Graphical Modeling of Ultrasound Propagation in Tissue for Automatic Bone Segmentation

Firat Ozdemir[(✉)], Ece Ozkan, and Orcun Goksel

Computer-Assisted Applications in Medicine, ETH Zurich, Zurich, Switzerland
ozdemirf@vision.ee.ethz.ch

Abstract. Bone surface identification and localization in ultrasound have been widely studied in the contexts of computer-assisted orthopedic surgeries, trauma diagnosis, and post-operative follow-up. Nevertheless, the (semi-)automatic bone surface segmentation methods proposed so far either require manual interaction or complex parametrizations, while failing to deliver accuracy fit for clinical purposes. In this paper, we utilize the physics of ultrasound propagation in human tissue by encoding this in a factor graph formulation for an automatic bone surface segmentation approach. We comparatively evaluate our method on annotated in-vivo ultrasound images of bones from several anatomical locations. Our method yields a root-mean-square error of 0.59 mm, far superior to state-of-the-art approaches.

1 Introduction

Radiography (e.g. X-ray, CT, fluoroscopy) is the conventional technique for imaging bones, however it involves radiation exposure. Ultrasound (US) has been proposed as a safe, real-time imaging alternative for certain applications such as bone surface localization for diagnosis and routine orthopedic controls, e.g. [1–4]; and for intra-operative guidance in computer-assisted orthopedic surgery (CAOS), e.g. [5,6]. Nevertheless, identifying bone surface is a challenging task, since US suffers from a range of different artifacts and presents low signal-to-noise-ratio (SNR) in general. The methods proposed in the literature require manual interaction or complex parametrizations limiting their generalizability.

Although ultrasound raw radio-frequency can be used to segment bones [7], its availability for routine clinical applications from commercial US machines is still quite limited. Considering conventional B-mode imaging, early work focusing on bone surface segmentation utilized intensity and gradient information, e.g. [8]. Hacihaliloglu et al. [1] exploited phase congruency from Kovesi [9] to introduce *phase symmetry* (PS) in 2D and 3D to identify bone fractures by aggregating log-Gabor filters at different orientations. This enhances bone surface appearance as seen in Fig. 1c. The tedious parameter selection phase of log-Gabor filters for PS was automated later in [10]. Inspired by *gradient energy tensor* from [11], PS was also used to define *local phase tensor* (LPT) metric and was studied for enhancing bone surface appearance for registering statistical shape models to 3D US images [5]. Despite its high sensitivity, the major drawback of PS is its
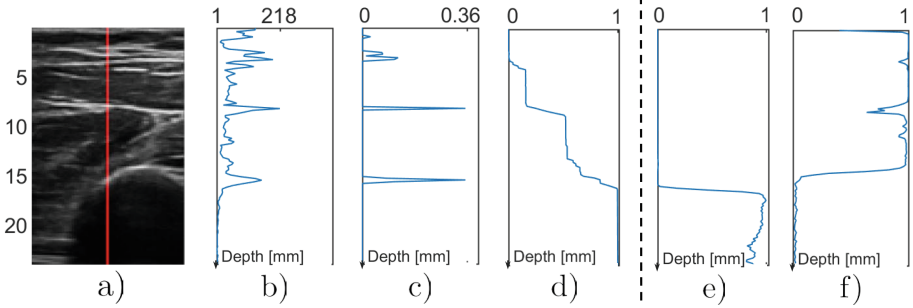
Fig. 1. (a) An in-vivo bone US. The red line indicates the mid-column, along which the plots show: (b) B-mode intensity, (c) PS [1], (d) shadowing feature [12], (e) shadow and (f) soft-tissue probabilities from the trained appearance model cf. Sect. 2.1.

low specificity; i.e. it gives false positives at interfaces between soft tissue layers (Fig. 1c). Therefore, most works using PS alone require manual interaction, e.g. selection of a region-of-interest (ROI) around expected bone surface, or post-processing to remove false positives. Note that PS is a hard-decision, giving almost binary (a very high dynamic range) response, from which post-processing may not always recover from, leading to suboptimal solutions. Alternatively, in [3] *confidence in phase-symmetry* (CPS) was introduced to enhance bone surfaces in US by uniformly weighting PS, attenuation and shadowing features; the latter two stemming from *confidence maps* [12] based on random walks. The shadowing feature is exemplified in Fig. 1d. These earlier works either lack a principled approach to combine the available information, e.g. image appearance and physical constraints of ultrasound, or rely strongly on PS for bone surface. In this paper, we propose a novel graphical model, which is robust to false-positive responses, by introducing physical constraints of ultrasound-bone interaction combined in a principled way with appearance information from a supervised learning framework.

2   Methods

Despite the fact that soft-tissue interfaces and bone surface may both appear as hyperechoic reflections, there is a fundamental difference at bone surfaces: Due to the relatively higher acoustic impedance of cortical bone, it causes an almost total reflection of transmitted ultrasound energy. This leads to a bright surface appearance and, behind this, a dark or incoherent appearance due to lack of ultrasound penetration. Accordingly, we categorize the US scene in three classes: *bone* surface (**B**), *shadow* behind this surface (**S**), and other (soft) *tissue* (**T**). We model their appearance using supervised learning with the following features.

2.1   Image Features and Learned Appearance Models

2D image patch and 1D image column features are employed, the latter approximating the axial propagation of focused beams. Features regarding statistical, textural, and random walks-based information are extracted at different scales as listed in Table 1, where *Scale-space* indicates kernel sizes, i.e. the edge length of square kernels or length of vector kernels. A subset of the features for a sample US image is depicted in Fig. 2. Below they are briefly summarized.
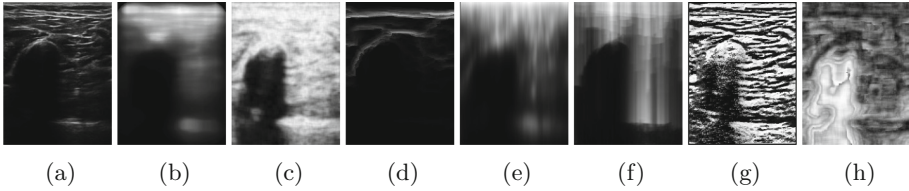

| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |

Fig. 2. Sample features from a US image (a), where $\cdot_n$ denotes filter kernel scale from fine "1" to coarse "3": (b) Median_3, (c) Entropy_2 (d) Attenuation, (e) Gauss_3, (f) Column-long σ, (g) LBP_1, (h) Rayleigh fit_error.

Local-patch statistics. Simple and higher-order statistical features are used.

Random-Walks. Features from the literature such as confidence maps $(m_{x,y})$ from [12]; and, based on this, attenuation $(a_{x,y} = \text{norm}\left(\sum_w (m_{x,y} - m_{\min})\right))$ and shadowing $(s_{x,y} = \text{norm}\left(\sum_w m_{x,y}/m_{\min}\right))$ from [3], where norm(.) is the unity-based normalization and w is number of pixels in the patch are applied.

Column-wise and *integral statistics.* Intuitive metrics motivated by the reflection and attenuation effects acting in a cumulative manner as ultrasound propagates are also employed from the far-side of the image to a point.

Local Binary Patterns. In order to capture textural (speckle) information visually, we used well-known *Local Binary Patterns* [13] and *Modified Census Transform* [14], which relate the intensity at a point to its neighbors.

Speckle characteristics. A last feature is included from an ultrasound physics perspective: It is known that the appearance of fully-developed speckle can be characterized locally by Rayleigh, Nakagami, or similar distributions. At locations where ultrasound SNR is low, e.g. behind bone surface, although it may be possible to get a high intensity (with high gain, etc.), the content would be mostly other (e.g. electrical) noise, which will follow a Gaussian or uniform distribution. Accordingly, we used the fit of a Rayleigh *probability density function* (pdf) to patch intensity histograms to quantify its speckle characteristics, i.e.:

$$\text{fit\_error} = || \, \text{pdf}_{\text{Rayleigh}} - \text{norm}(\text{hist}(\text{patch}_i)) \, || \tag{1}$$

where $\text{pdf}_{\text{Rayleigh}}$ is the maximum likelihood fitted distribution and the second term is the normalized histogram of patch intensities.

Table 1. Features extracted at different kernel space-scales for US transmit wavelength (λ) and pixels (px).

Filter group type	Filter names	Scale-space
Intensity	Pixel intensity (Fig. 2a)	1px
Local patch statistics	Mean, median (Fig. 2b), variance, standard deviation, skewness, kurtosis, entropy (Fig. 2c)	$3,6,12\lambda$
Random Walks	Confidence Map [12], Shadowing [3], logShadowing, Attenuation [3] (Fig. 2d)	1px
Column-wise statistics	0th order (Fig. 2e), 1st order, 2nd order, 3rd order	$2,5,11\lambda$
Integral statistics	Integral, weighted integral, standard deviation	$5,11,31\lambda$
Local Binary Patterns	Local Binary Patterns [13], Modified Census Transform [14]	-
Speckle characteristics	Rayleigh fit_error (Fig. 2h)	12λ

To capture scale-space information, patch-based features are extracted at multiple scales (see Table 1). At a point i, this leads to a feature vector of f_i of length 47 populated by the above-mentioned features. From these features extracted from all image locations of annotated sample images, two discriminative binary classifiers are then trained to construct independent probability functions $p(f_i \,|\, \text{label}_i)$ for classes $\mathbf{S}$ and $\mathbf{T}$, below and above the annotated bones respectively. For bone surface $\mathbf{B}$, we use phase symmetry PS, converted to a likelihood as $e^{\frac{\text{PS}}{-\sigma_0}}$. For a given test image, we cast the bone segmentation as a graph labeling problem shown below.
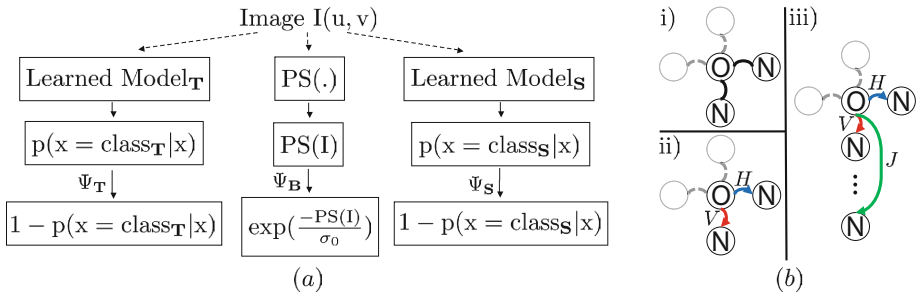
Fig. 3. (a) Unary cost calculation and (b) Pairwise edge connections for (i) 4-connected, (ii) directional 4-connected, (iii) proposed configuration. Horizontal, vertical and jump-edge connections are denoted with H, V and J respectively.

2.2   Encoding Ultrasound Physics on Graph Edges

For spatially consistent results and removing false local responses, *Markov Random Fields* (MRF) is a common regularization approach. In MRF, the image is represented by a graphical model, where pixels are the nodes and inter-pixel interaction (e.g. regularization) are encoded on the edges. A maximum-a-posteriori solution involves the minimization of a cost function in the following form:

$$\sum_i \Psi(i) \ + \ \mu \sum_i \sum_{j \in N_i} \Psi(i,j) \tag{2}$$

where $\Psi(\cdot)$ and $\Psi(\cdot,\cdot)$ are the unary and pairwise cost functions and N_i is the neighbourhood of node i. One can then obtain a regularized labeling (segmentation) solution, e.g., using common Potts potential for pairwise regularization and the label models above as unary costs, as seen in Fig. 3(a).

Table 2. Our pairwise cost definition for horizontal (H), vertical (V) and jump (J) edges.

(a)			
$\Psi^H(i,j)$	$\mathbf{T}(j)$	$\mathbf{B}(j)$	$\mathbf{S}(j)$
$\mathbf{T}(i)$	k_1	1	1
$\mathbf{B}(i)$	1	k_2	1
$\mathbf{S}(i)$	1	1	k_1

(b)			
$\Psi^V(i,j)$	$\mathbf{T}(j)$	$\mathbf{B}(j)$	$\mathbf{S}(j)$
$\mathbf{T}(i)$	k_2	k_3	∞_1
$\mathbf{B}(i)$	∞_1	k_2	k_3
$\mathbf{S}(i)$	∞_1	∞_1	k_2

(c)			
$\Psi^J(i,j)$	$\mathbf{T}(j)$	$\mathbf{B}(j)$	$\mathbf{S}(j)$
$\mathbf{T}(i)$	0	0	∞_2
$\mathbf{B}(i)$	∞_1	∞_3	0
$\mathbf{S}(i)$	∞_1	∞_1	0

MRF uses undirected edges as in Fig. 3(b.i), and thus can only encode bidirectional information. Regarding ultrasound, we know that it travels axially, thus different types of interaction occur between vertical (V) and horizontal (H) pixel neighbours in the image. Different pairwise costs for such neighbours can be set using a directed factor graph as in Fig. 3(b.ii). For horizontal edges, we use a Potts-like model in Table 2(a), where same labels on both ends are penalized less, with parameters k_1 and k_2 in range $(0, 1)$ since neighboring pixels shall be more likely to be of the same class. For vertical edges, what we know is following: 1. soft tissue $\mathbf{T}$ starts from the skin; 2. once the bone surface $\mathbf{B}$ is encountered, the rest of the image (below that location) should be shadow $\mathbf{S}$ (no more $\mathbf{T}$); and 3. $\mathbf{S}$ cannot start without encountering $\mathbf{B}$ first. Constraint 1 above is enforced by a unary constraint on top image pixels (skin), and the latter two are enforced using the vertical pairwise costs in Table 2(b), where ∞_1 prohibits transitions that violate these conditions. Consequently, starting from the transducer the encountered labels (downward) should be in this strict order: $\mathbf{T} \rightarrow \mathbf{B} \rightarrow \mathbf{S}$. Thanks to factor graphs, vertical transitions can also be penalized, if desired, differently than horizontal ones, controlled by parameter k_3 (=1 for isotropic penalty).

Reflection of ultrasound at bone surface generates a hyperechoic band, the thickness of which depends on various factors (e.g., ultrasound frequency).

Accordingly, after the label switching to bone surface **B**, it should not continue as **B** until the bottom of the image, but instead switch to **S** shortly after. We encode this with an additional so-called *jump* edge (J) connected from each pixel to the one l pixels below, as in Fig. 3b.iii (green). With the costs given in Table 2(c), this enforces the thickness of surface appearance to be exactly l pixels: ∞_2 prohibiting **S** below **T**, setting a lower bound of l; and ∞_3 prohibiting both ends from being **B**, setting an upper bound of l. For J, ∞_1 still enforces the right order of transmission. We call this novel connectivity and cost definition as *bone factor graph* (BFG). This is optimized by off-the-shelf tools to obtain segmentation.

3   Results and Discussion

37 US images were acquired using a SonixTouch machine (Ultrasonix, Richmond, Canada) with L14-5 transducer at depths $[3, 5]$ cm with frequencies $\{6.66, 10\}$ MHz (depending on body location). B-mode images had an isotropic pixel resolution of $230\,\mu$m. Collected data include bones in the forearm (radius, ulna), shoulder (acromion, humerus tip), leg (fibula, tibia, malleolus), hip (iliac crest), jaw (mandible, rasmus) and fingers (phalanges). Following [2], bone-surfaces were delineated in the images by an expert at locations where it can be distinguished with certainty; i.e. unannotated columns mean either no bone or not visible.
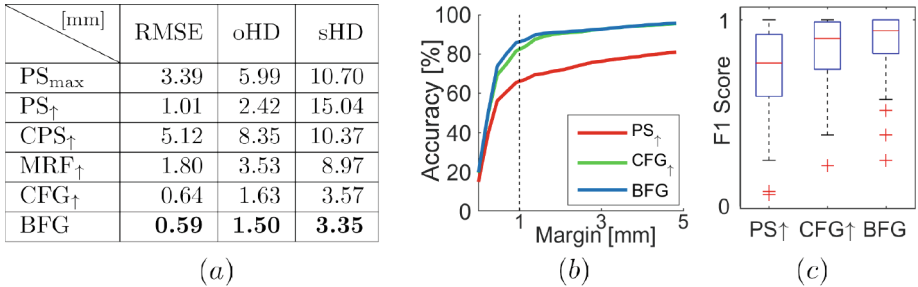
[mm]	RMSE	oHD	sHD
PS_{max}	3.39	5.99	10.70
$PS_\uparrow$	1.01	2.42	15.04
$CPS_\uparrow$	5.12	8.35	10.37
$MRF_\uparrow$	1.80	3.53	8.97
$CFG_\uparrow$	0.64	1.63	3.57
BFG	**0.59**	**1.50**	**3.35**

(a)

(b)

(c)

Fig. 4. (a) Comparison of algorithms (best scores are shown in bold), (b) average accuracy vs. tolerance margin, and (c) F1 score; where we propose $CFG_\uparrow$ & BFG.

We ran 6-fold cross-validation experiments. For learning probability models, L2-regularized logistic regression was used from LIBLINEAR library[1]. Factor graphs were implemented using OpenGM library[2]. For transmit wavelength λ at a given ultrasound frequency, $l = 4\lambda$, $\sigma_0 = 10^{-3}$, $\mu = 1$, $k_1 = k_2 = 0.5$ and $k_3 = 0.3$ are used for the experiments. We utilized the Sequential Tree-Reweighted Message Passing (TRW-S) algorithm for graph optimization.
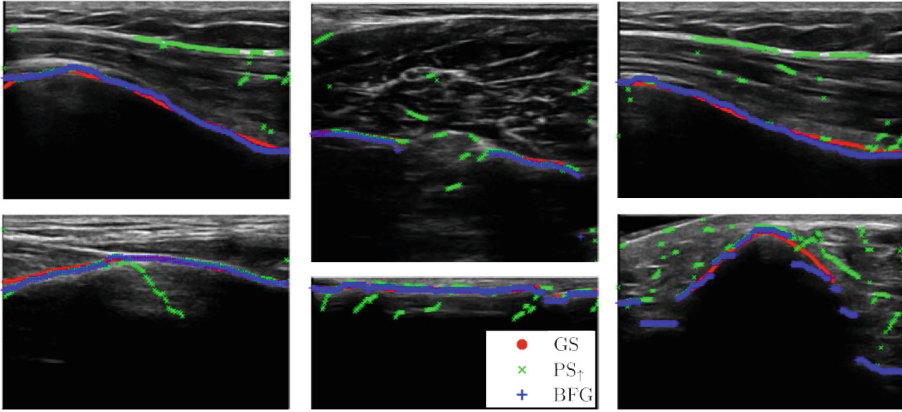
---

[1] https://www.csie.ntu.edu.tw/~cjlin/liblinear/.
[2] http://hci.iwr.uni-heidelberg.de/opengm2/.

Fig. 5. Sample qualitative results show robustness (top row) to false detections in soft tissue interfaces; (bottom, left&center) to shadowing and reverberation artifacts inside bone; (top, center) separate bone surfaces, e.g. radius and ulna; (right) images demonstrate typical failures.

To compare BFG with alternatives, we also implemented MRF with edge connectivity in Fig. 3(b.i) and Potts pairwise potentials, and conventional factor graphs (CFG) without the jump-edge potential with edge connectivity in Fig. 3(b.ii) and potentials in Table 2(a, b), and with parameters given above. As these implementations gave arbitrarily poor results for our evaluation metrics due to many false negative in comparison to BFG, we applied the following post-processing steps to improve these alternative methods to a comparable level. We first thinned the result to single pixel using morphological thinning [15]. Subsequently, if there are multiple occurrences of bone detection, only the lowermost pixel is kept to avoid false positives within the soft tissue. We denote these two post-processing steps with $(\cdot_\uparrow)$. Considering typical state-of-the-art PS methods, most require the selection of a ROI around actual bones, since multiple reflections are extracted. We compared our method with [10], where the highest PS response per column (PS_{max}) was proposed as an automatic way of identifying the bone surface. Since this yielded relatively poor results as it was, we also applied $(\cdot_\uparrow)$ to PS as an alternative technique, which we refer as state-of-the-art. We also compared with *confidence-weighted phase symmetry* (CPS) from [3]. This similarly yields many false negatives, so we report its post-processed version $CPS_\uparrow$. For BFG, simply the midpoint of l-thick **B** was output.

We used common bone-detection evaluation metrics: symmetric Hausdorff distance (sHD), one-way Hausdorff distance (oHD), and RMSE of detected bone surface to the closest gold standard (GS) point. Quantitative results averaged over 6-folds are seen in Fig. 4(a), indicating that our algorithm outperforms other approaches. We also looked at the classification accuracy of surface detections. We considered a detection pixel correct if it is within a tolerance margin around the gold standard annotation. Accordingly, we generated a classification result

(i.e. true/false positive/negative) for each column and computed accuracy score over those for the image. In Fig. 4(b), average accuracy of three best methods are seen as the tolerance is changed. The accuracy of BFG is 86 % whereas $PS_\uparrow$ is 65 % at 1 mm; 92 % vs. 72 % at 2 mm; and 95 % vs. 79 % at 4 mm margin, respectively. *BFG outperforms the others at all operating points.* According to [16], error tolerance in CAOS is 1 mm excluding operator error. Choosing this as an example operating point, we also calculated F1 scores as in Fig. 4(c). This shows the robustness of our method across all test images, compared to alternatives. A qualitative comparison between BFG, $PS_\uparrow$, and GS is seen in Fig. 5.

A computer with Intel i7 930 @ 2.80 GHz and 8 GB RAM is used for the experiments. Results were computed in 2 min on average with a non-optimized Matlab implementation, where the majority of time is taken by feature extraction; which can be in the future accelerated by parallel computation or feature selection, albeit was not the focus of this paper.

4   Conclusions

In this paper, we have presented a novel graph representation of ultrasound-bone interaction for a robust and fully-automatic segmentation of bone surfaces. Our method performs superior to alternative techniques, demonstrating clinically-relevant performance for a diverse range of anatomical regions. In the future, we will improve its speed for real-time surface detection, e.g. for registration of pre-operative models to real-time US data for navigation and guidance.

References

1. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A., Rohling, R.: Bone surface localization in ultrasound using image phase-based features. Ultrasound Med. Biol. **35**(9), 1475–1487 (2009)
2. Jain, A.K., Taylor, R.H.: Understanding bone responses in B-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework. In: Proceedings of SPIE, San Diego, USA, pp. 131–142 (2004)
3. Quader, N., Hodgson, A., Abugharbieh, R.: Confidence weighted local phase features for Robust bone surface segmentation in ultrasound. In: Linguraru, M.G., Oyarzun Laura, C., Shekhar, R., Wesarg, S., González Ballester, M.Á., Drechsler, K., Sato, Y., Erdt, M. (eds.) CLIP 2014. LNCS, vol. 8680, pp. 76–83. Springer, Heidelberg (2014). doi:10.1007/978-3-319-13909-8_10
4. Cheung, C.W.J., Law, S.Y., Zheng, Y.P.: Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): and system validation. In: EMBC, pp. 6474–6477 (2013)

5. Hacihaliloglu, I., Rasoulian, A., Rohling, R.N., Abolmaesumi, P.: Statistical shape model to 3D ultrasound registration for spine interventions using enhanced local phase features. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 361–368. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40763-5_45

6. Scepanovic, D., Kirshtein, J., Jain, A.K., Taylor, R.H.: Fast algorithm for probabilistic bone edge detection (FAPBED). In: Medical Imaging, pp. 1753–1765 (2005)

7. Hussain, M.A., Hodgson, A., Abugharbieh, R.: Robust bone detection in ultrasound using combined strain imaging and envelope signal power detection. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 356–363. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10404-1_45

8. Daanen, V., Tonetti, J., Troccaz, J.: A fully automated method for the delineation of osseous interface in ultrasound images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 549–557. Springer, Heidelberg (2004). doi:10.1007/978-3-540-30135-6_67

9. Kovesi, P.: Image features from phase congruency. J. Comput. Vis. Res. **1**(3), 1–26 (1999)

10. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A.J., Rohling, R.N.: Automatic adaptive parameterization in local phase feature-based bone segmentation in ultrasound. Ultrasound Med. Biol. **37**(10), 1689–1703 (2011)

11. Felsberg, M., Köthe, U.: GET: the connection between monogenic scale-space and Gaussian derivatives. In: Kimmel, R., Sochen, N.A., Weickert, J. (eds.) Scale-Space 2005. LNCS, vol. 3459, pp. 192–203. Springer, Heidelberg (2005). doi:10.1007/11408031_17

12. Karamalis, A., Wein, W., Klein, T., Navab, N.: Ultrasound confidence maps using random walks. Med. Image Anal. **16**(6), 1101–1112 (2012)

13. He, D.C., Wang, L.: Texture unit, texture spectrum, and texture analysis. IEEE Trans. Geosci. Remote Sens. **28**(4), 509–512 (1990)

14. Froba, B., Ernst, A.: Face detection with the modified census transform. In: Proceddings of International Conference on Automatic Face and Gesture Recognition (2004)

15. Lam, L., Lee, S.W., Suen, C.Y.: Thinning methodologies-a comprehensive survey. Pattern Anal. Mach. Intell. **14**(9), 869–885 (1992)

16. Phillips, R.: The accuracy of surgical navigation for orthopaedic surgery. Curr. Orthop. **21**(3), 180–192 (2007)