 2 \cdot \frac{p \cdot r}{p + r} \tag{9}$$

where p and r are respectively the *precision* and *recall* calculated as

$$p = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \tag{10}$$

$$r = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \tag{11}$$

3 Results and discussion

Tables 1–3 report the results of the automatic segmentation validation performed by expert operator #1, #2, and #3 respectively. Table 4 shows the values of ALE and RLE calculated according to (7) and (8), respectively. Table 5 summarises the accuracy of the segmentation algorithm as expressed by the F_1-score.

All the expert operators documented that the most frequent segmentation errors were 'false negatives' (i.e. frames that presented an undetected vertebral interface), whose average incidence was 18.9%. On the other hand, 'false positives', including both the wrong segmentation of structures different from

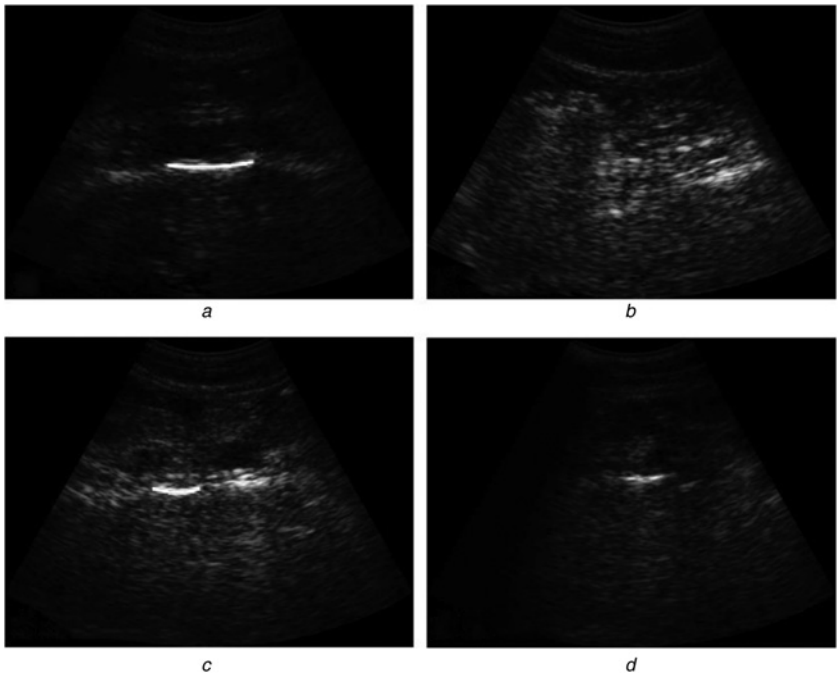

Fig. 6 *Examples of possible classifications attributed by experienced operators to PNG images resulting from the automatic segmentation*

a True positive
b True negative
c False positive
d False negative

Table 1 Results of the automatic segmentation validation performed by the expert operator #1

BMI range, kg/m^2	Number of subjects	Number of frames	True positives, %	True negatives, %	False positives, %	False negatives, %	Sensitivity, %	Specificity, %
<25	80	8000	43.6	32.4	2.0	21.9	66.5	94.1
25–30	48	4800	39.5	41.4	3.3	15.8	71.5	92.6
>30	22	2200	33.1	49.8	4.4	12.7	72.3	91.9
all	150	15,000	40.8	37.8	2.8	18.6	68.7	93.2

Table 2 Results of the automatic segmentation validation performed by the expert operator #2

BMI range, kg/m^2	Number of subjects	Number of frames	True positives, %	True negatives, %	False positives, %	False negatives, %	Sensitivity, %	Specificity, %
<25	80	8000	43.0	32.2	2.2	22.6	65.6	93.6
25–30	48	4800	39.3	41.6	3.4	15.7	71.4	92.4
>30	22	2200	33.2	51.7	4.4	10.7	75.7	92.1
all	150	15,000	40.4	38.1	2.9	18.7	68.4	92.9

Table 3 Results of the automatic segmentation validation performed by the expert operator #3

BMI range, kg/m^2	Number of subjects	Number of frames	True positives, %	True negatives, %	False positives, %	False negatives, %	Sensitivity, %	Specificity, %
<25	80	8000	42.1	33.4	2.0	22.6	65.1	94.4
25–30	48	4800	38.8	41.0	2.9	17.3	69.2	93.4
>30	22	2200	32.9	49.8	4.2	13.2	71.3	92.3
all	150	15,000	39.7	38.2	2.6	19.5	67.1	93.7

Table 4 Evaluation of the segmentation consistency by means of Absolute Length Error and Relative Length Error (respectively, (7) and (8)) performed by the three expert operators on 100 true positive frames randomly selected

Expert operator	Absolute length error (mean ± standard deviation), mm	Relative length error (mean ± standard deviation), %
1	2.5 ± 2.0	3.2 ± 2.8
2	3.1 ± 3.2	3.8 ± 2.6
3	3.2 ± 2.4	4.0 ± 3.4
mean	2.9 ± 2.1	3.7 ± 2.9

Table 5 Measurement of the segmentation accuracy by means of the F_1-score calculation as reported in (9)

Expert operator	F_1-score, %
1	79.2
2	78.9
3	78.2
mean	78.8

vertebral interfaces and the segmentation of a vertebral interface that was not the best one available in the considered frame, represented only 2.8% of the analysed frames. This is also visually emphasised in Fig. 7, highlighting the strong prevalence of correctly segmented frames (average accuracy = 78.3%) over the total segmentation errors (21.7%), which, in most of cases, are 'false negatives'. Therefore, not only the automatic segmentation process showed a good success rate, but also the mistakes had a negligible influence on the final diagnosis, since there were just few cases in which a non-vertebral structure was identified as a vertebra, and its effects were minimised by statistical analyses performed on samples being mostly composed by actual vertebrae. We also verified that the vertebral detection was not very sensitive to slight movements of the probe: once the operator learnt how to perform sagittal echographic acquisitions of the lumbar vertebrae, which was done during the 3-h specific training session, the vertebrae were segmented with a reproducible rate even in the presence of slight probe movements due, for instance, to patient breathing.

In fact, the proposed approach showed a very high specificity in vertebra detection (over 93% for each operator), coupled with a reasonable sensitivity (68.1% on the average). These results are also visually emphasised in Figs. 8 and 9, respectively.

In particular, Figs. 8 and 9 show that variations in algorithm performance as judged by different operators were essentially negligible, while both specificity and sensitivity presented slight but visible trends as a function of patient BMI. In fact, the average specificity gradually decreased from 94.0% for normal- or under-weight subjects (BMI < 25 kg/m^2) to 92.8% for over-weight ones (BMI in the range 25–30 kg/m^2), finally reaching the minimum value (92.1%) for obese women (BMI > 30 kg/m^2). On the other hand, the average sensitivity showed the opposite behaviour, being equal to 65.7, 70.7 and 73.1%, for normal-/under-weight, over-weight, and obese subjects, respectively.

These trends were expected because of the specific algorithm tuning, which was optimised, on one hand, to be highly specific (even at the expenses of some sensitivity decrements) and, on the other hand, to have a sensitivity increasing with patient BMI. These implementation choices were motivated by the precise reasons herein detailed.

First of all, there was the mentioned necessity of minimising the number of 'false vertebrae' detected (i.e. false positives), in order to assure the reliability of osteoporosis diagnoses obtained from different algorithms, separately developed within the same research project and taking as input data the coordinates of vertebral interfaces automatically segmented by the presently discussed algorithm [31, 33–35]. This explained the decision to favour specificity with respect to sensitivity, which was also due to the empirical observation that the opposite choice had a negative effect on the subsequent diagnostic algorithm, since it caused a significant extension of computational time without any benefit for final accuracy.

Simultaneously, it was important to ensure, for each investigated patient, the correct identification of a suitable number of vertebral interfaces, taking into account that the difficulty of the corresponding segmentation increased with patient BMI because of

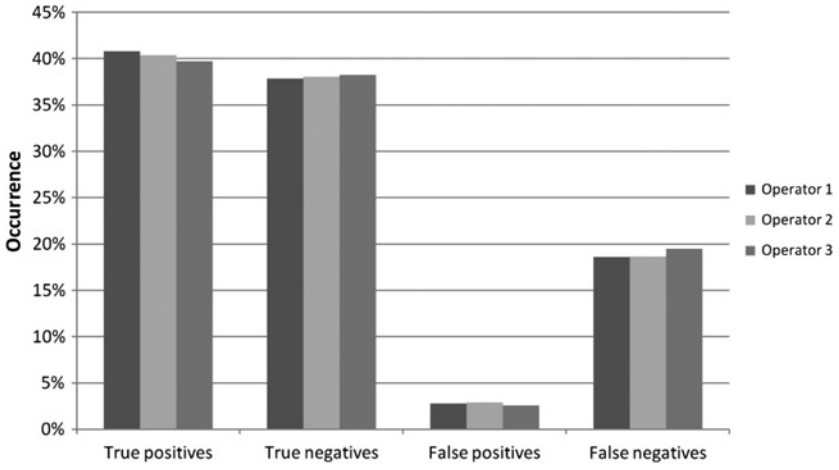

Fig. 7 *Relative occurrence of possible frame classifications for each experienced operator, referred to the whole sample of analysed frames (N = 15,000 for each operator)*

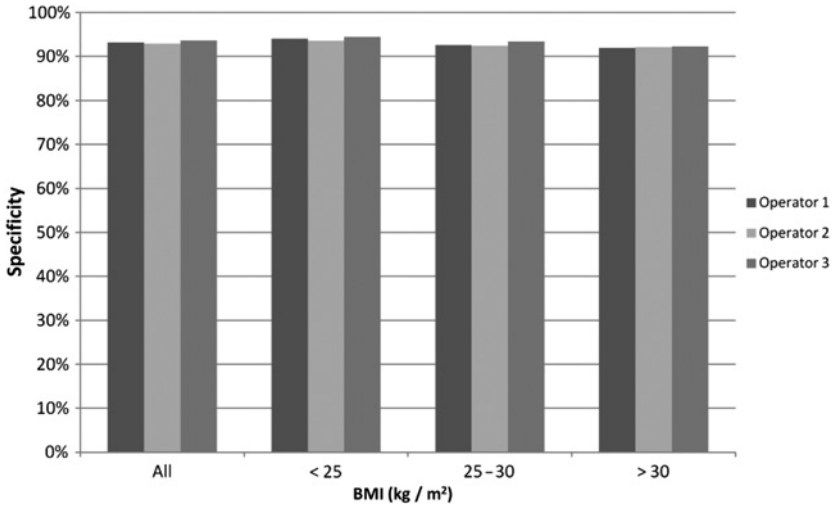

Fig. 8 *Specificity of the automatic segmentation as a function of BMI for each experienced operator*

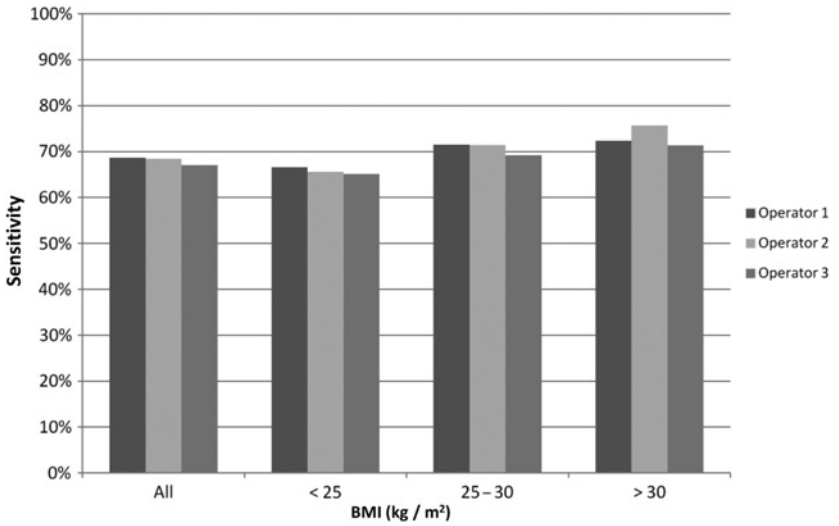

Fig. 9 *Sensitivity of the automatic segmentation as a function of BMI for each experienced operator*

the augmented thickness of intervening soft tissues. In fact, the actual percentage of frames containing a vertebral interface suitable for diagnostic analyses (given by the sum of 'true positives' and 'false nagatives' in Tables 1–3) decreased from 65.3% for normal-/underweight subjects to 55.5% for over-weight ones, and to 45.3% for obese women. Therefore, in order to

compensate for the progressively reduced availability of appropriate frames with increasing patient BMI, the automatic segmentation algorithm was opportunely tuned to provide a corresponding gradually better sensitivity (at the expenses of the reported slight decrements in specificity).

To evaluate the segmentation consistency, we used ALE and RLE, which were calculated according to (7) and (8), respectively. As reported in Table 4, the mean value for the three expert operators, evaluated on a random selection of 100 true positive frames, was 2.9 ± 2.1 mm for ALE and $3.7 \pm 2.9\%$ for RLE. Moreover the algorithm accuracy was evaluated by means of the F_1-score, which ranged from 78.2 to 79.2% for the three expert operators. Overall, the algorithm employed in this work was an optimised version of the one presented in [46], thanks to the implementation of two additional parameters in the ranking process of possible vertebral interfaces (*parallelism to probe surface* and *relative contrast*, see par. 2.2). Although the validation procedure adopted in the present paper was more severe than the previous one, since the number of enrolled patients was increased by 50% and the judging expert operators were three in place of just one, the global effectiveness of the proposed automatic segmentation was substantially confirmed: average values of specificity and sensitivity were 93.3 and 68.1%, respectively, against the corresponding 93.4 and 69.1% obtained in the referred work [46], which provided a F_1-score of 79.6%.

To further improve the algorithm performance, future studies will experimentally evaluate the opportunity of differentiating algorithm settings for each specific vertebra to be recognised in the lumbar tract L1–L4 and, possibly, using proper machine learning algorithms to improve frame classification in terms of specificity and sensitivity. Moreover, a proper real time processing could be tested to improve the B-Mode view and help the operator in the identification of deep vertebral interfaces.

4 Conclusions

The implemented method for automatic segmentation of vertebral interfaces in echographic images was validated by three experienced operators on 15,000 image frames acquired on 150 patients during routine clinical activity. The overall average results documented a very high specificity (93.3%), which was a very important requisite in order to avoid deteriorating the accuracy of subsequent analyses, and a reasonable sensitivity (68.1%), anyway representing a suitable compromise between the detection of a sufficient number of vertebrae for reliable diagnoses and the limitation of the corresponding computation time.

Importantly, the algorithm was specifically tuned to provide an improved sensitivity (up to 73.1%) with increasing patient BMI, in order to keep a suitable number of correctly detected vertebrae even when the acquisition was intrinsically more difficult because of the augmented thickness of abdominal soft tissues.

The proposed algorithm will represent an essential added value for all the developing US methods for the assessment of vertebral health assessment based on abdominal echographic acquisitions.

5 Acknowledgment

This work was partially funded by FESR P.O. Apulia Region 2007-2013 – Action 1.2.4 (grant no. 3Q5AX31: ECHOLIGHT Project).

6 References

1 Pike, C., Birnbaum, H.G., Schiller, M., *et al.*: 'Economic burden of privately insured non-vertebral fracture patients with osteoporosis over a 2-year period in the US', *Osteoporos. Int.*, 2011, **22**, pp. 47–56
2 Pisani, P., Renna, M.D., Conversano, F., *et al.*: 'Screening and early diagnosis of osteoporosis through X-ray and ultrasound based techniques', *World J. Radiol.*, 2013, **5**, (11), pp. 398–410
3 Kanis, J.A., McCloskey, E.V., Johansson, H., *et al.*: 'European guidance for the diagnosis and management of osteoporosis in postmenopausal women', *Osteoporos. Int.*, 2013, **24**, pp. 23–57
4 Baim, S., Leslie, W.D.: 'Assessment of fracture risk', *Curr. Osteoporos. Rep.*, 2012, **10**, pp. 28–41
5 Link, T.M.: 'Osteoporosis imaging: state of the art and advanced imaging', *Radiology*, 2012, **263**, pp. 3–17
6 Casciaro, S., Pisani, P., Soloperto, G., *et al.*: 'An innovative ultrasound signal processing technique to selectively detect nanosized contrast agents in echographic images', *IEEE Trans. Instrum. Meas.*, 2015, **64**, (8), pp. 2136–2145
7 Casciaro, S., Conversano, F., Casciaro, E., *et al.*: 'Automatic evaluation of progression angle and fetal head station through intrapartum echographic monitoring', *Comput. Math. Meth. Med.*, 2013, 8 pages, Article id 278978
8 Chiriacò, F., Conversano, F., Soloperto, G., *et al.*: 'Epithelial cells biocompatibility of silica nanospheres for contrast enhanced ultrasound molecular imaging', *J. Nanop. Res.*, 2013, **15**, (7), article number UNSP 1779
9 Soloperto, G., Conversano, F., Greco, A., *et al.*: 'Advanced spectral analyses for real time automatic echographic tissue-typing of simulated tumour masses at different compression stages', *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 2012, **59**, (12), pp. 2692–2701
10 Conversano, F., Soloperto, G., Greco, A., *et al.*: 'Echographic detectability of optoacoustic signals from low concentration PEG-coated gold nanorods', *Int. J. Nanomed.*, 2012, **7**, pp. 4373–4389
11 Conversano, F., Greco, A., Casciaro, E., *et al.*: 'Harmonic ultrasound imaging of nanosized contrast agents for multimodal molecular diagnoses', *IEEE Trans. Instrum. Meas.*, 2012, **61**, (7), pp. 1848–1856
12 Conversano, F., Franchini, R., Lay-Ekuakille, A., *et al.*: 'In vitro evaluation and theoretical modeling of the dissolution behavior of a microbubble contrast agent for ultrasound imaging', *IEEE Sens. J.*, 2012, **12**, (3), pp. 496–503
13 Conversano, F., Casciaro, E., Franchini, R., *et al.*: 'A quantitative and automatic echographic method for real-time localization of endovascular devices', *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 2011, **58**, (10), pp. 2107–2117
14 Casciaro, S., Conversano, F., Musio, S., *et al.*: 'Full experimental modelling of a liver tissue mimicking phantom for medical ultrasound studies employing different hydrogels', *J. Mat. Sci. Mat. M.*, 2009, **20**, (4), pp. 983–989
15 Demitri, C., Sannino, A., Conversano, F., *et al.*: 'Hydrogel based tissue mimicking phantom for in-vitro ultrasound contrast agents studies', *J. Biomed. Mater. Res. B Applied Biomaterials*, 2008, **87**, (2), pp. 338–345
16 Casciaro, S., Palmizio Errico, R., Conversano, F., *et al.*: 'Experimental investigations of nonlinearities and destruction mechanisms of an experimental phospholipid-based ultrasound contrast agent', *Invest. Radiol.*, 2007, **42**, (2), pp. 95–104
17 Breban, S., Padilla, F., Fujisawa, Y., *et al.*: 'Trabecular and cortical bone separately assessed at radius with a new ultrasound device, in a young adult population with various physical activities', *Bone*, 2010, **46**, pp. 1620–1625
18 Nayak, S., Olkin, I., Liu, H., *et al.*: 'Meta-analysis: accuracy of quantitative ultrasound for identifying patients with osteoporosis', *Ann. Intern. Med.*, 2006, **144**, pp. 832–841
19 Paggiosi, M.A., Barkmann, R., Gluer, C.C., *et al.*: 'A European multicenter comparison of quantitative ultrasound measurement variables: the OPUS study', *Osteoporos. Int.*, 2012, **23**, pp. 2815–2828
20 Pais, R., Campean, R., Simon, S.P., *et al.*: 'Accuracy of quantitative ultrasound parameters in the diagnosis of osteoporosis', *Centr. Eur. J. Med.*, 2010, **5**, pp. 478–485
21 Kaufman, J.J., Einhorn, T.A.: 'Ultrasound assessment of bone', *J. Bone Miner. Res.*, 1993, **8**, pp. 517–525
22 Njeh, C.F., Boivin, C.M., Langton, C.M.: 'The role of ultrasound in the assessment of osteoporosis: a review', *Osteoporos. Int.*, 1997, **7**, (1), pp. 7–22
23 Schnitzer, T.J., Wysocki, N., Barkema, D., *et al.*: 'Calcaneal quantitative ultrasound compared with hip and femoral neck dual-energy X-ray absorptiometry in people with a spinal cord injury', *PM R*, 2012, **4**, pp. 748–755
24 El Maghraoui, A., Morjane, F., Mounach, A., *et al.*: 'Performance of calcaneus quantitative ultrasound and dual-energy X-ray absorptiometry in the discrimination of prevalent asymptomatic osteoporotic fractures in postmenopausal women', *Rheumatol. Int.*, 2009, **29**, pp. 551–556
25 Iida, T., Chikamura, C., Aoi, S., *et al.*: 'A study on the validity of quantitative ultrasonic measurement used the bone mineral density values on dual-energy X-ray absorptiometry in young and in middle-aged or older women', *Radiol. Phys. Technol.*, 2010, **3**, pp. 113–119
26 Kwok, T., Khoo, C.C., Leung, J., *et al.*: 'Predictive values of calcaneal quantitative ultrasound and dual energy X-ray absorptiometry for non-vertebral fracture in older men: results from the MrOS study (Hong Kong)', *Osteoporos. Int.*, 2012, **23**, pp. 1001–1006
27 Liu, J.M., Ma, L.Y., Bi, Y.F., *et al.*: 'A population-based study examining calcaneus quantitative ultrasound and its optimal cut-points to discriminate osteoporotic fractures among 9352 Chinese women and men', *J. Clin. Endocrinol. Metab.*, 2012, **97**, pp. 800–809
28 Moayyeri, A., Adams, J.E., Adler, R.A., *et al.*: 'Quantitative ultrasound of the heel and fracture risk assessment: an updated meta-analysis', *Osteoporos. Int.*, 2012, **23**, pp. 143–153
29 Conversano, F., Casciaro, E., Franchini, R., *et al.*: 'A new ultrasonic method for lumbar spine densitometry'. Proc. IEEE Int. Ultrasonics Symp. (IUS), Prague, Czech Republic, July 2013, pp. 1809–1812
30 Conversano, F., Franchini, R., Greco, A., *et al.*: 'A novel ultrasound methodology for estimating spine mineral density', *Ultrasound Med. Biol.*, 2015, **41**, (1), pp. 281–300
31 Casciaro, S., Renna, M.D., Conversano, F., *et al.*: 'Clinical evaluation of a novel ultrasound-based methodology for osteoporosis diagnosis on overweight and obese women'. Proc. Third Imeko TC13 Symp. Measurements Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 45–49

32 Chiriacò, F., Conversano, F., Quarta, E., *et al*.: 'Preliminary clinical validation of a new ultrasound-based methodology for femoral neck densitometry'. Proc. Third Imeko TC13 Symp. Measurements Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 58–61

33 Pisani, P., Greco, A., Renna, M.D., *et al*.: 'An innovative ultrasound-based method for the identification of patients at high fracture risk'. Proc. Third Imeko TC13 Symp. Measurements Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 50–53

34 Renna, M.D., Conversano, F., Pisani, P., *et al*.: 'Diagnostic accuracy of a novel ultrasound-based methodology for spinal densitometry on a cohort of normal- and under-weight women'. Proc. Third Imeko TC13 Symp. Meas. Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 40–44

35 Greco, A., Pisani, P., Soloperto, G., *et al*.: 'Comparison between ultrasound Fragility Score and FRAX® for the assessment of osteoporotic fracture risk'. Proc. Third Imeko TC13 Symp. Meas. Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 54–57

36 Villani, V., Conversano, F., Aventaggiato, M., *et al*.: 'Implementation of a model database for a novel ultrasonic approach to bone evaluation'. Proc. Third Imeko TC13 Symp. Meas. Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 13–17

37 Guglielmi, G., Diacinti, D., Van Kuij, C., *et al*.: 'Vertebral morphometry: current methods and recent advances', *Eur. Radiol.*, 2008, **18**, pp. 1484–1496

38 Stern, D., Njagulj, V., Likar, B., *et al*.: 'Quantitative vertebral morphometry based on parametric modeling of vertebral bodies in 3D', *Osteoporos. Int.*, 2013, **24**, (4), pp. 1357–1368

39 Zeng, Y., Nixon, M.S., Allen, R.: 'Automated segmentation of lumbar vertebrae in digital videofluoroscopic images', *IEEE Trans. Med. Imaging*, 2004, **23**, (1), pp. 45–52

40 Alvarez Ribeiro, E., Nogueira-Barbosa, M.H., Rangayyan, R.M., *et al*.: 'Detection of vertebral plateaus in lateral lumbar spinal X-ray images with Gabor filters'. Conf. Proc. IEEE Engineering Medicine Biology Society, 2010, pp. 4052–4055

41 Iglesias, J.E., de Bruijne, M.: 'Semiautomatic segmentation of vertebrae in lateral X-rays using conditional shape model', *Acad. Radiol.*, 2007, **14**, pp. 1156–1165

42 Zukic, D., Vlasak, A., Dukatz, T., *et al*.: 'Segmentation of vertebral bodies in MR images', in Goesele, M., Grosch, T., Preim, B., *et al.* (Eds.): 'Vision, modeling and visualization' (Eurographics Association, Geneva, Switzerland, 2012), pp. 135–142

43 Ghosh, S., Alomari, R.S., Chaudhary, V., *et al*.: 'Automatic lumbar vertebra segmentation from clinical CT for wedge compression fracture diagnosis', SPIE Proceedings Vol. 7963, Medical Imaging 2011: Computer-Aided Diagnosis, Article ID 796303

44 Roberts, M., Cootes, T., Adams, J.: 'Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models', *Invest. Radiol.*, 2006, **41**, pp. 849–859

45 Lecron, F., Mahmoudi, S.A., Benjelloun, M., *et al*.: 'Heterogeneous computing for vertebra detection and segmentation in X-ray images', *Int. J. Biomed. Imaging*, 2011, **2011**, 12 pages, Article id 640208

46 Aventaggiato, M., Conversano, F., Casciaro, E., *et al*.: 'Automatic segmentation of vertebral interfaces in echographic images'. Proc. Third Imeko TC13 Symp. Meas. Biol. Med. 'New Frontiers in Biomedical Measurements', Lecce, Italy, April 2014, pp. 18–22