Complete Fetal Head Compounding from Multi-View 3D Ultrasound

Robert Wright[1], Nicolas Toussaint[1], Alberto Gomez[1], Veronika Zimmer[1], Bishesh Khanal[3,1], Jacqueline Matthew[1], Emily Skelton[1], Bernhard Kainz[2], Daniel Rueckert[2], Joseph V. Hajnal[1], and Julia A. Schnabel[1]

[1] School of Biomedical Engineering & Imaging Sciences, King's College London, UK
[2] Department of Computing, Imperial College London, UK
[3] Nepal Applied Mathematics and Informatics Institute for Research (NAAMII)
robert.wright@kcl.ac.uk

Abstract. Ultrasound (US) images suffer from artefacts which limit its diagnostic value, notably acoustic shadow. Shadows are dependent on probe orientation, with each view giving a distinct, partial view of the anatomy. In this work, we fuse the partially imaged fetal head anatomy, acquired from numerous views, into a single coherent compounding of the full anatomy. Firstly, a stream of freehand 3D US images is acquired, capturing as many different views as possible. The imaged anatomy at each time-point is then independently aligned to a canonical pose using an iterative spatial transformer network (iSTN), making our approach robust to fast fetal and probe motion. Secondly, images are fused by averaging only the best (most salient) features from all images, producing a more detailed compounding. Finally, the compounding is iteratively refined using a groupwise registration approach. We evaluate our compounding approach quantitatively and qualitatively, comparing it with average compounding and individual US frames. We also evaluate our alignment accuracy using two physically attached probes, that capture separate views simultaneously, providing ground-truth. Lastly, we demonstrate the potential clinical impact of our method for assessing cranial, facial and external ear abnormalities, with automated atlas-based masking and 3D volume rendering.

1 Introduction

Ultrasound (US) technology is inexpensive and widely available, however, acquired images suffer from artefacts which limit its diagnostic value. Moreover, images acquired from different probe orientations have very different appearances, further confounding interpretation. For example, the strongest reflections are from tissue interfaces that are normal to the beam direction, therefore different tissues are more visible at different probe orientations. Additionally, big differences in acoustic impedance between tissues, like bone interfaces, create shadows that obscure the anatomy behind. Lastly, resolution is not uniform over image volumes, with lateral resolution worsening at greater depths. Thus, it takes expertise to navigate through the anatomy and recognise anomalies.

Image compounding can improve interpretability of US images by fusing information from multiple images to produce a single more informative image, better optimised for human perception. For example, compounding has been used to eliminate speckle [1], remove shadows [2, 3] and extend the field of view [3, 4]. In the beamforming literature, compounding is carried out by averaging images acquired with the same transducer from the same view. This improves image quality, but does not get rid of view-dependent artefacts, such as shadows [5]. In this work, we align images from multiple views and average only the best (most salient) image features, suppressing image regions that are blurry or in shadow. This allows the whole anatomy to be recovered in maximum detail.

Accurate image alignment is critical to image fusion quality. Poorly aligned images introduce blurriness and may even distort the correct geometric relationship between anatomical structures. Previous methods rely on careful acquisition of data, where slow, smooth sweeping motions are made with the probe. The anatomy captured in consecutive frames is assumed to have a similar pose, providing a good initialisation for registration. Images may be compounded on the fly, registering each new frame to the current fused representation [4, 6]. However, with 3D fetal imaging, this assumption is easily violated by fast fetal or probe motion, or when probe contact is lost. A single failed registration can cause misalignment of all subsequent frames, resulting in an erroneous compounding. This limits the number of frames / different views that can be successfully aligned, which is critical for our aim of reconstructing the whole fetal head. Aligning US images acquired from different view angles is particularly challenging as the heterogeneous appearances make registration via traditional intensities-based techniques ill-suited, e.g. by voxel-based similarity measures. For this reason, previous methods for multi-view fetal head compounding rely on manual registration [7], or by transforming imaging data into a view-independent domain, for example, by extracting surface models of the head [6].

The contribution of this paper is a processing pipeline to reconstruct a complete coherent 3D representation of the fetal head anatomy by compounding large numbers of 3D US images acquired from different views, each only showing a partial view of the anatomy due to shadow artefacts. This is made possible by a novel image alignment strategy, where an iterative spatial transformer network (iSTN) is used to independently reorient each frame directly to a canonical pose, making our method robust to fast fetal/probe movements, where previous methods fail. This provides a good initial compounding in a standard pose, that can be further refined using e.g. block matching. We also introduce an efficient multi-scale image fusion method which selects the best image features to fuse, producing a more detailed compounding.

2 Methods

Our approach for complete fetal head reconstruction involves four steps (see Fig. 1). **1) Multi-view freehand acquisition**: 3D US volumes are acquired from as many different probe orientations as possible. **2) Pose correction**:

acquired images are independently aligned to a canonical pose. **3) Saliency weighted compounding**: the best image features from the aligned images are fused, revealing the whole anatomy in greater detail. **4) Groupwise registration**: image alignment is improved by registration with the compounded volume. Steps 3 and 4 are repeated iteratively for a sharper compounding. Our methodology is now described in detail, in the rest of this section.

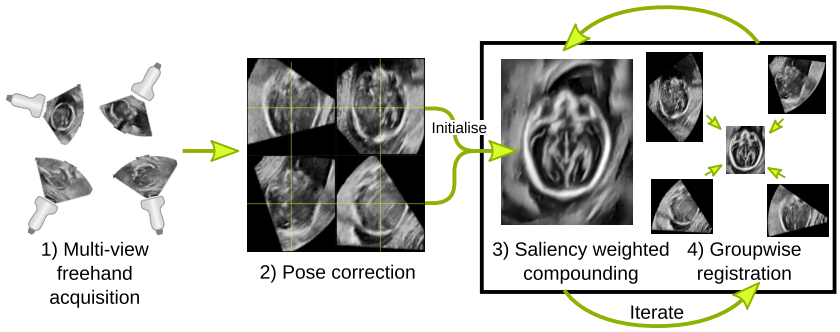

Fig. 1. Processing pipeline for complete fetal head compounding.

2.1 Pose Correction

Aligning brain images to a common atlas space is a staple of neuroimage analysis. Typically, an image is registered to a template (a mean representation of the population), using a voxel-wise image similarity measure or loss function. Many local minima exist in the cost landscape, thus a good initialisation is essential to successfully register images, which may not be available. Previously we proposed a robust alternative [8], where an iSTN is used to precisely align images of the fetal head to a canonical pose, given any initial orientation. We use this approach to reorient each frame independently, making our method robust to rapid fetal or US probe movement, unlike previous methods, which rely on a successful chain of registrations, which can easily break down in theses circumstances [4, 6]. Crucially, this allows us to align a much larger number of views/volumes, to reconstruct the complete fetal head anatomy. Furthermore, acquisition is not restricted to slow smooth sweeps; any detected head images from an examine may be used. This speeds up acquisition, removes the need for operator care, and allows multiple probes to acquire different views simultaneously.

2.2 Saliency-based image compounding

Each anatomical structure in the head is best visualised from a particular probe orientation, typically where tissues interfaces are perpendicular to the US beam.

For sub-optimal orientations, a structure may be less distinct, blurry or in shadow. Therefore, a simple average of the aligned images, results in a sub-optimal compounding, where the most detailed image features are degraded. [9] proposed to select and average only the best (salient) image features from a subset of images in order to maximise information content. Their rule-based fusion scheme, however, is computationally expensive and formulated for three images only. In this section, we propose an efficient alternative, inspired by [10], which may be used for compounding hundreds of images, capturing all views of the head. We use a Laplacian pyramid [11] multi-scale approach, which is well suited to US image compounding where resolution varies across an image volume. We also use a measure of "feature consistency" inspired by [2] for artefact suppression, making our method robust to outliers.

For n images, let $P^j = (L_0^j, L_1^j, L_2^j, I_2^j)$ be the Laplacian pyramid representation of the jth image, I^j, a tuple consisting of high L_0^j, medium L_1^j and low resolution L_2^j difference images, as well as I_2^j, a low resolution version of I^j. The average pyramid representation, $\bar{P} = (\bar{L_0}, \bar{L_1}, \bar{L_2}, \bar{I_2})$ is computed by taking a weighted average of each difference image and low resolution image independently, $\bar{L_i} = \sum_{j=1}^{n} W_i^j \circ L_i^j$ and $\bar{I_2} = \sum_{j=1}^{n} W_2^j \circ I_2^j$. Here, W_i^j is a voxel-wise weight map for the jth image pyramid, at level i, and $\circ$ denotes element-wise multiplication. Our final compounding $\bar{I}$ may then be reconstructed recursively from $\bar{P}$.

The difference images computed in a Laplacian pyramid are very similar to the result of a Laplacian of Gaussian (LoG) filter, and thus provide an edge-detection map. We take the absolute value of these difference images to give a voxel-wise measure of image "saliency" at different levels. Let x be a vectorized voxel index, then $W_i^j(x) = |L_i^j(x)| \ \forall x$. To preserve relative image intensities across the compounding we normalize the total weight for each voxel so that it sums to one, i.e. we force $\sum_{j=1}^{n} W^j(x) = 1 \ \forall x$.

Weighting images by saliency alone can result in an artefacted compounding as sharp edges are upweighted indiscriminately. Exponentiating weight maps gives a method for controlling the influence of the computed saliency in our fusion e.g. $(W(x))^\alpha \ \forall x$. Higher values of α will result in a sharper compounding but gives rise to more artefacts and reduces SNR, as only a subset of images that contain strong edges will have high weights at each voxel. As α approaches zero, the compounding will reduce to the mean of all images. Instead of a fixed value for α, for all voxels, we use a voxel-wise feature consistency measure to automatically determine the exponentiation at a voxel level, allowing artefacts to be suppressed.

For each image, feature consistency is computed at each pyramid level/ resolution independently, with respect to a saliency-weighted compounding constructed from all other images. This can be seen as a kind of leave-one-out (LOO) cross validation. To measure feature consistency we use local normalized cross correlation (LNCC). For the jth image, at pyramid level i, a feature consistency map C_i^j is computed $C_i^j = \mathrm{LNCC}(I_i^j, \hat{I}_i^j)$, where $\hat{I}_0^j$ is a compounding constructed from all other images except j, and $\hat{I}_1^j$, $\hat{I}_2^j$ are downsampled com-

poundings. As well as feature consistency for each image we also compute a mean feature consistency map, $\bar{C}_i = \frac{1}{n}\sum_{j=1}^{n} C_i^j$. This term allows greater smoothing to be applied in highly inconsistent regions across all images, such as outside of the head. Our final feature consistent, saliency-weighted fusion scheme is thus formulated:

$$\bar{L}_i = \sum_{j=1}^{n} \mathrm{pow}(W_i^j, \alpha\, C_i^j \circ \bar{C}_i) \circ L_i^j,$$
$$\bar{I}_2 = \sum_{j=1}^{n} \mathrm{pow}(W_2^j, \alpha\, C_2^j \circ \bar{C}_2) \circ I_2^j.$$

Note we retain α as a non-critical compounding parameter that may be tweaked to balance the trade off between image sharpness and SNR, if desired (we found $\alpha = 1.5$ struck a good balance).

2.3 Groupwise registration (iterative refinement)

An initial compounding is produced by fusing pose corrected frames as described above. Frame alignment is then refined iteratively with a groupwise registration approach [12], whereby all frames are rigidly registered simultaneously to the current compounding (reference), using a block-matching approach (nifty-reg)[13], before recompounding the frames. This forms a *virtuous circle* where our compounding becomes sharper due to better frame alignment, which then provides a better reference for registration, etc. This process may be repeated until no improvement is seen in image sharpness. To improve registration accuracy and robustness we mask the region of interest in which to evaluate block matches. This is easily achieved using a predefined mask in our pose corrected common-space (note, the pose correction method also estimates scale of the head). This has two major benefits: firstly, non head tissues are excluded from the matching process increasing accuracy, and secondly, the registration speed is increased by reducing the image size and the search domain for matching blocks.

3 Experiments and Results

To evaluate our approach we constructed head compoundings for 43 fetuses with an average of 212 ($\pm$78) 3D US volumes acquired for each. Selected examples are shown in Fig. 2. Note that individual frames contain shadows and speckle interference that occlude parts of the anatomy (Fig. 2, bottom row). An average compounding of multiple volumes acquired from different probe orientations reveals the whole anatomy (Fig. 2, middle row), however areas which are more frequently in shadow appear darker in the compounding, for example the back of the skull for subject 3. Our proposed compounding method successfully mitigates this problem (Fig. 2, top row). Note the enhancement of the deep grey matter in subject 1 and the uniform intensity of the skull for all subjects compared to the average compoundings. Image sharpness, estimated by the variance of the image Laplacian (Fig. 2), was also enhanced by our method compared to average compounding and individual frames.

Additionally, for 20 subjects, images were acquired from two physically attached probes, simultaneously capturing separate views of the head, 30 degrees apart. This provided ground-truth data for assessing frame alignment accuracy. For this, we estimated the transformation between the two probes, for each pair of images (1874 in total), by composing their transformations to and from the common space. We then computed, for each pair, the absolute rotation and translation errors from the mean. We computed these errors after pose correction and after each iteration of groupwise registration refinement. Fig. 3 shows that an initial median accuracy of 4.6° and 2.6mm was found after pose correction with these values falling to 3.8° and 2.3mm after 10 iterations of groupwise registration.

We demonstrate the potential clinical impact of our method for assessment of cranial, facial and external ear abnormalities through volumetric rendering of compounded image volumes in Fig. 4. Here, we automatically masked the volumes using atlas-based segmentation.

4 Discussion and Conclusions

In this work, we proposed a pipeline to fuse large numbers of 3D US images, each containing a partial view of the fetal head anatomy, into a single coherent compounding of the full anatomy. We introduced a novel image alignment strategy, using an iSTN to independently register images to a common pose, making our approach robust to fast fetal or probe movement, where previous methods fail. This allowed us to align and compound many more volumes/views than was possible previously: 212 on average, compared to 16 for previous state-of-the-art [6], ensuring the whole skull/face anatomy is captured (Fig. 4). We also introduced an efficient image compounding method which uses feature saliency to select the best image features from all frames to fuse, removing shadows and increasing the sharpness of the compoundings (Fig. 2).

We demonstrated the potential of our method for assessment of facial and cranial anomalies with automatic masking and volume rendering, something which is often time consuming and tedious for expert sonographers. On average, it takes around 10 minutes to navigate to a good view of the fetal face (see Fig. 4, bottom row) as it may be obscured by the placenta or uterus wall. In these circumstances, the sonographer may have to manually remove this tissue from a volume before rendering. The best compoundings that we observed were constructed from large numbers of images acquired from multiple probes, simultaneously, as this increased the number of views acquired and reduced the likelihood that parts of the anatomy were missing from the compounding. Deployment of our approach in a clinical setting would require real-time feedback for a sonographer (or machine) of the views acquired, to guarantee full coverage and also to minimise acquisition time. As our implementation is computationally efficient (pose correction and basic saliency-weighted compounding both run in real-time), we are currently exploring options for this, such as real-time volume rendering of the current compounding.

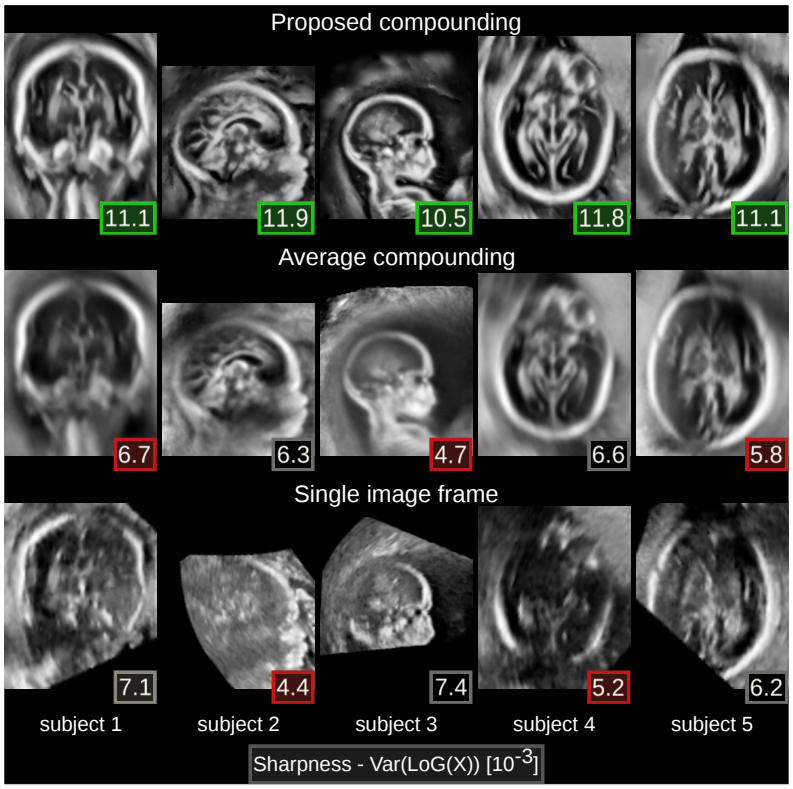

Fig. 2. Compounding examples. Results for 5 subjects are shown using our proposed method (top row). Average compoundings (middle row) and single images frames (bottom row) are shown for comparison. Image sharpness is estimated by computing the variance of the LoG filter (σ=2), on normalised images (unit variance, zero mean). Green denotes the sharpest images for each of the subjects, red the blurriest.

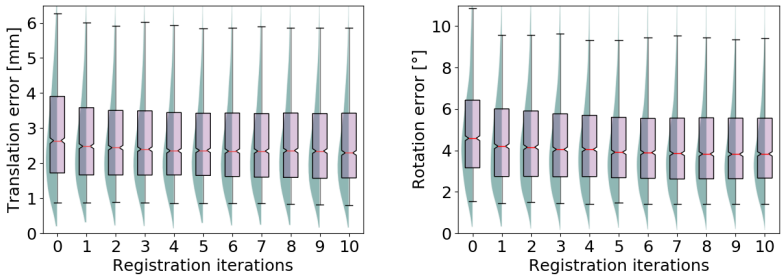

Fig. 3. Frame alignment accuracy. Rotation and translation errors are shown for each iteration of groupwise registration and also after initialisation via pose correction (iteration 0). Kernel density estimates (green) are overlaid with boxplots (purple). Note, whisker caps show 5th and 95th percentiles.

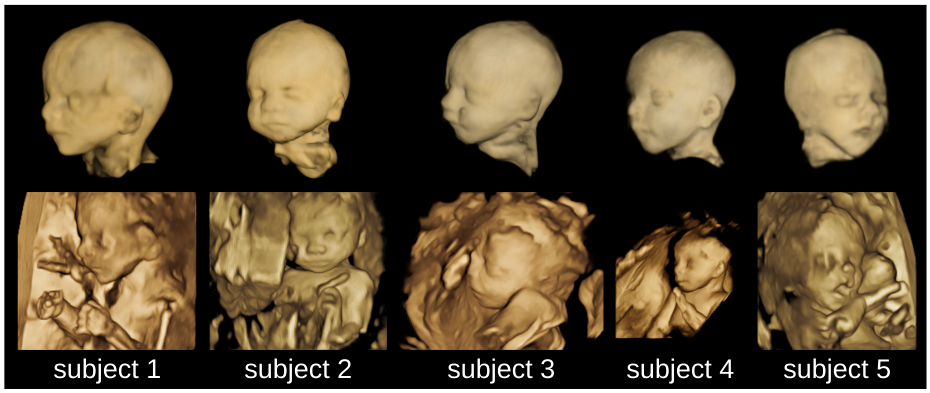

Fig. 4. Complete head reconstructions. Renderings for compounded volumes are shown for 5 subjects using our proposed method (top row). The best single view snapshot acquired by an expert sonographer during each subject's US exam is presented for comparison (bottom row).

Acknowledgements This work was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z], Wellcome Trust IEH Award [102431] and NVIDIA with the donation of a Titan Xp GPU.

References

1. Perperidis, A., Cusack, D., White, A., McDicken, N., MacGillivray, T., Anderson, T.: Temporal Compounding: A Novel Implementation and Its Impact on Quality and Diagnostic Value in Echocardiography. UMB **41**(6), 1749–1765 (2015)
2. Yao, C.: 3D Echocardiography Image Compounding. Ph.D. thesis, King's College London (2012)
3. Zimmer, V.A., Gomez, A., Noh, Y., Toussaint, N., Khanal, B., Wright, R., Peralta, L., van Poppel, M., Skelton, E., Matthew, J., Schnabel, J.A.: Multi-view Image Reconstruction: Application to Fetal Ultrasound Compounding. In: MICCAI-DATRA/PIPPI, pp. 107–116 (2018)
4. Gomez, A., Bhatia, K., Tharin, S., Housden, J., Toussaint, N., Schnabel, J.: Fast Registration of 3D Fetal Ultrasound Images Using Learned Corresponding Salient Points. In: MICCAI-FIFI. pp. 33–41 (2017)
5. Contreras Ortiz, S.H., Chiu, T., Fox, M.D.: Ultrasound image enhancement: A review. BSPC **7**(5), 419–428 (2012)
6. Khanal, B., Gomez, A., Toussaint, N., McDonagh, S., Zimmer, V., Skelton, E., Matthew, J., Grzech, D., Wright, R., Gupta, C., Hou, B., Rueckert, D., Schnabel, J., Kainz, B.: EchoFusion: Tracking and Reconstruction of Objects in 4D Freehand Ultrasound Imaging Without External Trackers. In: MICCAI-DATRA/PIPPI. pp. 117–127 (2018)
7. Perez-Gonzalez, J., Armbula-Coso, F., Guzmn, M., Camargo, L., Gutierrez, B., Mateus, D., Navab, N., Medina-Bauelos, V.: Spatial Compounding of 3-D Fetal Brain Ultrasound Using Probabilistic Maps. UMB **44**(1), 278 – 291 (2018)

8. Wright, R., Khanal, B., Gomez, A., Skelton, E., Matthew, J., Hajnal, J.V., Rueckert, D., Schnabel, J.A.: LSTM Spatial Co-transformer Networks for Registration of 3D Fetal US and MR Brain Images. In: MICCAI-DATRA/PIPPI, pp. 149–159 (2018)
9. Grau, V., Noble, J.: Adaptive Multiscale Ultrasound Compounding Using Phase Information. In: MICCAI. pp. 589–596 (2005)
10. Li, S., Kang, X., Hu, J.: Image Fusion With Guided Filtering. TIP **22**(7), 2864–2875 (2013)
11. Burt, P., Adelson, E.: The Laplacian Pyramid as a Compact Image Code. TCOM **31**(4), 532–540 (1983)
12. Vercauteren, T., Perchant, A., Pennec, X., Ayache, N.: Mosaicing of Confocal Microscopic *In Vivo* Soft Tissue Video Sequences. In: MICCAI (2005)
13. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. JMI **1**(2), 024003 (2014)