TITLE PAGE

Title

CLINICAL FEASIBILITY OF QUANTITATIVE ULTRASOUND TEXTURE ANALYSIS: A ROBUSTNESS STUDY USING FETAL LUNG ULTRASOUND IMAGES

Manuscript category

Original research

Author names and affiliation

Alvaro Perez-Moreno[1,2], MS, Mara Dominguez[1], MS, Federico Migliorelli[2,3], MD, Eduard Gratacos[2,3,4], MD, PhD, Montse Palacio[2,3,4], MD, PhD, Elisenda Bonet-Carne[1,5], PhD

[1] Transmural Biotech S.L. Barcelona, Spain.

[2] Fetal i+D Fetal Medicine Research Center, IDIBAPS, University of Barcelona, Spain.

[3] BCNatal – Barcelona Center for Maternal-Fetal and Neonatal Medicine (Hospital Clínic and Hospital Sant Joan de Deu), Barcelona, Spain

[4] Center for Biomedical Research on Rare Diseases (CIBER-ER), Barcelona, Spain

[5] UCL - Microstructure Imaging Group, Centre for Medical Image Computing (CMIC), London, United Kingdom

Short titlte

Clinical feasibility of ultrasound texture analysis

Corresponding author

Alvaro Perez-Moreno

alvaro.perez@transmuralbiotech.com

Phone +34 931 190 929

Sabino Arana, 38 1r 1a – 08028 Barcelona - Spain

ABSTRACT

Objectives. To compare the robustness of several methods based on quantitative ultrasound texture analysis in order to evaluate its feasibility for extracting features from ultrasound images on its use for diagnosis in a clinical tool.

Methods. We compared, ranked and validated the robustness of five texture-based methods for extracting textural features from ultrasound images acquired under different conditions. For comparison and ranking purposes, we used 13.171 non-ultrasound images from widely known available databases (OUTEX and PHOTEX); specifically acquired under different controlled parameters (illumination, resolution and rotation) from 103 textures. The robustness of those methods with better results using the non-ultrasound images were validated using 666 fetal lung ultrasound images acquired from singleton pregnancies. In this study, two similarity measurements (Correlation and Chebyshev distances) were used to evaluate the repeatability of the features extracted from the same tissue images.

Results. Three of the five methods presented a favorably robustness performance using the non-ultrasound database. In fact, these methods showed similarity values close to 0 for the acquisition variations and delineations. Results from ultrasound database confirmed robustness for all the evaluated methods when comparing the same texture obtained from different regions of the image (proximal/distal lungs and ultrasound machine brand stratification).

Conclusions. Our results confirmed that texture analysis can be robust (high similarity for different condition acquisitions) with potential to be included in a clinical tool.

Keywords

Quantitative ultrasound, Ultrasonography, Texture analysis, Image processing, Robustness

FULL TEXT

Introduction

Development of non-invasive and reliable methodologies to report pathophysiological process status is still an elusive goal in modern medicine. Texture analysis methods have been extensively investigated on medical images, as they possess a vast amount of texture information relevant to clinical practice.[1] This phenomenon occurs because medical images contain physical properties of tissues; the signal producing the image changes according to modifications of tissue microstructure and composition. Texture analysis methods allow quantification of these subtle changes in the image.[1]

Over the years, a large number of powerful texture-based methods have been developed thanks to improvements in computation capacity and image resolution.[2,3,4] Specifically, texture analysis in ultrasound images extracts information related to the speckle characteristics of the ultrasound image. Oosterveld et al.[5] showed the close relation between speckle and the "density" of the ultrasound scatter within a medium. In that study, Oosterveld et al.[5] suggested that ultrasound texture analysis could quantify the effective number density of tissues, as well as pathological changes of this parameter. Thus, the principle goal of applying ultrasound texture analysis is to characterize speckle variation between ultrasound images in order to distinguish those tissues altered as a consequence of the pathology.

The ability of texture-based methods for extracting relevant texture features from medical ultrasound images and quantifying subtle changes in human tissues, non-visible to the human eye, have been widely demonstrated.[6,7,8] One of the first studies based on ultrasound texture analysis[7] presented a perspective on tissue characterization features to extract diagnostic information. Later, Tunis et al.[8] corroborated that textural information in ultrasound images is related to pathophysiological processes. Thus, the potential clinical application of quantitative ultrasound texture analysis has been investigated in different medical fields .[9,10,11,12]

Sujana et al.[13] used ultrasound texture analysis and classification methods for characterizing certain liver lesions, Chen et al.[10] for classification of breast tumors and even Vince et al.[14] for characterizing coronary plaques. In the fetal-maternal field, ultrasound texture analysis was introduced to evaluate association of brain textures with neurobehavioral outcome in preterm newborns.[12]

Research in other quantitative ultrasound-based techniques reasserts a clinical trend in obtaining information related to tissue microstructure taking advantage of its acoustical properties. These techniques include elastography, flow estimation through Doppler, shear wave imaging, spectral-based parameterization of ultrasound signals and envelope statistics. [15,16] Despite some of these techniques have shown promising results for diagnosis purposes, most of them require specific devices and training for its integration into a clinical setting.[16] We introduce quantitative ultrasound texture analysis as a technique that might be easily implemented into clinical practice as it might provide reliable information from standard ultrasound.

Up to the present, most of the studies have applied texture-based methods as part of a classification system, where ultrasound texture features fed the classifier, evaluating its performance to predict the clinical outcome.[17,18] There have been few application-oriented studies aimed to evaluate the relative powers of the texture-extractor methods before any classification or retrieve system. In fact, none of them have considered whether ultrasound texture features are robust enough (i.e. repeatable regardless of different image acquisition parameters, such as illumination or resolution) to be used in a clinical setting. In particular, any have used a huge number of ultrasound images of the same tissue acquired under different conditions. It is worth to consider that speckle characteristics may be affected due to different acquisition conditions including but not restricted to those induced by operators, biological samples or ultrasound system settings. Some quantitative ultrasound-based approaches have attempted to characterize pathological tissues in a robust way[19,20,21,22] but these require

following complicated acquisition protocols to provide repeatable acquisitions conditions in order to replicate the results. Furthermore, there are new texture-based methods that have not been widely applied for characterizing ultrasound texture in the literature[23,24] even though they might be useful because they compute local textural features related to local information[25]. Finally, a fundamental step in the use of texture-based methods is the region of interest (ROI), which identifies the region of the image that corresponds to the piece of tissue that will be analyzed. Most studies overlook this step when evaluating texture analysis whereas it is a fundamental step as delineation (selection of the ROI) would be performed by different operators and, therefore, will be different each time. This might also affect the robustness of the specific textural features. For all the above, a robustness assessment to variations in the ultrasound acquisition conditions and delineations of same type of tissue would represent a step forward in the exploration of the use of quantitative ultrasound texture analysis for clinical purposes.

We aimed to compare, rank and validate the robustness of several texture-based methods in order to evaluate its feasibility as texture feature extractors in ultrasound images on its use for analysis in a clinical tool. Particularly, we compared methods that compute local information. We included those methods most commonly found in literature for ultrasound texture classification and newer methods as an alternative. To evaluate the methods, we acquired different ultrasound images of the same texture acquired under different conditions. Nevertheless, two main limitations were observed: (1) not all parameters can be modified through the whole range when scanning real textures due to clinical limitations. For instance, different ultrasound wave absorption exists when crossing distinct tissues such as fat or bone causing acoustic shadows; sometimes these artifacts cannot be avoided when the organ of interest is fixed and distant to the transducer (fetal evaluation); and (2) it is not possible to change acquisition parameters in a precise and controlled way especially due to operator variability when positioning the transducer. Thus, we decided to use an approach inspired with

131 the Image Quality Transfer (IQT) one which first selects and configures the methods using images
132 obtained with a different source but that are easier to be acquired in a controlled setting and
133 later the method is refined using real images.[26] Concretely, we used two sets of images for this
134 study: (1) a controlled set of images, non-ultrasound available images acquired under controlled
135 acquisition parameters (i.e. illumination, rotation angle) emulating the acquisition conditions of
136 medical ultrasound setting, thus evaluating a huge number of images for each texture, and (2)
137 an ultrasound image set comprising ultrasound images of fetal lungs acquired under similar
138 conditions to those of a clinical setting. Hence, (1) different texture-based methods were
139 compared and ranked using the controlled sample set and (2) the most robust methods were
140 validated using clinically acquired ultrasound images of fetal lungs.
141

Materials and methods

In this section, we briefly describe both (1) image data sets and (2) its characteristics (image acquisition and image labeling) in order to determine which information related to acquisition conditions was evaluated. We also describe (3) the ROIs to evaluate the robustness when different regions of the same tissue are delineated. Then, we introduce (4) the texture-based methods and (5) the metrics used to compare, rank and validate robustness of the methods for acquisitions and delineations. Finally, we describe (6) the experiments' design used in this study.

Data Sets

Controlled sample set. Images with different textures were obtained from widely known available databases that previously have been used for testing classification methods[27], OUTEX[28] and PHOTEX[29]. These databases provide pictures of the same texture acquired under different conditions varying (1) illumination, (2) spatial resolution and (3) rotation parameters, thus emulating the differences between ultrasound textures when acquired at different conditions in a controlled way. Three parameters whose changes might affect ultrasound speckle patterns and used to be indirectly adjusted by the radiologist when performing ultrasound scanning: (1) illumination, which is related to gain parameter or image contrast and possible attenuation of the acoustic wave that have to cross different tissues till arriving to the desired tissue to be analyzed; we also used illumination for the ultrasound system's colormaps that can be different for different systems since it is inherent to the ultrasound system; (2) spatial resolution that is related to frequency, depth, zoom and the aperture of the transducer and (3) rotation determined by the unpredictable position of the organ and the transducer when performing a scan.

Clinical ultrasound images. Fetal lung ultrasound images were acquired from singleton pregnancies attending the Maternal-Fetal Medicine Department at Hospital Clinic in Barcelona

for routine pregnancy ultrasound scans. Multiple pregnancies and structural/chromosomal anomalies were excluded from the study. Ultrasound images of the same lung tissue acquired at different conditions were not available for all patients since it was not feasible to acquire images with the whole range of acquisition parameters in a precise and controlled way. The study protocol was approved by the local Ethics Committee (ID 3823-2007) and pregnant women provided written informed consent.

Image acquisition and labeling

Each Data Set was acquired and labeled as follows:

Controlled. OUTEX and PHOTEX databases were downloaded from the links specified in Hossain et al.[27] For the purpose of this study, only those textures that could be similar to the ultrasound patterns (i.e. granulated, dotted, flecked, etc.) were selected by visual inspection. An example of the selected textures is shown in **Figure 1**. Additionally, only those images that presented similar histograms to the ones computed from the real ultrasound textures were selected (see an example in **Figure S.1** in the supplementary material). Analysis of variance (ANOVA) was computed to compare the mean, skewness and kurtosis of the histograms computed from the Controlled and Clinical data sets. All images were digitally stored in Portable Network Graphics (PNG) and Tagged Image File (TIF) formats and converted to gray scale values within a range between 0 and 255 values. Then, texture images were labeled according to controlled acquisition parameters.

A total of 69 textures were selected and labeled from OUTEX database obtaining a total of 11178 images. Specifically, OUTEX textures were labeled according to different illuminations (horizon, inca and TL84), that emulate differences in the gain and ultrasound system's colormaps used for ultrasound image representation, the resolution levels (100, 120, 300, 360, 500 and 600 dpi) and rotation degrees (0^0, 5^0, 10^0, 15^0, 30^0, 45^0, 60^0, 75^0 and 90^0) obtaining 162 images per texture. Changes in resolution and rotation degrees emulate different acquisition conditions that used to be present between ultrasound images due to frequency, depth and/or organ

position changes. Regarding the PHOTEX database, a total of 1993 images were labeled from 34 tissues selected for the purpose of this study. PHOTEX database images were labeled according to rotation degrees and tilt angles of illumination since they emulate changes in ultrasound textures due to transducer and/or organ position when insonating an organ. The acquisition parameters (rotation and tilt illumination) were controlled but differed for each texture.

Clinical. Ultrasound images of fetal lungs were acquired in an axial section of the fetal thorax at the level of the cardiac four-chamber view. Acquisition settings as gain, zoom, frequency and time-gain compensation were not fixed and were adjusted according clinical criteria. Depth and the aperture of the transducer were adjusted to magnify the fetal thorax so that the thorax occupied about two thirds of the screen. Aperture might change for each ultrasound system, operator and the unpredictable position of the fetus during acquisition. Changes in aperture and frequency are related to changes in spatial resolution (see its distribution in **Figure S.2** in the supplementary material). Scans were performed by certified radiologists using a Siemens Sonoline Antares (Siemens Medical Systems, Malvern, Pa., USA), Voluson 730 Pro, Voluson 780 Pro (GE Medical Systems, Milwaukee, Wisc., USA), ALOKA Prosound Alpha-7 (Hitachi Aloka Medical, Ltd., Tokyo, Japan) and Toshiba Aplio (Toshiba Medical Systems, Tokyo, Japan) ultrasound system. All machines were equipped with curved linear transducer with a frequency range from 3 to 7.5 MHz. All images were collected digitally in the original Digital Imaging and Communication in Medicine (DICOM) format and then inspected by EB and AP for image quality control. Images were considered non-eligible if fetal thorax occupied less than two thirds of the screen, or if color Doppler, calipers or pointers were used. Furthermore, images were excluded if they presented any of the following characteristics as they can directly alter the values of the ultrasound features: presence of obvious acoustic shadows from the fetal ribs, saturation or any type of post-processing (such as smoothing). Image quality control was done manually assisted by an ad-hoc graphical user interface (GUI) that: (1) computed the proportion of fetal thorax in the image by semi-automatically delineating

an ellipse over the thorax, (2) showed images in order to check the use of calipers, color Doppler or any type of post-processing, and (3) plotted acoustic shadows in green and saturated regions in red (pixel values close to 0 and 255, respectively).

A total of 713 ultrasound images were acquired from 385 fetuses. 47 images were discarded resulting in 666 useful images from a total of 355 patients after image quality control. Images were labeled according to rotation angle, fetal spine position (left or right) and the proximal lung (the lung close to the transducer) as left of right. The same GUI developed for image quality control was used to label the fetal lungs. By means of the GUI, a clinical expert (FM) semi-automatically calculated rotation angle indicating the orientation of the fetal spine respect to the atrio-ventricular bundle of the heart (see rotation angle distribution in **Figure S.3** in the supplementary material). Additionally, the clinical expert also indicated the fetal spine position and the proximal lung as defined above. The same GUI was used for delineation.

Image delineation

Once images were labeled, different delineations were performed in each image for each Data Set:

Controlled. An automatic delineation was performed for each texture image considering (1) 25 non-overlapped and (2) 28 overlapped but with different size ROIs. In this manner, different regions of the same texture were evaluated as it is shown in **Figures 2** and **3**, respectively.

Clinical. Two operator dependent delineations of both fetal lungs were considered, (1) manually and (2) semi-automatically ROIs, which were performed by a clinical expert (FM) (**Figure 4**). Manual delineations included the largest possible homogenous area of the fetal lung, avoiding the heart, gross vessels and surrounding areas. Semi-automatic delineations were performed indicating a size-fixed squared region, following the same criteria than for manual delineations. After the operator dependent delineations were performed, smaller ROIs were

created automatically, eroding repeatedly the manual and semi-automatic delineations (**Figure 5**) until reaching the limit of 100 pixels for the smallest ROI.

Texture-based methods

The texture-based methods used for this study are expected to be able to extract gray-scale, multi-resolution and/or rotation invariant local features from ultrasound images, as robustness for these characteristics will be required for their use in a clinical application. Additionally, the number of textural features obtained by each method should not be dependent on the ROI size or location within the same type of tissue. Textural image features were computed by several texture-based methods, widely known for texture classification in the computer vision field.[2,3,4,23,24] For each texture-based method different sets of textural features were extracted for each ROI and image. The used texture-based methods are detailed below (see a summary of the texture-based methods in **Table S.1** in the supplementary material):

Gray-Level Co-occurrence Matrices (GLCM)

GLCM has been widely used to characterize textures in ultrasound images.[30,31] This method counts pairs of horizontally adjacent pixels in a grayscale version of the image as defined by Haralick et al.[2] Characteristics of the features extracted by this method are described in detail elsewhere.[2] In our experiments, one adjacency direction 0^0 and 8 gray levels when scaling the grayscale values in the image were used to compute GLCM. Thus, there were 64 possible ordered combinations of values for each pair of pixel corresponding to the final 64 textural features.

Local Binary Patterns (LBP)

LBP has been recently applied for texture characterization in ultrasound images.[32,33] This method computes the distribution of binary patterns in the circular neighborhood of each pixel, which is characterized by a radius R and a number of neighbors P. The principle is to threshold neighboring pixels, compared to the central pixel. Thus, for each pixel a binary pattern is

obtained. A LBP code at pixel p is computed by the scalar product between the binary pattern and a vector of powers of two,

$$LBP(p) = \sum_{i=0}^{P-1} 2^i \, \delta\big(f(q_i) - f(p)\big) \qquad (1)$$

Where $f(q_i)$ and $f(p)$ are gray levels of pixels q_i and p, respectively, and δ is the Kronecker function. Then, the histogram of the LPB is used as texture features. The LPB method presents some variants that have been widely used as texture features for medical images.[34] In particular, we worked with the multi-resolution gray-scale and rotation invariant approach based on recognizing those binary patterns that occur more often in a texture image than others. These frequent patterns are called uniform patterns and are explained in more detail in Ojala et al.[4] In our study, uniform patterns were defined with P = 16 equally spaced pixels on a circle of radius R = 1 resulting in 18 specific texture features.

Histogram of Oriented Gradients (HOG)

HOG might obtain information about the anisotropy of a texture, to determine the predominant directions of a texture.[35] Recent studies have applied HOG to characterize textures in ultrasound images.[36,37] But up to the present the main purpose of applying this method on ultrasound images has been macrostructure detection such as nuchal translucency[38] or motion estimation[39]. We decided to include HOG method in our study since it may provide useful information related to tissue histology. HOG counts frequencies of gradient orientation values in localized portions of an image. The gradient orientation is estimated at every pixel and histogram is computed in order to tell how often the respective gradient direction is present in the image. The specific textural features computed by this method are explained in Junior et al.[3] For this study, each image to be analyzed (ROI) was divided in 3x3 cells (or portions) of the same size and the number of histogram bins was N_b = 9, obtaining 81 textural features.

Local Phase Quantization (LPQ)

LPQ computes quantized phase information of the Discrete Fourier Transform (DFT) but it has not been extensively applied in texture classification for medical images and, even less, for

characterizing ultrasound textures. It uses the local phase information extracted by Short Term Fourier Transform (STFT) computed over a rectangular MxM neighborhood N_p at each pixel position p of the image $f(p)$. The way of obtaining the features is explained in more detail in Ojansivu et al.[23] The same number of specific textural features is always computed, obtaining a total of 256 features for this study.

Rotation invariant LPQ (riLPQ)

The riLPQ acronym corresponds to the rotation invariant approach derived from the LPQ method. riLPQ compensates the rotation of the image that has to be analyzed considering the direction of the characteristics in the examination of the local phase. In this manner, the final textural features extracted should be the same regardless of the image rotation. For more detail, the specific features computed by this method are described in Ojansivu et al.[24] A total of 256 features are obtained by this method.

Similarity measurements / metric distances

Robustness was evaluated and validated measuring similarity (or dissimilarity) between two sets of specific textural features, extracted from two images of the same texture acquired at different conditions or acquired under the same conditions (the same image) with different ROI. We used Correlation and Chebyshev distances to compare the texture features because they provide different similarity information that might be useful in order to construct a classification algorithm when developing a clinical application.

Correlation distance measures the similarity between the relative shapes of the two features sets. This distance is defined as a measure of statistical dependence between two random sets of features. In our study, the scale of Correlation similarity values was inverted for comparison purposes. Consequently, lower distance indicated more similarity (robustness); if the features were dependent, this measure was 0. Conversely, the features were independent when this measure was 1. The Correlation distance used in this study can be expressed as

$$D_{CR}(X,Y) = 1 - \frac{\sum_{i=1}^{n}(X_i - \acute{X})(Y_i - \acute{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \acute{X})^2 \sum_{i=1}^{n}(Y_i - \acute{Y})^2}}, \quad (2)$$

where $X = \{X_0, X_1, \ldots X_{n-1}\}$ and $Y = \{Y_0, Y_1, \ldots Y_{n-1}\}$ are the features vectors extracted from images acquired under different conditions or different delineations considered statistically independent.

Chebyshev distance measures similarity between absolute values. In this study, we normalized distance between 0 and 1 for comparison purposes, in this manner two sets of features were similar (robust) if the distance was close to 0 or not (distance close to 1). This similarity measurement can be expressed as

$$D_{CH}(X,Y) = \frac{\underset{0 \le i \le N}{max}\{|X_i - Y_i|\}}{max\{D_{CH}(X,Y)\}} \quad (3)$$

where $X = \{X_0, X_1, \ldots X_{n-1}\}$ and $Y = \{Y_0, Y_1, \ldots Y_{n-1}\}$ are the features vectors extracted from images acquired at different conditions or different delineations.

Experiments

Experiments were designed following a similar approach to the IQT one.[26] First, the controlled sample set was used to determine reference values for comparison purposes when using Correlation and Chebyshev distance. Concretely, the best three methods were selected and then reference values for Correlation and Chebyshev distances were determined. Once methods were selected, we evaluated the robustness of the selected methods using the clinical sample set by comparing the results with the measures previously obtained. A summary of the experiments, including number of images for both sample sets, is displayed in **Figure 6**. The texture-based methods (GLCM, LBP, HOG LPQ and riLPQ) were ranked according to the robustness assessed (1) with the controlled sample set. Then, only those methods that presented better robustness were validated (2) with the clinically acquired ultrasound images. The experiments are explained in more detail below.

Texture-based methods ranking using the controlled sample set

For each texture and texture-analysis method the similarity measures (Correlation and Chebyshev distances) were computed using the controlled databases (OUTEX and PHOTEX). Robustness for each acquisition parameter was assessed, the parameter of interest was not fixed to any value while the rest of the acquisition parameters were fixed resulting in different acquisition scenarios. Then, both similarity measures were computed between the different textural features of the same texture acquired at different settings of the same parameter of interest. In this manner, the robustness for each acquisition parameter was isolated. This procedure was repeated for each parameter of interest till all the acquisition parameters were unfixed once. Finally, to summarize the robustness for each acquisition parameter and texture mean and standard deviation were computed over fixed parameters (different scenarios) for each similarity measurement resulting in a unique value [mean±std]. For instance, to assess illumination robustness using OUTEX database samples (illumination had 'horizon', 'inca' and 'TL84' labels), resolution and rotation were fixed resulting in a total of 54 scenarios (6 resolution levels and 9 rotation degrees) for each texture (**Figure 7**). Then, mean and standard deviation were computed for each similarity measurement over the 54 scenarios. In this example, a total of 3 similarity values [mean±std] from 2 similarity measures for 3 different labels were obtained for each texture. In order to compare robustness of the texture-based methods for each acquisition parameter, for each similarity measure the mean among similarity values was computed for each texture and then among all textures. In this manner, a unique value for each similarity measure, acquisition parameter, database (OUTEX and PHOTEX) and texture-based method was obtained.

The same approach was used to assess robustness regarding the different delineations; similarity measures were computed for the overlapped but different size ROIs and the non-overlapped ROIs delineated in the same texture image. Mean and standard deviation were computed over overlapped and non-overlapped delineations for each similarity measure resulting in a unique value for each texture image. Then, robustness for non-overlapped and

overlapped delineations was compared between the different texture-based methods computing the mean among similarity values [mean±std] for each similarity measure and each texture, and then among all selected textures. A unique similarity value was obtained for each similarity measure, the non-overlapped and overlapped delineations, each database and texture-based method.

Those texture-based methods that presented lower similarity values in regards of acquisition parameters and delineations were considered the most robust methods. Based on this criterion, methods were ranked from the most to the least robust in relation to acquisitions and delineations for each database (OUTEX and PHOTEX) first. Then, each texture-based method was globally ranked according to the number of times it ranked the best. The first three methods were elected for validation using clinical images.

Validation of the robust methods using the clinically acquired ultrasound images

Robustness of those methods that obtained better results using the controlled sample set was validated using fetal lung ultrasound images. Different experiments were performed as detailed below.

First, we assumed that left and right lungs of the same patient have the same type of tissue and in consequence images of both lungs acquired at different conditions should show the same or similar textural features. Based on this, robustness for illumination, resolution and rotation was indirectly validated by computing similarity measurements between proximal and distal lungs that were at different depth positions. Different illumination and resolution conditions of the same tissue were indirectly achieved since lateral speckle size is strongly dependent on the depth within the tissue and acoustic attenuation is dependent on depth.[5,40] Robustness for rotation was also assessed using the fetal lung ultrasound images acquired with different fetal spine orientations. In this manner, the same ultrasound tissue at different rotation conditions (with respect to proximal and distal lungs) was achieved. For each texture-

based method, mean and standard deviation of the Correlation and Chebyshev distances were computed among ultrasound fetal lung images for manual and semiautomatic delineations.

Second, in order to validate the robustness dependence of the selected texture-based methods to ultrasound systems, robustness results for illumination, resolution and rotation were stratified for the different ultrasound systems brands used in our clinical setting. No dependence to systems was considered when similar robustness was obtained between ultrasound systems of different brands. ANOVA was computed over the stratified values (Siemens, General Electrics, Toshiba and Aloka).

Finally, robustness for different delineations was assessed for each texture-based method. Similarity measurements were computed between the eroded ROIs from the manual and semiautomatic delineations. Mean and standard deviation of the similarities were computed among all the proximal and distal lungs for each method and the manual and semiautomatic delineations.

All computations in this study were performed using MATLAB R2014b (version 8.4.0.150421; MATLAB; The MathWorks Inc., Natick, Mass., USA).

Results

Selection of non-ultrasound images (Controlled data set)

No significant differences were shown between the mean, skewness and kurtosis of the histograms computed from the non-ultrasound selected images and the histograms computed from the fetal lung ultrasound textures.

Texture-based methods ranking

Similarity results will be presented in the form of mean (with standard deviation (SD) of). Similarity results between features extracted from each texture acquired at different illumination, resolution and rotation labels are given in **Table 1**. Regarding OUTEX database, most methods showed high robustness when illumination acquisition parameter was left free ('horizon', 'inca' and 'TL84'). For illumination in PHOTEX database, GLCM, LBP and riLPQ texture-based methods presented more robustness in comparison with the rest of the methods (HOG and LPQ). Specifically, HOG and LPQ method resulted in a correlation distance of 0.36 (0.15) and 0.29 (0.16 SD), respectively. GLCM, LBP and riLPQ were the most robust methods for resolution and rotation parameters stratified in OUTEX database while HOG and LPQ methods performed poorly for these parameters. HOG and LPQ methods presented less robustness for rotation in PHOTEX database than the other methods as well.

Similarity results for different delineations in the OUTEX and PHOTEX databases are displayed in **Table 2**. HOG and LPQ methods resulted the worst in terms of robustness for different delineations using both databases. Maximum similarity values between textural features extracted by HOG and LPQ in different overlapped ROIs were 0.32 (0.13) and 0.42 (0.21), respectively, and 0.27 (0.13) and 0.42 (0.30) for the non-overlapped ones. On the other hand, LBP and riLPQ performed better for the non-overlapped delineations than the other methods.

Overall, robustness performance for GLCM, LPB and riLPQ texture-based methods resulted favorably when compared with HOG and LPQ. In fact, these methods showed similarity

values close to 0 for the acquisition variations in almost all acquisition parameters and delineation from both controlled databases (OUTEX and PHOTEX). **Table 3** shows the ranking of the robustness of the texture-based methods in relation to acquisition conditions and delineations for each Data Set.

Validation of the robust methods

Table 4 displays similarity results between proximal and distal lungs of all images. Overall results confirmed robustness for all the evaluated methods (LBP, riLPQ and GLCM) depending on the similarity measure and the two operator dependent delineations (manual and semiautomatic). The highest similarity was shown for the riLPQ method using the manual delineation but overall the LBP method performed the best. The GLCM resulted in the worst method in terms of robustness when using semiautomatic delineations and measuring Correlation distance although Chebyshev distance resulted close to 0.

Stratified results by ultrasound brand are shown in **Table 4**. A total of 198, 392, 56 and 20 fetal lung ultrasound images were acquired using Siemens, General Electrics, Toshiba and Aloka ultrasound systems, respectively. Similar results were shown when comparing robustness stratified by ultrasound brands. Results demonstrated that variations in indirect illumination, resolution and rotation were not dependent on the ultrasound system. No significant differences ($p>0.05$) were found for the GLCM, LBP and riLPQ texture-based methods after stratifying by ultrasound brands.

Similarity results between textural features extracted from different ROIs are displayed in Table 5. Mean similarity values were computed among all proximal and distal lungs. Results confirmed robustness for delineations for all selected methods evaluated in the controlled setting (LBP, riLPQ and GLCM).

Discussion

This study provides evidence that texture analysis can be used to extract robust information from ultrasound images acquired under different conditions. This supports the use of texture analysis to obtain reliable features from ultrasound images, which is required to use those features for clinical purposes in a classification or grading systems.

Different quantitative ultrasound-based techniques have been explored to extract information from the signals causing speckle that are associated to the underlying tissue microstructure.[15,16] These techniques have shown promising results such as transient elastography for the staging of liver fibrosis,[41] spectral-based quantitative ultrasound parameters to characterize breast cancer and detect response of breast cancer to therapy[42,43]and most recently shear wave elasticity imaging for the assessment of cervical softening[44]. Some of these techniques are implemented on specific devices and have demonstrated to be invariant to different operators and systems.[16] Despite this, some of them have not been capable of detecting specific pathologies that still being prevalent in general population. Perhaps, because its approaches are inadequate and are not able to obtain relevant information from any tissue. Quantitative ultrasound texture analysis might become a new clinical tool that might provide new insight for clinical diagnosis.

Several attempts have been made to obtain clinical information related to a pathophysiological process using quantitative ultrasound texture analysis in a robust way. Oosterveld et al.[20] analyzed the texture of B-mode images to differentiate diffuse liver diseases and evaluated its reproducibility over a 5 days period. In that study, the B-mode images were reconstructed by radiofrequency signals that were corrected by attenuation to remove the depth. Results showed the possibility to correct the depth dependencies of the B-mode texture. Garra et al.[19] used quantitative analysis of ultrasound image texture to distinguish benign from malignant breast lesions showing promising results. Nonetheless, Garra et al.[19] concluded that the method presented ultrasound system dependence. Previous methodologies showed

promising results but not its feasibility for clinical practice. Other studies demonstrated a high diagnostic accuracy for detection of subtle changes in affected tissues non-visible for the human eye. However, no perspective studies have been conducted to validate its robustness in a clinical setting.

To our knowledge, this is the first study reporting accurate robustness of quantitative ultrasound texture analysis considering only the specific textural features and not the prediction rate for a clinical event, using machine learning algorithms. The main difference between this study and the previous ones is that robustness of ultrasound texture features was assessed using a large number of controlled (non-medical) images. The data sets used in this study emulate ultrasound acquisition conditions, which are usually present in a clinical setting. Additionally, several ROIs were performed to assess robustness when delineating. Our study shows that the LBP, riLPQ and GLCM methods were the three most robust methods for extracting information from images acquired under different conditions and different delineations in the controlled setting (**Table 1, 2** and **3**). It should be noticed that LBP and riLPQ methods were the most robust in both databases (OUTEX and PHOTEX). These methods have not been widely used for ultrasound texture classification in literature. Thus, this finding opens the possibility to explore new methods to develop ultrasound texture-based tools. Then, the most robust methods (LBP, riLPQ and GLCM) were validated using clinically acquired ultrasound images acquired by several ultrasound machines and operators. Our results validated robustness in relation to acquisition conditions using LBP, riLPQ and GLCM and showed to be invariant against ultrasound machines (**Table 4**). Concretely, LBP performed the best; the riLPQ and GLCM methods presented low similarity values in relation to acquisitions according to the delineation mode (manual or semiautomatic) and the similarity measure (Correlation and Chebyshev). Robustness against multiple delineations was also validated using clinically acquired ultrasound images. All methods resulted in low similarity values according to the delineation mode or the similarity measure

(**Table 5**). These results confirm that a texture-based tool that integrates a classification system could be developed using any of the tested methods.

Even though three of all the texture-based methods, LBP, riLPQ and GLCM showed robustness using clinically acquired ultrasound images, the use of these methods to develop a clinical tool needs to be demonstrated. Our results do not evidence the suitability of these methods to assess pathophysiological conditions involved in most of the tissues, it will depend on the intrinsic properties of textural features extracted by each texture analysis method. In fact, a method that always gives the same values will be the most robust method but completely useless. Additionally, robustness was assessed in the controlled setting over all acquisition conditions discretely and not considering specific ranges. In some cases, depending on the organ to be scanned (i.e. carotid artery or fetal heart), acquisition protocols might include repeatable acquisitions with acquisition parameters fixed within particular ranges. Therefore, the discarded texture-based methods might obtain repeatable features within specific ranges and provide useful information related to the underlying pathophysiological process. Moreover, it should be noticed that robustness was validated comparing proximal versus distal lungs. Robustness of the methods that presented higher similarities when comparing both fetal lungs would be improved using a focal configuration and evaluating tissues within the same depth. Hence, when exploring texture ultrasound analysis to develop a clinical tool, an acquisition protocol should be designed to obtain the most repeatable acquisitions.

The main strength of our study is that feasibility of texture analysis to obtain ultrasound features in a robust way was tested using non-ultrasound images acquired under controlled conditions similar to ultrasound and clinically acquired fetal lung ultrasound images. On the one hand, the non-ultrasound set provides different images of the same tissue acquired in a very precise way in contrast to whichever ultrasound setting that depends on the ability of the radiologist. This opens the possibility to evaluate a higher number of images of the same texture acquired under different conditions than in the theoretical case of evaluating real ultrasound

images. Furthermore, images were acquired combining parameters with the whole range, thus emulating possible acquisition conditions of whichever ultrasound setting where textures are scanned from any organ. On the other hand, testing ultrasound texture-based methods robustness using fetal lung ultrasound images expands opportunities to explore the same methods for quantifying textural changes in other organs, even in adult scans where acquisition conditions might be more repeatable. Another strength of our study is the use of the fetal lung ultrasound images to compare the same lung tissue at different depths (proximal and distal fetal lungs). Our results represent a forward step in relation to the study published by Thijssen[25]. Thijssen[25] suggested that texture analysis based on second order statistics should be used in the axial direction exclusively since speckle size changes according to depth and attenuation strongly. Finally, several ultrasound systems were used to acquire our clinical images. Speckle patterns might be related to system since wave propagation fundamentals, such as wavelength or gain, are post-processed in the system. In our study, we demonstrated that it is possible to configure similar settings in different ultrasound systems without affecting robustness of the selected methods (LBP, riLPQ and GLCM).

This study has some limitations that should be acknowledged. First, non-controlled resolution images in PHOTEX database might affect robustness evaluation between non-overlapped delineations. We believe that non-overlapped ROIs (of the same image) present different textural content between them when the resolution is high. For instance, the GLCM method resulted in a high dissimilarity (Correlation distance) only for non-overlapped delineations in PHOTEX database (**Table 2**) where resolution was not controlled. Second, we used clinically acquired ultrasound images of the fetal lungs to validate the robustness of the selected texture-based methods, but only robustness for different lungs (proximal versus distal) and delineations of the same tissue were assessed. In fact, for this study we assumed that proximal and distal lungs of the same patient present the same tissue without being previously demonstrated in the literature. Ideally, the robustness evaluation should be performed using

different controlled acquisitions of the same organ and patient. Although different ultrasound images of a same patient were acquired in some cases, acquisition conditions were similar since they were acquired for clinical purposes using a similar setting. To evaluate robustness for ultrasound images acquired under different conditions in a controlled way, a robustness study using different ultrasound images of the same tissue (i.e. from carotid artery or liver in adults) should be performed. Third, this study evaluated the repeatability of specific textural features obtained from images acquired under different conditions and different delineations without demonstrating its ability to detect differences against a clinical outcome of interest. We acknowledge that an additional study to compare the prediction of a clinical outcome with the same ultrasound tissue acquired at different conditions should be performed. Nonetheless, the use of texture analysis to develop a robust clinical tool has been recently demonstrated by Palacio et al.[45] In that study, a prospective multicenter study in 20 centers worldwide was undergone including a total of 730 samples for the final analysis, different operators and different ultrasound systems. The results showed that quantitative ultrasound of fetal lung texture predicted neonatal respiratory morbidity with a sensitivity, specificity, positive predictive value and negative predictive value of 74.3%, 88.6%, 51.6% and 95.5%, respectively. These promising results support our findings, suggesting that texture analysis may provide robust and relevant information useful for clinical diagnosis.

In summary, this study provides evidence that ultrasound tissues can be characterized by quantitative texture analysis in a robust way allowing its use for diagnostic purpose in clinical practice. These results should be confirmed in larger clinical images of the same tissue acquired under different controlled conditions and validated using this information to examine the ability to detect differences against a clinical outcome in a reliable manner.

598 **Supplementary material**

599 An example of the selection of the non-ultrasound images and histograms is shown in
600 **Figure S.1**.

601 The distribution of the resolution and rotation angle of the fetal lung ultrasound images
602 used for this study is displayed in **Figure S.2** and **Figure S.3**, respectively.

603 A summary table of the texture-based methods used for this study is shown in **Table S.1**.

604

ACKNOWLEDGEMENTS

This work was partially supported by grants from “la Secretaria d’Universitats I Recerca del Departament d’Economia i Coneixement de la Generalitat de Catalunya (A.P: 2014DI083)”, The Cerebra Foundation for the Brain Injured Child (Carmarthen, Wales, UK) and Obra Social “la Caixa”. Additionally, this publication has been funded with support of the Erasmus + Programme of the European Union (Framework Agreement number: 2013-0040). This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

1. Castellano G, Bonilha L, Li LM, Cendes F. Textures analysis of medical images. Clin Radiol 2004; 59:1061-9.
2. Haralick RM, Shanmugam K. Textural Features for Image Classification. IEEE Trans Syst Man, Cybern, Syst 1973; 3:610-21.
3. Junior OL, Delgado D, Gonçalves V, Nunes U. Trainable classifier-fusion schemes: An application to pedestrian detection. InIntelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on 2009 Oct 4 (pp. 1-6). IEEE.
4. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002; 24:971-87.
5. Oosterveld BJ, Thijssen JM, Verhoef WA. Texture of B-mode echograms: 3-D simulations and experiments of the effects of diffraction and scatterer density. Ultrason Imaging 1985; 7:142-60.
6. Bergen JR, Adelson EH. Theories of visual texture perception. Spat Vision 1991; 10:114-34.
7. Insana MF, Garra BS, Rosenthal SJ, Hall TJ. Quantitative ultrasonography. Med Prog Technol 1988; 15:141-153.
8. Tunis AS, Czarnota GJ, Giles A, Sherar MD, Hunt JW, Kolios MC. Monitoring structural changes in cells with high-frequency ultrasound signal statistics. Ultrasound Med Biol 2005; 31:1041-9.
9. Bonet-Carne E, Palacio M, Cobo T, et al. Quantitative ultrasound texture analysis of fetal lungs to predict neonatal respiratory morbidity. Ultrasound Obstet Gynecol 2015; 45:427-33.

637 10. Chen DR, Chang RF, Kuo WJ, Chen MC, Huang YL. Diagnosis of breast tumors with
638 sonographic texture analysis using wavelet transform and neural networks. Ultrasound
639 Med Biol 2002; 28:1301-10.
640 11. Hartman PC, Oosterveld BJ, Thijssen JM, Rosenbusch GJ. Variability of quantitative
641 echographic parameters of the liver: intra- and interindividual spread, temporal- and
642 age- related effects. Ultrasound Med Biol 1991; 17:857-67.
643 12. Tenorio V, Bonet-Carne E, Figueras F, et al. Correlation of quantitative texture analysis
644 of cranial ultrasound with later neurobehavior in preterm infants. Ultrasound Med Biol
645 2014; 40:2285-94.
646 13. Sujana H, Swarnamani S, Suresh S. Application of artificial neural networks for the
647 classification of liver lesions by image texture parameters. Ultrasound Med Biol 1996;
648 22:1177-81.
649 14. Vince DG, Dixon KJ, Cothren RM, Cornhill JF. Comparison of texture analysis methods for
650 the characterization of coronary plaques in intravascular ultrasound images. Comput
651 Med Imaging Graph 2000; 24:221-9.
652 15. Oelze ML. Emerging quantitative ultrasound applications: From diagnosing disease to
653 monitoring of therapy. The Journal of the Acoustical Society of America. 2016
654 Oct;140(4):3136-.
655 16. Oelze ML, Mamou J. Review of quantitative ultrasound: Envelope statistics and
656 backscatter coefficient imaging and contributions to diagnostic ultrasound. IEEE
657 transactions on ultrasonics, ferroelectrics, and frequency control. 2016 Feb;63(2):336-
658 51.
659 17. Christodoulou CI, Pattichis CS, Pantziaris M, Nicolaides A. Texture-based classification of
660 atherosclerotic carotid plaques. IEEE Trans Med Imaging 2003; 22:902-12.
661 18. El-Gayar MM, Soliman H. A comparative study of image low level feature extraction
662 algorithms. Egyptian informatics Journal 2013; 14:175-81.

19. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrason Imaging 1993; 15:267-85.
20. Oosterveld BJ, Thijssen JM, Hartman PC, Romijn RL, Rosenbusch GJ. Ultrasound attenuation and texture analysis of diffuse liver disease: methods and preliminary results. Phys Med Biol 1991; 36:1039.
21. Schmitz G, Ermert H, Senge T. Tissue-characterization of the prostate using radio frequency ultrasonic signals. IEEE Trans Ultrason Ferroelectr Freq Control 1999; 46:126-38.
22. Serizawa M, Maeda K. Noninvasive fetal lung maturity prediction based on gray level histrogram width. Ultrasound Med Biol 2010; 36:1998-2003.
23. Ojansivu V, Heikkilä J. Blur insensitive texture classification using local phase quantization. In International conference on image and signal processing 2008 Jul 1 (pp. 236-243). Springer Berlin Heidelberg.
24. Ojansivu V, Rahtu E, Heikkila J. Rotation invariant local phase quantization for blur insensitive texture analysis. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on 2008 Dec 8 (pp. 1-4). IEEE.
25. Thijssen JM. Ultrasonic speckle formation, analysis and processing applied to tissue characterization. Pattern Recognit Lett 2003; 24:659-75.
26. Alexander DC, Zikic D, Zhang J, Zhang H, Criminisi A. Image quality transfer via random forest regression: applications in diffusion MRI. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2014 Sep 14 (pp. 225-232). Springer, Cham.
27. Hossain S, Serikawa S. Texture databases-a comprehensive survey. Pattern Recognit Lett 2013; 34:2007-22.

688 28. Ojala T, Maenpaa T, Pietikainen M, Viertola J, Kyllonen J, Huovinen S. Outex-new
689 framework for empirical evaluation of texture analysis algorithms. In Pattern
690 Recognition, 2002. Proceedings. 16th International Conference on 2002 (Vol. 1, pp. 701-
691 706). IEEE.
692 29. Targhi AT, Geusebroek JM, Zisserman A. Texture classification with minimal training
693 images. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on 2008
694 Dec 8 (pp. 1-4). IEEE.
695 30. Giesen RJ, Huynen AL, Aarnink RG, de la Rosette JJ, Debruyne FM, Wijkstra H.
696 Construction and application of hierarchical decision tree for classification of
697 ultrasonographic prostate images. Med Biol Eng Comput 1996; 34:105-9.
698 31. Poonguzhali S, Ravindran G. Automatic classification of focal lesions in ultrasound liver
699 images using combined texture features. Information Technology Journal 2008; 7:205-
700 9.
701 32. Baños N, Perez-Moreno A, Migliorelli F, et al. Quantitative analysis of the cervical texture
702 by ultrasound and correlation with gestational age. Fetal Diagn Ther 2017; 41:265-72.
703 33. Keramidas EG, Iakovidis DK, Maroulis D, Dimitropoulos N. Thyroid texture
704 representation via noise resistant image features. In Computer-Based Medical Systems,
705 2008. CBMS'08. 21st IEEE International Symposium on 2008 Jun 17 (560-565). IEEE.
706 34. Nanni L, Lumini A, Brahnam S. Local binary patterns variants as texture descriptors for
707 medical image analysis. Artif Intell Med 2010; 49:117-25.
708 35. Beyerer J, León FP, Frese C. Machine vision: Automated visual inspection: Theory,
709 practice and applications. Springer, 2015 Oct. 1
710 36. Pazinato DV, Stein BV, de Almeida WR, et al. Pixel-level tissue classification for
711 ultrasound images. IEEE J Biomed Health Inform 2016; 20:256-67.

37. Ravishankar H, Annangi P, Washburn M, Lanning J. Automated kidney morphology measurements from ultrasound images using texture and edge analysis. In SPIE Medical Imaging 2016 Apr 1 (pp. 97901A-97901A). International Society for Optics and Photonics.
38. Deng Y, Wang Y, Chen P, Yu J. A hierarchical model for automatic nuchal translucency detection from ultrasound images. Comput Biol Med 2012; 42:706-13.
39. Tenbrinck D, Schmid S, Jiang X, Schäfers K, Stypmann J. Histogram-based optical flow for motion estimation in ultrasound imaging. J Math Imaging Vis 2013; 47:138-50.
40. Huisman HJ, Thijssen JH. An in vivo ultrasonic model of liver parenchyma. IEEE Trans Ultrason Ferroelectr Freq Control 1998; 45:739-50.
41. Friedrich–Rust M, Ong MF, Martens S, Sarrazin C, Bojunga J, Zeuzem S, Herrmann E. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology. 2008 Apr 1;134(4):960-74.
42. Sadeghi-Naini A, Papanicolau N, Falou O, Zubovits J, Dent R, Verma S, Trudeau M, Boileau JF, Spayne J, Iradji S, Sofroni E. Quantitative ultrasound evaluation of tumor cell death response in locally advanced breast cancer patients receiving chemotherapy. Clinical Cancer Research. 2013 Feb 20.
43. Sannachi L, Tadayyon H, Sadeghi-Naini A, Tran W, Gandhi S, Wright F, Oelze M, Czarnota G. Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters. Medical image analysis. 2015 Feb 1;20(1):224-36.
44. Rosado-Mendez IM, Carlson LC, Woo KM, Santoso AP, Guerrero QW, Palmeri ML, Feltovich H, Hall TJ. Quantitative assessment of cervical softening during pregnancy in the Rhesus macaque with shear wave elasticity imaging. Physics in Medicine & Biology. 2018 Apr 19;63(8):085016.

45. Palacio M, Bonet-Carne E, Cobo T, et al. Prediction of neonatal respiratory morbidity by quantitative ultrasound lung texture analysis: a multicenter study. American Journal of Obstetrics and Gynecology. 2017 Mar 23.

TABLES

Table 1. Similarity results for images acquired under different conditions in the controlled setting. Lower distance (values close to 0) indicates similarity (robustness); higher distance (values close to 1) indicates dissimilarity.

		Illumination		Resolution		Rotation	
Database	Methods	Corr	Cheb	Corr	Cheb	Corr	Cheb
OUTEX	GLCM	0.05 (0.02)	0.17 (0.03)	0.01 (0.01)	0.06 (0.03)	0.00 (0.00)	0.02 (0.01)
	LBP	0.02 (0.01)	0.07 (0.02)	0.06 (0.02)	0.11 (0.02)	0.00 (0.00)	0.02 (0.01)
	HOG	0.04 (0.03)	0.05 (0.02)	0.15 (0.06)	0.11 (0.02)	0.39 (0.11)	0.17 (0.04)
	LPQ	0.04 (0.03)	0.02 (0.01)	0.21 (0.09)	0.04 (0.01)	0.20 (0.11)	0.04 (0.02)
	riLPQ	0.01 (0.01)	0.01 (0.00)	0.07 (0.01)	0.03 (0.01)	0.01 (0.01)	0.01 (0.01)
PHOTEX	GLCM	0.01 (0.01)	0.13 (0.10)	-	-	0.03 (0.01)	0.15 (0.05)
	LBP	0.01 (0.01)	0.02 (0.01)	-	-	0.02 (0.01)	0.11 (0.02)
	HOG	0.36 (0.15)	0.23 (0.11)	-	-	0.37 (0.11)	0.36 (0.06)
	LPQ	0.29 (0.16)	0.04 (0.02)	-	-	0.43 (0.07)	0.10 (0.02)
	riLPQ	0.03 (0.02)	0.03 (0.01)	-	-	0.17 (0.05)	0.06 (0.02)

Data is given as mean (SD). Corr, Correlation distance. Cheb, Chebyshev distance. GLCM, Gray-Level Co-occurrence Matrices. LBP, Low Binary Patterns. HOG, Histogram of Oriented Gradients. LPQ, Local Phase Quantization. riLPQ, rotation invariant Local Phase Quantization.

Table 2. Similarity results for different delineations in the controlled setting. Lower distance (values close to 0) indicates similarity (robustness); higher distance (values close to 1) indicates dissimilarity.

		Overlapped		Non-overlapped	
Database	Methods	Corr	Cheb	Corr	Cheb
OUTEX	GLCM	0.02 (0.05)	0.01 (0.00)	0.05 (0.11)	0.01 (0.01)
	LBP	0.02 (0.03)	0.06 (0.04)	0.02 (0.03)	0.06 (0.05)
	HOG	0.18 (0.15)	0.32 (0.13)	0.19 (0.17)	0.27 (0.13)
	LPQ	0.25 (0.11)	0.06 (0.03)	0.17 (0.15)	0.04 (0.04)
	riLPQ	0.10 (0.04)	0.04 (0.02)	0.03 (0.04)	0.02 (0.01)
PHOTEX	GLCM	0.01 (0.04)	0.01 (0.00)	0.21 (0.30)	0.03 (0.03)
	LBP	0.01 (0.02)	0.03 (0.02)	0.02 (0.06)	0.03 (0.03)
	HOG	0.13 (0.17)	0.31 (0.15)	0.12 (0.13)	0.23 (0.13)
	LPQ	0.42 (0.21)	0.04 (0.02)	0.42 (0.30)	0.03 (0.02)
	riLPQ	0.09 (0.05)	0.05 (0.03)	0.04 (0.05)	0.04 (0.02)

Data is given as mean (SD). Corr, Correlation distance. Cheb, Chebyshev distance. GLCM, Gray-Level Co-occurrence Matrices. LBP, Low Binary Patterns. HOG, Histogram of Oriented Gradients. LPQ, Local Phase Quantization. riLPQ, rotation invariant Local Phase Quantization.

Table 3. Ranking of the texture-based methods robustness. Methods are ranked from the most (1) to the least (5) robust in relation to acquisitions and delineations for each database (OUTEX and PHOTEX).

		AcquisitionConditions		Delineations	
Methods	General ranking	OUTEX	PHOTEX	OUTEX	PHOTEX
LBP	1	2	1	2	1
riLPQ	2	1	2	3	2
GLCM	3	3	3	1	3
LPQ	4	4	4	4	5
HOG	5	5	5	5	4

GLCM, Gray-Level Co-occurrence Matrices. LBP, Low Binary Patterns. HOG, Histogram of Oriented Gradients. LPQ, Local Phase Quantization. riLPQ, rotation invariant Local Phase Quantization.

Table 4. Similarity results for proximal and distal fetal lungs in the clinical ultrasound set. Lower distance (values close to 0) indicates similarity (robustness); higher distance (values close to 1) indicates dissimilarity.

		Manual		Semiautomatic		
Methods	US brand	Corr	Cheb	Corr	Cheb	*p value*
LBP	All	**0.05 (0.04)**	**0.13 (0.10)**	**0.07 (0.06)**	**0.04 (0.04)**	0.5971
	Siemens	0.04 (0.03)	0.11 (0.09)	0.05 (0.04)	0.04 (0.04)	
	GE	0.06 (0.05)	0.15 (0.11)	0.08 (0.06)	0.04 (0.04)	
	Toshiba	0.08 (0.05)	0.07 (0.04)	0.09 (0.05)	0.02 (0.02)	
	Aloka	0.06 (0.05)	0.04 (0.02)	0.08 (0.07)	0.01 (0.01)	
riLPQ	All	**0.00 (0.00)**	**0.14 (0.11)**	**0.17 (0.23)**	**0.23 (0.18)**	0.9956
	Siemens	0.00 (0.01)	0.16 (0.12)	0.12 (0.21)	0.22 (0.16)	
	GE	0.00 (0.00)	0.13 (0.11)	0.20 (0.25)	0.24 (0.19)	
	Toshiba	0.00 (0.00)	0.11 (0.10)	0.17 (0.22)	0.23 (0.17)	
	Aloka	0.00 (0.00)	0.22 (0.14)	0.06 (0.14)	0.27 (0.22)	
GLCM	All	**0.16 (0.14)**	**0.14 (0.10)**	**0.51 (0.31)**	**0.04 (0.03)**	0.9656
	Siemens	0.12 (0.12)	0.14 (0.12)	0.36 (0.27)	0.04 (0.03)	
	GE	0.19 (0.14)	0.15 (0.10)	0.59 (0.30)	0.04 (0.03)	
	Toshiba	0.13 (0.11)	0.09 (0.04)	0.44 (0.29)	0.03 (0.01)	
	Aloka	0.19 (0.11)	0.05 (0.04)	0.48 (0.28)	0.02 (0.01)	

Data is given as mean (SD). The results for all images are in bold. Corr, Correlation distance. Cheb, Chebyshev distance. GLCM, Gray-Level Co-occurrence Matrices. LBP, Low Binary Patterns. HOG, Histogram of Oriented Gradients. LPQ, Local Phase Quantization. riLPQ, rotation invariant Local Phase Quantization.

Table 5. Similarity results for different delineations in the clinical ultrasound set. Lower distance (values close to 0) indicates similarity (robustness); higher distance (values close to 1) indicates dissimilarity.

	Manual		Semiautomatic	
Methods	Corr	Cheb	Corr	Cheb
LBP	0.08 (0.05)	0.05 (0.01)	0.23 (0.09)	0.02 (0.00)
riLPQ	0.00 (0.00)	0.15 (0.08)	0.02 (0.03)	0.23 (0.11)
GLCM	0.03 (0.04)	0.05 (0.01)	0.12 (0.09)	0.01 (0.00)

Data are given as mean (SD). Corr, Correlation distance. Cheb, Chebyshev distance. GLCM, Gray-Level Co-occurrence Matrices. LBP, Low Binary Patterns. HOG, Histogram of Oriented Gradients. LPQ, Local Phase Quantization. riLPQ, rotation invariant Local Phase Quantization.

Table S.1. Summary of the texture-based methods.

Summary of the methods					
Acronym	**Name**	**Basis**	**Parameters**	**Features Number**	**Reference**
GLCM	Gray-Level Co-occurrence Matrices	Co-occurrence Matrix	Adjacency direction = 0 Gray levels = 8	64	Haralick et al.
LBP	Local Binary Patterns	Uniform patterns	Radius (R) = 1 Number of neighbors (P) = 16	18	Ojala et al.
HOG	Histogram of Oriented Gradients	Gradient orientation values frequencies	Number of cells = 3x3 Number of histogram bins = 9	81	Junior et al.
LPQ	Local Phase Quantization	Short Term Fourier Transform	Window size = 9x9	256	Garra et al.
riLPQ	rotation invariant LPQ	Compensates image rotation considering direction of local phase characterization	Window size = 9x9 Number of angles = 36	256	Oosterveld et al.

FIGURE LEGENDS

Figure 1. Example of a selected (a) and a non-selected (b) texture image for the controlled Data Set.

Figure 2. Non-overlapped regions of interest (ROIs) divisions (dotted lines) of a texture image in the controlled setting (texture from PHOTEX database). The original image is divided into 25 different ROIs.

Figure 3. 28 Overlapped (with different sizes) ROIs of a texture image in the controlled setting (texture from PHOTEX database).

Figure 4. Manual (a) and semi-automatic (b) delineations of the proximal (1) and distal (2) lungs. Ultrasound scan of fetal lungs, 4 cardiac chamber views at 37.0 weeks+days of gestational age.

Figure 5. Eroded ROIs from manual (a) and semiautomatic (b) delineations of the fetal lungs (clinical data set). Original ROIs from a distal/proximal lung at 37.0 weeks+days of gestational age.

Figure 6. Flowchart of the experiment design.

Figure 7. Flowchart of the robustness evaluation in relation to an acquisition parameter using a texture (from OUTEX databased) acquired under different illumination conditions as example. For each similarity measurement (Chebyshev and Correlation), a mean similarity value [mean±std] in relation to illumination is obtained for texture T and each texture-based method ($z = 1…5$). Then, for each similarity measurement the mean among all textures will be computed obtaining a unique value for illumination and each method.

Figure S.1. Example of a fetal lung ultrasound texture and its histogram (a), and a selected (b) and a non-selected (c) texture image and the corresponding histograms for the controlled Data Set.

862 **Figure S.2**. Distribution of the resolution of the clinically acquired ultrasound images. Resolution
863 values are given as mm.

864 **Figure S.3**. Distribution of the rotation of the clinically acquired ultrasound images. Spine
865 orientation angle values are given as degrees.