Ultrasound Image Analysis: Challenges and Opportunities

J Alison Noble FREng

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford
alison.noble@eng.ox.ac.uk
www.ibme.ox.ac.uk
Overview

- The changing world of ultrasound imaging
- Ultrasound physics – why do ultrasound images look the way they do?
- Improving image quality
- Ultrasound image quantification
- Learning to interpret ultrasound images
- Some future directions
Ultrasound Imaging 2001 versus 2012

Poor anatomical definition but a rich source of information....
The Changing Role of Ultrasound in Clinical Medicine

- Enabling technical advances
  - Miniaturization
  - Towards (RF) open-architecture
  - Transducer quality
  - Digitization
  - 3D/4D
  - Emergence of elastography and its variants

- Role of Ultrasound (US) is changing
  - Information probe
  - Dynamic image analysis (vs static)
  - Adjunct to other imaging modalities
  - Co-registration across time
  - Quantitative tool
Ultrasound Imaging 1995 versus 2010
Modulus of the detected backscatter US signal

Variable resolution & sampling

Field-of-view

Interfaces

Attenuation

Signal dropout

Speckle

Real-time
Different quality fetal abdomen ultrasound images at various gestational ages (17, 21, 28, 33 and 38 weeks). The highest quality images (top row) show a clear boundary of the abdominal wall and the standard anatomical landmarks (stomach and umbilical vein) are clearly visible. The medium quality images (middle row) have partial occlusion of the landmarks and/or less distinguishable borders. In the lowest quality images (bottom row), there is reduced contrast due to shadowing and poor visibility of the landmarks and borders.

(With thanks to Bahbibi Rahatullah for preparing)
Improving image quality – ultrasound fusion

Aligned Single-view Images

Fused Multiview (Wavelet)

Fused Multiview (Averaging)


Clinical abstracts include at ESC, AHA 2007, ASE, EuroEcho 2009, EuroEcho 2010)
3D Fusion Echocardiography
(Scientific Highlights, ASE 2009, JACC: Cardiovascular Imaging 2010)

Contrast to noise ratio (CNR)
Cardiac Phantom

Contrast to noise ratio (CNR)
Participants

P<0.001

Endocardial border definition

Visualization grading scale
- Good
- Intermediate
- Poor
- Out of sector

Before Fusion (RT3DE) (%)
- 2.9
- 29.9
- 36.9
- 30.3

After Fusion (4DFE) (%)
- 3.9
- 0.4
- 23.6
- 72.1

Number of RT3DE full volumes used for fusion
Ultrasound Beam

(a) Aligned single view 1
(b) Aligned single view 2
(c) Fused view
(d) Segmentation from four single views over the fused image

(Yaqub, Ioannao, Papageorhiou and Noble, submitted for publication 2012, Yaqub PhD 2011)
Validation with CT

Table 5. The manually measured volume of the femur of a stillborn baby on all single views, fused view and the CT. Volumes are in mm³.

<table>
<thead>
<tr>
<th>Single 1</th>
<th>Single 2</th>
<th>Single 3</th>
<th>Single 4</th>
<th>Single 5</th>
<th>Fused</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>343</td>
<td>368</td>
<td>366</td>
<td>314</td>
<td>375</td>
<td>407</td>
<td>405</td>
</tr>
</tbody>
</table>

US fetal femur  
CT fetal femur
Ultrasound image quantification is hard ....

....for all the reasons noted before...

- Missing boundaries
- Speckle
- Intensity gradients do no characterise object boundaries which can be edge or ridge-like and do not necessarily have high contrast
- To visually interpret we make use of temporal continuity and interpolate across missing boundaries
- Can we do better than intensity gradient to represent interesting features in ultrasound images?
Robyn Owens’ analysis of the Fourier components of “features” of signals

Step up: all Fourier components have phase zero

Ridge up and down: all Fourier components have phase $\pi/2$ or $3\pi/2$

A feature is defined as a point where there is local phase congruency

Estimating local phase via the monogenic signal (Felsberg and Sommers 2001)

$$I^b(x, y) = I(x, y) * b(x, y)$$

$$h_1(x, y) * I^b(x, y)$$

$$h_2(x, y) * I^b(x, y)$$

$$E(x, y) = \sqrt{I_1^2 + (h_1 * I_1)^2 + (h_2 * I_1)^2}$$

local amplitude or energy

$$\tan \varphi(x, y) = \frac{I_1}{\sqrt{(h_1 * I_1)^2 + (h_2 * I_1)^2}}$$

local phase

$$\tan \theta(x, y) = \frac{(h_1 * I_1)}{(h_2 * I_1)}$$

local orientation
Characterise fetal nutritional state with soft tissue measurements

MALNOURISHED       OVERNOURISHED

IUGR - 2845g       Macrosomia - 4368g
39 weeks           38 weeks

- Attempt to reproduce methodology used to assess neonatal nutrition.
- Measure of fat deposition from 2D/3D ultrasound images is a good indicator of fetal nutritional state.
- Most common measurements: cheek-to-cheek, abdominal subcutaneous fat, soft tissue thigh/arm.

(Sylvia Rueda, Caroline Knight, Aris Papageorghiou, Alison Noble, U, Oxford UK)
- Region-based method adapted to segment ultrasound images.

- **Local Phase**: invariant to contrast, describes structure.

- **Fuzzy Connectedness**: considers the fuzziness of the region of interest. Assigns a global connectivity value to any pair of pixels.
Table 1. Comparison of methods with manual segmentation. (FC: Feature-based FC - C: completion - MCF: regularisation)

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>Dice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC [6]</td>
<td>87.91 ± 6.24</td>
<td>80.31 ± 5.03</td>
<td>83.57 ± 2.63</td>
</tr>
<tr>
<td>FC+MCF</td>
<td>90.30 ± 5.64</td>
<td>79.42 ± 4.93</td>
<td>84.17 ± 2.18</td>
</tr>
<tr>
<td>FC+C</td>
<td>89.08 ± 5.68</td>
<td>82.28 ± 4.65</td>
<td>85.25 ± 2.69</td>
</tr>
<tr>
<td>FC+MCF+C</td>
<td>89.83 ± 5.66</td>
<td>83.37 ± 4.63</td>
<td>86.17 ± 2.51</td>
</tr>
<tr>
<td>FC+C+MCF</td>
<td>88.75 ± 5.88</td>
<td>85.17 ± 4.56</td>
<td>86.62 ± 2.73</td>
</tr>
</tbody>
</table>

Fig. 3. Adipose tissue (mm²) vs gestational age (weeks) measured with the proposed approach applied to 20 cross-sectional images of the fetal arm, representing normal growth.
Mutual Information of Local Phase

We have seen that we can use local phase instead of intensity as the basis for feature detection. What about image registration?

Replacing the intensity-based mutual information with the local phase one (Mellor and Brady, MedIA, 2005):

\[
MI(\phi_i, \phi_j) = \sum p(\phi_i, \phi_j) \ln \left( \frac{p(\phi_i, \phi_j)}{p(\phi_i)p(\phi_j)} \right)
\]

- Detect a meaningful structural relationship between the local signal shapes of an image pair;
- Makes no assumption of a like-with-like correspondence.
MR-Echocardiography Fusion (image alignment)

Data is acquired on different machines at different times.

This leads to:

- Different image appearance
- Different temporal sampling

A non-rigid registration solution is required…..

(Zhang, Brady, Noble, MICCAI 2008, IPMI 2007, PMB 2011)
Non-rigid image alignment is based on image phase not intensity

A poly-affine transformation estimated using an iterative algorithm

(Zhang & Noble, IPMI 2007)
MR-Echocardiography Fusion

Learning to Interpret Ultrasound Images

Why use machine learning?

Sonographers interpret ultrasound images very well. Ultrasonic patterns are repeatable, but do not match to image features used in classic model-based image processing paradigms. Many clinical problems need to handle population and image acquisition variability which makes them well-suited to machine learning approaches. At least in some clinical areas there are large databases of images which is a requirement for machine learning training.
Oxford Biomedical Image Analysis Lab – Machine learning research

[Lempitsky et al 09] 3D heart segmentation

[Verhoek et al 11] 2D heart tracking

[Rahmatulah et al 11] Fetal scan plane detection

[Yaqub et al 12] Neuro-sonography feature detection

[Yaqub et al 10] Efficient 3D segmentation

[Flaccavento et al 11] Cell counting
Random Forests (Breiman 2001)

A forest consists of $T$ decision trees
A feature vector is classified by descending each tree

Random Forests is a machine learning solution toolkit
Key design choices (given a training/testing dataset) are:
The number of trees $T$
The depth of a tree
The choice of features/decision made at a node (a weak classifier)
Problem with the conventional Random Forests for 3D medical image segmentation

• In conventional training
  – Huge feature pool size \( n \) (unary3D, binary3D, rect3D, Haar3D, position3D etc)
  – Many irrelevant features
  – Approximately \( \sqrt{n} \) are selected randomly at each node\(^1\)
    e.g., for fetal femur \( n \approx 10^8 \rightarrow \) selected features \( n' = 10^4 \)
  – Features have to be selected in a more efficient manner

• In testing
  – Ideally, each tree should provide a decision based on the "strength" of the features on each path and the "number" of these features
  – Trees should have unequal votes (weighted)

\(^1\) Breiman 2001

(Yaqub PhD Thesis 2010, MICCAI MLMI 2011)
Selecting “strong” features

- Offline feature selection
- Lookup table of “good” feature
- Training a node
  - Load the lookup table of size $m$ where $m<<n$
  - Select $m'$ random keys from the lookup table
  - Choose the feature $f_k$ that has the maximum global score ($\Delta E$) from the $m'$ random ones
  - Apply the chosen feature $f_k$ to the training examples
  - Find the best threshold $th$ for $f_k$
  - Set $\text{Best\_Classifier} = (f_k, th, \Delta E)$
Segmentation of Individual Decision Trees

Segmentation accuracy by different trees. Top left is a 2D fetal femur slice. The remaining five subfigures are the segmentation results of 5 separate trees.
Weighted Testing:

• Use the scores $S$ (e.g., use the information gain for each $f_i, t_i$ calculated during training)
• Final distribution for $v_i$ is the weighted sum of distributions over all trees

$$\alpha_t = \frac{1}{F_t} \sum_{f=1}^{F_t} \frac{1}{F_j} \sum_{f=1}^{F_j} \text{Score}_f(\text{tree}_j)$$

Normalise the weight → divide by the sum of all $\text{score(tree}_{1:T})$

$$p(c \mid v_i, RF) = \sum_{t=1}^{T} \alpha_t p(c \mid v_i, \text{tree}_t)$$

Weighted vote from each tree

Score(tree$_t$) = $(s_1+s_2+s_3)/3$

1 Yaqub et al. 2010a
Results

(a) 2D ultrasound slice from the 3D volume.

(b) Segmentation from conventional RF.

(c) Segmentation from Fast RF.

(d) Segmentation from Fast Weighted RF.
Results

Linear regression analysis of 51 femoral volumes between the manually segmented (GT) and (a) the conventional RF, (b) the fast RF with even vote (FRF) and (c) the fast weighted RF (FWRF). (Yaqub et al. MICCAI-MLMI 2011).

Table 1. The mean ± standard deviation of precision, recall, dice similarity and Jaccard index over 51 cases. These measurements are computed between manually segmented fetal femur volumes and the automatically segmented ones using conventional RF, fast RF and fast weighted RF. (Yaqub et al. MICCAI-MLMI 2011)

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>Dice</th>
<th>Jaccard</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.68±0.18</td>
<td>0.89±0.11</td>
<td>0.75±0.11</td>
<td>0.61±0.13</td>
</tr>
<tr>
<td>FRF</td>
<td>0.79±0.14</td>
<td>0.81±0.11</td>
<td>0.78±0.08</td>
<td>0.65±0.10</td>
</tr>
<tr>
<td>FWRF</td>
<td>0.83±0.09</td>
<td>0.82±0.08</td>
<td>0.81±0.04</td>
<td>0.69±0.06</td>
</tr>
</tbody>
</table>
Typical segmentation results on a brain MRI. (a) shows a 2D manually segmented slice from the 3D volume. (b) shows the segmentation using the classic RF. (c) shows the segmentation using the Fast RF with equal voting. (d) shows the segmentation using the Fast weighted RF. (a-d) Black, dark gray, gray & white represent background, CSF, GM, and WM, respectively. (Yaqub et al. MICCAI-MLMI 2011).

Table 2. Comparison between the proposed and several state of the art techniques. This table is taken from [5]. Mean Jaccard indices are shown for segmenting CSF, GM and WM.

<table>
<thead>
<tr>
<th>Method</th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive MAP</td>
<td>0.069</td>
<td>0.564</td>
<td>0.567</td>
</tr>
<tr>
<td>Biased MAP</td>
<td>0.071</td>
<td>0.558</td>
<td>0.582</td>
</tr>
<tr>
<td>Fuzzy c-means</td>
<td>0.048</td>
<td>0.473</td>
<td>0.567</td>
</tr>
<tr>
<td>Maximum-a-posteriori (MAP)</td>
<td>0.071</td>
<td>0.550</td>
<td>0.534</td>
</tr>
<tr>
<td>Maximum-likelihood</td>
<td>0.062</td>
<td>0.533</td>
<td>0.551</td>
</tr>
<tr>
<td>Tree-Structure k-means</td>
<td>0.049</td>
<td>0.477</td>
<td>0.571</td>
</tr>
<tr>
<td>MPM-MAP</td>
<td>0.227</td>
<td>0.662</td>
<td>0.683</td>
</tr>
<tr>
<td>BSE/BFC/PVC</td>
<td>—</td>
<td>0.593</td>
<td>0.644</td>
</tr>
<tr>
<td>Constrained GMM</td>
<td>—</td>
<td>0.680</td>
<td>0.660</td>
</tr>
<tr>
<td>Spatial-varying GMM</td>
<td>—</td>
<td>0.765</td>
<td>0.734</td>
</tr>
<tr>
<td>Coupled surface</td>
<td>—</td>
<td>0.701</td>
<td>—</td>
</tr>
<tr>
<td>FSL [38]</td>
<td>—</td>
<td>0.756</td>
<td>—</td>
</tr>
<tr>
<td>SPM [30]</td>
<td>—</td>
<td>0.790</td>
<td>—</td>
</tr>
<tr>
<td>MAP with histograms [5]</td>
<td>0.549</td>
<td>0.814</td>
<td>0.710</td>
</tr>
<tr>
<td>Classic RF [5]</td>
<td>0.614</td>
<td>0.838</td>
<td>0.731</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>CSF</th>
<th>GM</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic RF</td>
<td>0.604±0.08</td>
<td>0.811±0.03</td>
<td>0.714±0.04</td>
</tr>
<tr>
<td>Fast RF</td>
<td>0.635±0.07</td>
<td>0.840±0.03</td>
<td>0.762±0.05</td>
</tr>
<tr>
<td>Fast-weighted RF</td>
<td>0.648±0.08</td>
<td>0.831±0.03</td>
<td>0.784±0.03</td>
</tr>
</tbody>
</table>

Table 3. Times to compute feature scores, training and testing for the classic RF, fast RF and fast-weighted RF on the brain MRI dataset. (Yaqub et al. MICCAI-MLMI 2011)

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature Selection</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0</td>
<td>78 hours</td>
<td>11 seconds</td>
</tr>
<tr>
<td>FRF</td>
<td>26 hours</td>
<td>6 hours</td>
<td>15 seconds</td>
</tr>
<tr>
<td>FWRF</td>
<td>26 hours</td>
<td>6 hours</td>
<td>23 seconds</td>
</tr>
</tbody>
</table>

Table 3. Times to compute feature scores, training and testing for the classic RF, fast RF and fast-weighted RF on the 3D fetal femur ultrasound dataset. (Yaqub et al. MICCAI-MLMI 2011)

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature Selection</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0</td>
<td>16 hours</td>
<td>1.3 second</td>
</tr>
<tr>
<td>FRF</td>
<td>11 hours</td>
<td>95 minutes</td>
<td>1.5 second</td>
</tr>
<tr>
<td>FWRF</td>
<td>11 hours</td>
<td>95 minutes</td>
<td>2.3 second</td>
</tr>
</tbody>
</table>
Learning Optical Flow Propagation Strategies using Random Forests

- Optical flow (OF) motion estimation is used to propagate single-frame segmentation results of a Random Forest classifier from one frame to the next. The best strategy for propagating between frames is learned on a per-frame basis.

(Verhoek et al, MICCAI-MLMI 2011)
Fig. 4 Area under the ROC curves for all sequences, for (a) each of the basic OF propagation strategies and the non-propagated RF segmentation result and (b) the four best and two worst of the combination strategies, as a function of the frame number. Best results: frame 1: o2+o7; frame 2: o5+o6; frame 6: o6+o8; other frames: o6+o7.
Anatomical structures are not always visible by eye due to the acquisition angle and other factors. How often can a software-based ML expert detect key anatomical structures in a 3D neurosonography scan?

(Yaqub et al. ISBI 2012)
Neuro-sonography plane finding

Table 3. 2D detection from 3D classification. Comparison between the automatic detection and the manual delineation on each 2D slice from the 3D volumes.

<table>
<thead>
<tr>
<th></th>
<th>CP</th>
<th>PVC</th>
<th>CSP</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Negative</td>
<td>733</td>
<td>784</td>
<td>813</td>
<td>830</td>
</tr>
<tr>
<td>False Positive</td>
<td>32</td>
<td>44</td>
<td>48</td>
<td>47</td>
</tr>
<tr>
<td>False Negative</td>
<td>39</td>
<td>45</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>True Positive</td>
<td>196</td>
<td>127</td>
<td>106</td>
<td>89</td>
</tr>
<tr>
<td>Precision</td>
<td>86.0%</td>
<td>74.3%</td>
<td>68.8%</td>
<td>65.4%</td>
</tr>
<tr>
<td>Recall</td>
<td>83.4%</td>
<td>73.8%</td>
<td>76.3%</td>
<td>72.4%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>92.9%</td>
<td>91.1%</td>
<td>91.9%</td>
<td>91.9%</td>
</tr>
</tbody>
</table>

(Yaquib et al. ISBI 2012)
Summary

- Ultrasound is the most widely used imaging modality in clinical practice (cardiology, oncology, early life in particular).
- As a research tool it has a role in diagnosis, interventional and potentially therapeutic research in the future.
- It can be used for 2D and 3D quantification.
- It can be used for characterising tissue health.
- Advances in medical image analysis underpin many of the current advances in 2D and 3D ultrasound.
- 3D ultrasound is already cheaper than its competitors and the cost of 3D transducers will come down.

How far are we off from 3Decho on a smartphone?
Low-cost ultrasound

MobiSante smartphone–based ultrasound system MobiUS

Interson SeeMore Ultrasound imaging probes

Cover image: Philips' compact CX50 in use in a medical clinic in Uganda. Portable ultrasound has the potential to have a major impact on maternal healthcare in developing countries. (see pages 477–489; image courtesy of Dr. Kristen DeStigter, Imaging The World).
Publications (1)

Review Articles
JA Noble, N Navab, H Becher, Review: Ultrasonic image analysis and image-guided interventions. Interface Focus, 16 May 2011 (Published online 15 June 2011)

Technical Articles
Richard V. Stebbing, John E. McManigie, J. A. Noble. Interpreting edge information for improved endocardium delineation in echocardiograms. IEEE International Symposium on Biomedical Imaging (ISBI) 2012. Accepted. WE-PO.PB.292
Publications (2)

Technical Articles (cont’d)


M Yaqub, K Javaid, C Cooper, JA Noble, Improving the classification accuracy of the classic RF method by intelligent feature selection and weighted voting of trees with application to medical image segmentation, International Workshop on Machine Learning in Medical Imaging (MLMI), Toronto, 2011.


Thank you for your attention.
Any questions?