



Multi-view Image Reconstruction: Application to Fetal Ultrasound Compounding

Veronika A. Zimmer^(✉), Alberto Gomez, Yohan Noh, Nicolas Toussaint,
Bishesh Khanal, Robert Wright, Laura Peralta, Milou van Poppel,
Emily Skelton, Jacqueline Matthew, and Julia A. Schnabel

School of Biomedical Engineering and Imaging Sciences,
King's College London, London, UK
veronika.zimmer@kcl.ac.uk

Abstract. Ultrasound (US), a standard diagnostic tool to detect fetal abnormalities, is a direction dependent imaging modality, i.e. the position of the probe highly influences the appearance of the image. View-dependent artifacts such as shadows can obstruct parts of the anatomy of interest and degrade the quality and usefulness of the image. If multiple images of the same structure are acquired from different views, view-dependent artifacts can be minimized.

In this work, we propose a new US image reconstruction technique using multiple B-spline grids to enable multi-view US image compounding. The B-spline coefficients of different control point grids adapted to the geometry of the data are simultaneously optimized at every resolution level. Data points are weighted depending on their view, position and intensity. We demonstrate our method on the compounding of co-planar 2D fetal US images acquired from multiple views. Using quantitative and qualitative evaluation scores, we show that the proposed method outperforms other multi-view compounding methods.

1 Introduction

Ultrasound (US) is an imaging technique using high-frequency sound waves to visualize soft tissues and organs inside the body. US is used as a routine diagnostic tool to detect fetal abnormalities. The diagnostic value of US images is limited by the expertise of the operator and the image quality. View-dependent artifacts such as shadows can obstruct parts of the anatomy of interest and degrade the quality and usefulness of the image.

The position of the probe highly influences the appearance of the image. Focal depth is typically set such that the center of the image achieves higher quality. Some of the most degrading artifacts are acoustic shadows (Fig. 1(a)/(b)), which obscure regions of the image, and changes in pixel intensity with depth due to tissue attenuation, which cannot always be compensated for using time gain

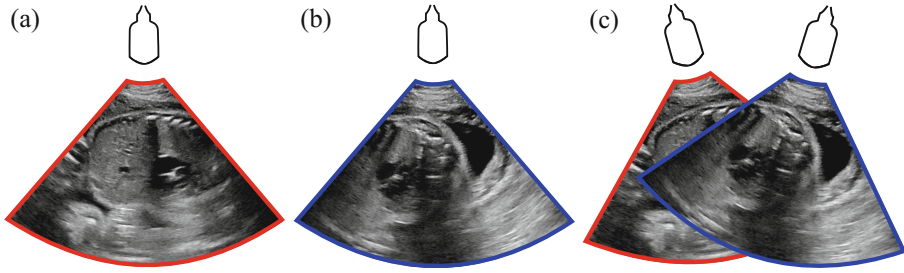


Fig. 1. (a)/(b) US images from different view directions with shadow artifacts; (c) co-planar alignment of both views, which are acquired with two active transducers.

compensation (TGC) accurately. If multiple images of the same structure are acquired from different views, view-dependent artifacts can be minimized. This can yield an easier and improved delineation of the detailed fetal anatomy by the sonographers.

Previous work has focused on compounding of multi-view 3D volumes, where there is some overlap of the fields of view (FoV) [1–3]. However, 2D imaging provides better image quality and higher frame rate and is the main imaging mode in fetal screening protocols. But obtaining a coincident imaging plane for multi-view compounding with a freehand 2D transducer is nearly impossible in practice.

In this work, we focus on the compounding of fetal 2D multi-view US images. To this end, we use a custom-made modification to a standard ultrasound system to connect two active transducers, and a physical device to maintain them on the same imaging plane, see Fig. 1(c).

To compound the multi-view images, we propose a new B-spline based [4] image reconstruction method. Due to the lack of a ground truth, different compounding methods were compared and rated qualitatively by experts, indicating a higher image quality when using multiple polar grids and a data point weighting.

Our main contributions are three-fold. First, we define multiple, view-dependent B-spline grids, adapted to the intrinsic polar geometry of US images. The US signal is measured in a polar coordinate system and only afterwards scan converted to Cartesian coordinates and interpolated for visualization. To obtain a single multi-view image, the B-spline coefficients of the grids are then determined simultaneously. Second, we introduce a data point weighting in the B-spline formulation based on the position (not only on the beam angles as in [5]) and on the intensities. And third, we evaluate our method on a dataset of 2D fetal US images acquired from multiple co-planar views.

2 Methods

2.1 Classical B-Spline Approximation

Let $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N \in \Omega \subset \mathbb{R}^2$ with $\mathbf{x}_n = (x_n, y_n)$ be a set of N image sampling points and $\mathbf{f} = f_n \in \mathbb{R}$ corresponding image intensities. The aim is to find a function $\mathcal{S}(\mathbf{x})$ such that $\mathcal{S}(\mathbf{x}_n) \approx f_n$. Using B-splines, this function can be expressed as

$$\mathcal{S}(\mathbf{x}; \mathbf{w}) = \sum_{p,q} \beta\left(\frac{x}{a} - p\right)\beta\left(\frac{y}{b} - q\right)w_{p,q},$$

where p, q are the indices of the grid control points, $w_{p,q}$ their coefficients, a, b the grid spacings along x - and y -direction with grid size $N^p \times N^q$, and $\beta(\cdot)$ is the B-spline basis function of degree d . Now, one has to find the coefficient vector $\mathbf{w}^* = (w_{p,q})$ such that

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_n \|\mathcal{S}(\mathbf{x}_n; \mathbf{w}) - f_n\|^2 + \lambda R(\mathcal{S}(\mathbf{x}; \mathbf{w})),$$

where \mathcal{R} is a regularization term and $\lambda \in \mathbb{R}$ a weighting parameter accounting for the trade-off between the reconstruction accuracy and the smoothness of the function \mathcal{S} .

For each point \mathbf{x}_n , the B-spline expansion \mathcal{S} can be expressed in matrix form as $\mathcal{S}(\mathbf{x}_n) = B_n \mathbf{w}$ with $B_n = [b_{0,0}(\mathbf{x}_n) \ b_{0,1}(\mathbf{x}_n) \ \dots \ b_{N^p, N^q}(\mathbf{x}_n)] \in \mathbb{R}^{N^p \cdot N^q}$ and $b_{p,q} = \beta(\frac{x}{a} - p)\beta(\frac{y}{b} - q)$. For all image points, this can be written as $\mathbf{f} = B\mathbf{w}$, where the n th row of $B \in \mathbb{R}^{N \times (N^p \cdot N^q)}$ is B_n , corresponding to image point \mathbf{x}_n . The coefficient vector \mathbf{w}^* is then calculated by [6]

$$\mathbf{w}^* = (B^T B + \lambda R)^{-1} B^T \mathbf{f}. \quad (1)$$

A widely used strategy, adopted in this work, is to compute the B-spline expansion on multiple resolution levels $l = 0, \dots, L$ [4]. On the coarsest level $l = 0$, the function \mathcal{S}_l is approximating the image intensities \mathbf{f} . On all subsequent levels $l > 0$, $\mathcal{S}_l(\mathbf{x}_n)$ is fitted against the residual $r_n = \mathbf{f}_n - (\sum_{i=1}^L \mathcal{S}_i(\mathbf{x}_n))$. The coefficients for each level are summed up for the final B-spline reconstruction.

2.2 Data Point Weighting Scheme

The contribution of each image point n can be weighted by a scalar $c_n \in \mathbb{R}^+$, $\sum_n c_n = N$. By arranging these weights in the diagonal of a weight matrix $C \in \mathbb{R}^{N^p \cdot N^q} \times \mathbb{R}^{N^p \cdot N^q}$, the weights can be incorporated into Eq. (1) as

$$\mathbf{w}^* = (B^T C B + \lambda R)^{-1} B^T C \mathbf{f}. \quad (2)$$

Our proposed weighting scheme is motivated by the widely used maximum compounding technique, where for the fusion of two images always the pixel value with maximum intensity is selected. Therefore, the weights in Eq. (2) are chosen such that data points with a strong signal have higher weights: $c_n = \frac{N}{\sum_i f_i} f_n$.

Additionally, we propose to take into account the position of a data point in the image. At acquisition time, image settings are optimized to get the best quality in the center, where the object of interest will be. We formulate the weight of data point \mathbf{x}_n as a function of the depth with respect to the probe position $\mathbf{b} \in \mathbb{R}^2$ and the beam angle $\alpha_n \in \mathbb{R}$:

$$g_n = g(\mathbf{x}_n, \alpha_n, \mathbf{b}) = \frac{1}{2\pi} \exp\left(-\left(\frac{\|\mathbf{x}_n - \mathbf{b}\|^2}{2\sigma_1^2} + \frac{\alpha_n}{2\sigma_2^2}\right)\right)$$

$$c_n = \frac{N}{\sum_i^N g_i f_i} g_n f_n$$
(3)

with standard deviations $\sigma_1, \sigma_2 \in \mathbb{R}$. Using the Gaussian kernel $g(\mathbf{x}_n, \alpha_n, \mathbf{b})$, a higher weight is given to data points closer to the transducer and with small beam angles. σ_1 and σ_2 were chosen to get high weights at the center of the image.

2.3 Multi-view Image Reconstruction

The matrix formulation of the B-spline approximation problem is convenient for the incorporation of multiple grids of different geometry.

Particularly, we propose to use multiple polar B-spline grids, which are adapted to the US acquisition geometry. Single polar grids have been used before for example for cardiac US registration [7]. Polar coordinates (r, θ) can be parameterized as $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{x} = (x, y)^T$: $x = r \sin(\theta)$ and $y = r \cos(\theta)$.

US images from different views do not share the same polar coordinate system. To account for this, we propose to use a separate grid for each view (as illustrated in Fig. 2(b)/(c) for two views) and optimize the coefficients of all grids simultaneously at each resolution level.

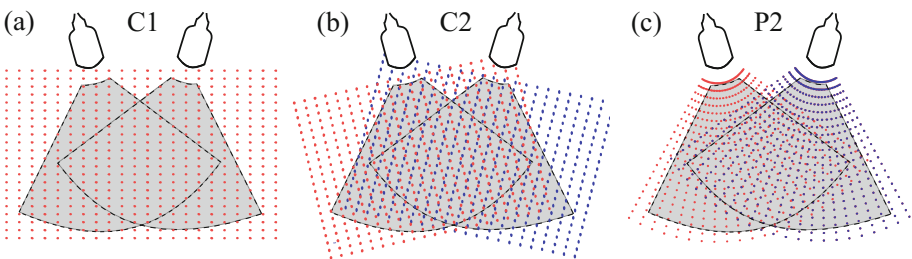


Fig. 2. Geometry of control point grids. (a) C1, single uniform (Cartesian) grid; (b) C2, two uniform (Cartesian) grids; (c) P2, two polar grids.

We consider T US views of the same object, acquired from different directions. The spatial transformations $\phi_t : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $t = 1, \dots, T$, align the T views. Those transformations can be obtained for example using image registration, tracker information or are known a priori due to special system settings.

At resolution level l , we construct T B-spline matrices $B_t, t = 1, \dots, T$, with $B_t = [b_{0,0}(\phi_t(\mathbf{x}_n)) \ b_{0,1}(\phi_t(\mathbf{x}_n)) \ \dots \ b_{N_t^p, N_t^q}(\phi_t(\mathbf{x}_n))] \in \mathbb{R}^{N_t}$. Here, $N_t = N_t^p \cdot N_t^q$ is the number of control points for view t with grid size $N_t^p \times N_t^q$. For each view, a separate coefficient vector $\mathbf{w}_t \in \mathbb{R}^{N_t}$ has to be calculated. This is done by concatenating the B_t 's to a single matrix $B \in \mathbb{R}^{N \times (N_1 + N_2 + \dots + N_T)}$ as $B = [B_1 \ B_2 \ \dots \ B_T]$.

With the regularization matrix $R \in \mathbb{R}^{(N_1 + N_2 + \dots + N_T) \times (N_1 + N_2 + \dots + N_T)}$

$$R = \begin{pmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & & \\ & & \ddots & 0 \\ 0 & & 0 & R_T \end{pmatrix},$$

Equation (1) is solved and the coefficient vectors \mathbf{w}_t are optimized simultaneously.

3 Materials and Experiments

3.1 Data Acquisition

We use a custom-made US signal multiplexer which allows to connect multiple US transducers to a standard US system, and switches rapidly between them so that images from each transducer are acquired alternatively. If the frame rate is high (as is generally in 2D mode, typically > 20 Hz), the images from both transducers are acquired nearly at the same time. We use a physical device that keeps the transducers' imaging planes co-planar and that ensures a large overlap in the center of the images to capture the region of interest from two different view angles (see Figs. 1 and 2). The relative position of the images is constant and known by calibration. If fetal motion occurred during the alternating transducer switch, images were discarded. 25 image pairs from five patients (gestational age 20–30w) were acquired using a Philips EPIQ 7g and two x6-1 transducers in 2D mode.

US images are acquired in polar coordinates. As a post-processing step, the recorded US signals are scan converted to a Cartesian coordinate system and spatially interpolated to form a 2D image. We use the scan converted but not interpolated data as input to our method to reduce interpolation artifacts.

3.2 Experiments

B-Spline Fitting Using Data Geometry. We evaluated the effect of using control point grids of different geometry for B-spline fitting of single views (Cartesian vs. polar). For a fair comparison, we ensured that the spacing of the grid points is similar in the center of the image. The grid spacing of the last and finest resolution level was 0.89×1.23 mm for the Cartesian grid and for the polar grid 0.89×0.22 mm (close to the probe), 0.89×1.01 mm (center of image) and 0.89×1.77 mm (furthest to the transducer).

Multi-view Image Compounding. We compared different multi-view B-spline reconstructions. The methods differ in the number of control point grids, T (see Sect. 2.3), the geometry of the grids and the data point weighting. We compared the following grid (compare Fig. 2) and weighting configurations:

- C1: A single uniform (Cartesian) grid of control points (Fig. 2(a)).
- C2: Two uniform (Cartesian) grids of control points transformed rigidly according to the alignment of the two views (Fig. 2(b)).
- P2: Two polar grids of control points transformed rigidly according to the alignment of the two views (Fig. 2(c)).
- W0: No data point weighting.
- W1: Data point weighting according to Eq. (3).

Accordingly, the method C1W0 denotes a B-spline fitting with a single Cartesian grid and without data point weighting. In total, six methods are compared.

3.3 Evaluation

Quantitative Evaluation. We selected four complementary quality measures to compare reconstructions I to a reference image J (available only for the first experiment): the Mean Square Error (MSE, compares the intensities of two images), the Peak Signal to Noise Ratio (PSNR, accesses the noise level of an image w.r.t. a reference image), the Structural Similarity Index (SSIM, compares structural information, such as luminance and contrast [8]), and the Variance of the Laplacian (VarL, estimates the amount of blur in an image [9]). Given two images $I, J \in \mathbb{R}^{M_1 \times M_2}$, the measures MSE, PSNR, SSIM and VarL are defined as:

$$\begin{aligned} \text{MSE}(I, J) &= \frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} (I(i, j) - J(i, j))^2, \\ \text{PSNR}(I, J) &= 10 \log_{10} \left(\frac{\max(I)}{\text{MSE}(I, J)} \right), \\ \text{SSIM}(I, J) &= \frac{(2\mu_I \mu_J + c_1)(2\sigma_{IJ} + c_2)}{(\mu_I^2 + \mu_J^2 + c_1)(\sigma_I^2 + \sigma_J^2 + c_2)}, \\ \text{VarL}(I) &= \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} (|L(i, j)| - \bar{L})^2, \end{aligned}$$

where $\mu_I, \mu_J, \sigma_I, \sigma_J, \sigma_{IJ} \in \mathbb{R}$ are the means, standard deviation and cross-covariance for images I, J , $c_1, c_2 \in \mathbb{R}$ small constants close to zero, $L \in \mathbb{R}^{M_1 \times M_2}$ the Laplacian image of I and $\bar{L} = \frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} |L(i, j)|$.

Qualitative Evaluation. No ground truth is available for the compounding of multiple views and only VarL scores can be computed. Therefore, we additionally designed a qualitative evaluation strategy. We asked seven experts (three

clinical and four US engineering experts) to evaluate as follows: at a time, two compounded images obtained by different methods from the same image pair are presented to the rater and he/she has to select which one is best, or if they have equal quality. Each rater selects from a different randomization of the six methods. The result is a quality score Q for each method, that indicates how often (in %) a method was selected as best, when it was presented to the rater as part of an image pair. No instructions were given to the experts on which features of the image to concentrate on for the quality rating. Inter-rater variability between those two groups was measured using Pearson’s r .

4 Results

4.1 B-Spline Fitting Using Data Geometry

Table 1 shows the results when reconstructing US images using the classical B-spline fitting scheme in Eq. (1) with Cartesian and polar grids. MSE, PSNR and SSIM values are computed using the original scan converted and interpolated images as reference. Using geometry-adapted (polar) grids, lower MSE and higher PSNR, SSIM and VarL values are obtained suggesting higher quality in the reconstructions compared with Cartesian grids.

Table 1. Mean square error (MSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Variance of Laplacian (VarL) of B-spline reconstructions with single Cartesian and polar grids.

	MSE	PSNR	SSIM	VarL
Cartesian	395.62 ± 143.30	22.44 ± 1.44	0.76 ± 0.03	113.35 ± 48.05
Polar	238.99 ± 162.24	25.01 ± 2.13	0.78 ± 0.05	139.24 ± 51.10

4.2 Multi-view Image Compounding

Table 2 reports the VarL values and Q -scores on the six different methods described in Sect. 3. It can be seen, that P2W1 (two view-dependent polar grids

Table 2. Evaluation of multi-view B-spline reconstructions using the Variance of Laplacian (VarL) and a qualitative Q -score obtained by the rating procedure explained in Sect. 3.2. C1: cartesian with one grid; C2: cartesian with two grids; P2: polar with two grids; W0: no weighting; W1: weighting as detailed in Eq. (3).

	C1W0	C1W1	C2W0	C2W1	P2W0	P2W1
VarL	48.6 ± 12.4	93.7 ± 17.4	53.9 ± 14.5	94.9 ± 20.4	92.0 ± 28.1	139.7 ± 33.6
Q	4.0	26.9	24.6	54.9	71.4	96.0

with data point weighting) received the highest score of $Q = 96$, i.e. the image obtained by P2W1 was chosen best in 96% of the cases. The “second best” method was P2W0 with $Q = 70.7$, further demonstrating the importance of the geometry-adapted grids to the final result. This is also reflected in the VarL values. High values, indicating sharper images, are obtained for P2W0 and P2W1.

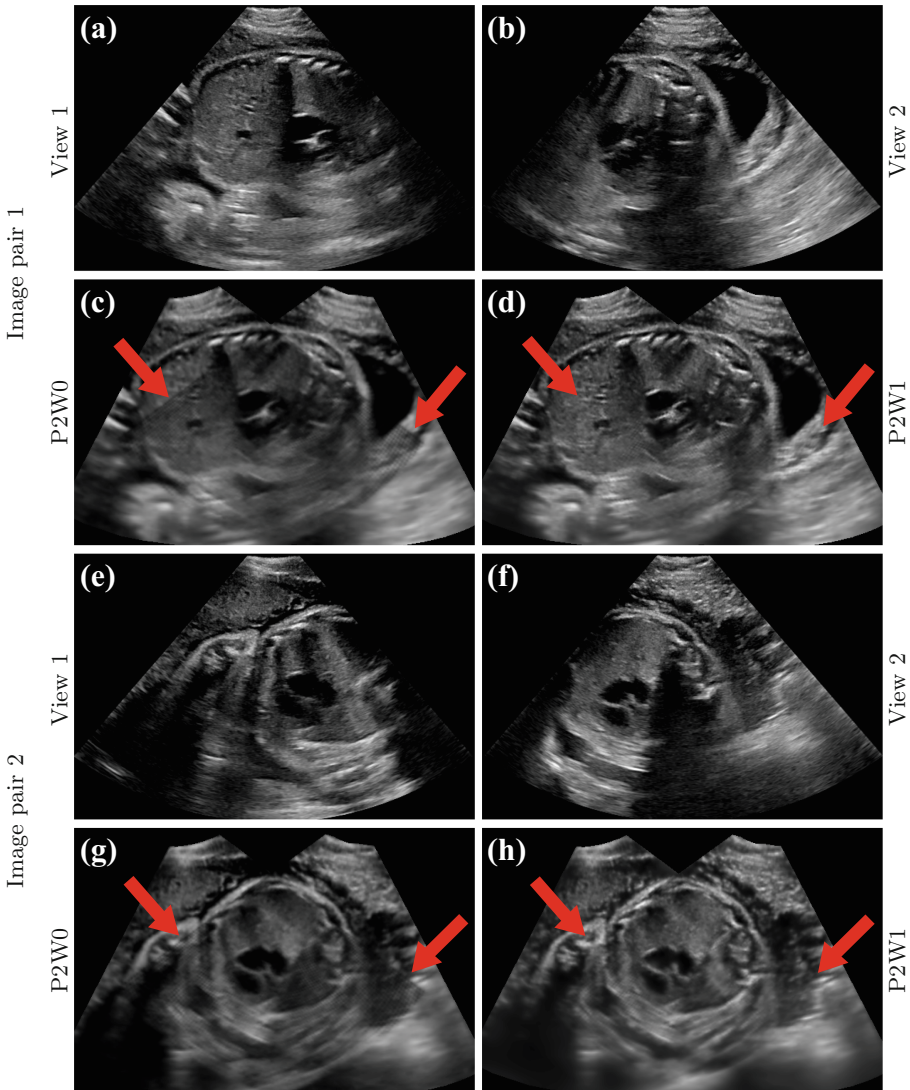


Fig. 3. (a)–(d) Original images of the two views; (e)/(g) compounded image with two polar grids, without data point weighting; (f)/(h) compounded image with two polar grids and data point weighting according to Eq. (3). (Color figure online)

For all grid configurations, the weighting improved both the ValL and Q-scores. While the best ValL values are achieved with all three grid configurations with data point weighting (C1W1: 93.7 ± 17.4 , C2W1: 94.0 ± 20.4 , P2W1: 139.7 ± 33.6), the highest Q scores are obtained with the polar grid configuration.

Overall, the inter-rater variability between all raters was low. The correlation measured with Pearson’s r is $r = 0.93$ for all experts, when comparing how often each expert selected a specific method as best. The variability when only considering the US engineers was higher ($r = 0.89$) than considering only the clinical experts ($r = 0.95$).

Two examples for the multi-view image compounding are shown in Fig. 3. By combining two views, shadow artifacts are reduced and the field-of-view is extended. By incorporating the data point weighting, artifacts due to varying intensities in both views are reduced (red arrows in Fig. 3 (e)–(h)). Those artifacts were, next to contrast and sharpness of image features, the main aspects the majority of the experts concentrated on for the quality assessment.

5 Discussion and Conclusions

We proposed a method for multi-view US image compounding, that uses multiple geometry-adapted B-spline grids that are simultaneously optimized at multiple levels. Furthermore, we introduced a data point weighting for reducing artifacts arising from different signal intensities in multiple views. Our results on co-planar US image pairs (acquired with two transducers simultaneously and held in the same plane) show that using adapted grids and our proposed weighting system yields better results qualitatively and quantitatively.

Due to the lack of a ground truth for compounded 2D US images, we designed a rating procedure evaluating the quality of the images by experts. There is some disagreement between the VarL scores and the quality rating Q score regarding the different grid and weighting configurations. This raises the question what makes out a good compounding of two US views. The sharpness or blurring, as measured by VarL, is not sufficient to rate the quality of compounding.

Motion was disregarded in our study because by using a rigid physical device, we can ensure that the images are co-planar and the transformation for aligning them is known a priori. However, fetal motion can occur in the small time gap between image acquisition from two transducers. For future work, we plan to incorporate a registration step in our framework to correct for fetal motion.

It is straightforward to generalize our framework to 3D. However, in the real-time 3D mode the frame rate decreases significantly and the assumption of no motion between the two transducer acquisitions does not hold anymore. A registration step becomes inevitable.

The proposed method is not restricted to B-splines for interpolation, and other gridded functions such as Gaussian functions are also possible. The ability to perform multi-view image reconstruction opens several possibilities, for example further reduction of acoustic shadows or other artifacts, or the inclusion of the orientation as additional dimension for image representation [2].

Acknowledgements. This work was supported by the Wellcome Trust IEH Award [102431]. This work was also supported by the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z]. The research was also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Yao, C., Simpson, J.M., Schaeffter, T., Penney, G.P.: Multi-view 3D echocardiography compounding based on feature consistency. *Phys. Med. Biol.* **56**(18), 6109–6128 (2011)
2. Hennersperger, C., Baust, M., Mateus, D., Navab, N.: Computational sonography. In: *Proceedings of MICCAI*, pp. 459–466 (2015)
3. Banerjee, J., et al.: A log-Euclidean and total variation based variational framework for computational sonography. In: *Proceedings of SPIE Medical Imaging*, vol. 10574 (2018)
4. Lee, S., Wolberg, G., Shin, S.Y.: Scattered data interpolation with multilevel B-splines. *IEEE Trans. Vis. Comp. Graph.* **3**(3), 228–244 (1997)
5. Ye, X., Noble, J.A., Atkinson, D.: 3-D freehand echocardiography for automatic left ventricle reconstruction and analysis based on multiple acoustic windows. *IEEE Trans. Med. Imag.* **21**(9), 1051–1058 (2002)
6. Arigovindan, M., Suhling, M., Hunziker, P., Unser, M.: Variational image reconstruction from arbitrarily spaced samples: a fast multiresolution spline solution. *IEEE Trans. Imag. Proc.* **14**(4), 450–460 (2005)
7. Porras, A.R., et al.: Improved myocardial motion estimation combining tissue Doppler and B-mode echocardiographic images. *IEEE Trans. Med. Imag.* **33**(11), 2098–2106 (2014)
8. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Imag. Proc.* **13**(4), 600–612 (2004)
9. Pech-Pacheco, J.L., Cristobal, G., Chamorro-Martinez, J., Fernandez-Valdivia, J.: Diatom autofocusing in brightfield microscopy: a comparative study. *Proc. ICPR* **3**, 314–317 (2000)