

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Spatiotemporal registration of multiple three-dimensional echocardiographic recordings for enhanced field of view imaging

Adriyana Danudibroto
Jørn Bersvendsen
Olivier Gérard
Oana Mirea
Jan D'hooge
Eigil Samset

SPIE.

Adriyana Danudibroto, Jørn Bersvendsen, Olivier Gérard, Oana Mirea, Jan D'hooge, Eigil Samset, "Spatiotemporal registration of multiple three-dimensional echocardiographic recordings for enhanced field of view imaging," *J. Med. Imag.* **3**(3), 037001 (2016), doi: 10.1117/1.JMI.3.3.037001.

Spatiotemporal registration of multiple three-dimensional echocardiographic recordings for enhanced field of view imaging

Adriyana Danudibroto,^{a,b,*} Jørn Bersvendsen,^{a,c} Olivier Gérard,^a Oana Mirea,^b Jan D'hooge,^b and Eigil Samset^b

^aGE Vingmed Ultrasound, Gaustadalléen 21, Oslo 0349, Norway

^bKU Leuven, Department of Cardiovascular Sciences, Cardiovascular Imaging and Dynamics Lab, UZ Herestraat 49, Box 7003, Leuven 3000, Belgium

^cUniversity of Oslo, Department of Informatics, Gaustadalléen 23 B, Oslo 0373, Norway

Abstract. The use of three-dimensional (3-D) echocardiography is limited by signal dropouts and narrow field of view. Data compounding is proposed as a solution to overcome these limitations by combining multiple 3-D recordings to form a wide field of view. The first step of the solution requires registration between the recordings both in the spatial and temporal dimension for dynamic organs such as the heart. Accurate registration between the individual echo recordings is crucial for the quality of compounded volumes. A temporal registration method based on a piecewise one-dimensional cubic B-spline in combination with multiscale iterative Farneback optic flow method for spatial registration was described. The temporal registration method was validated on *in vivo* data sets with annotated timing of mitral valve opening. The spatial registration method was validated using *in vivo* data and compared to registration with Procrustes analysis using manual contouring as a benchmark. The spatial accuracy was assessed in terms of mean of absolute distance and Hausdorff distance between the left ventricular contours. The results showed that the temporal registration accuracy is in the range of half the time resolution of the echo recordings and the achieved spatial accuracy of the proposed method is comparable to manual registration. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.3.3.037001]

Keywords: rigid registration; three-dimensional Farneback; spatial alignment; temporal alignment; three-dimensional echocardiography; multiscale iterative optic flow.

Paper 16026RR received Feb. 12, 2016; accepted for publication Jun. 20, 2016; published online Jul. 8, 2016.

1 Introduction

Three-dimensional (3-D) echocardiography (echo) has brought added value to ultrasound imaging in clinical practice. In comparison to its two-dimensional counterpart, 3-D echo can provide a more vivid view of complex 3-D anatomical structures as well as the spatial relation between structures. It assists in the diagnostic process and helps in training new echocardiographers.¹ Its real-time capability makes 3-D echo an attractive imaging modality for intraoperative monitoring.¹

Despite the added value of 3-D echo, it also shows some inherent limitations. First, signal dropout is quite common, particularly in the presence of strongly reflecting structures. Second, image appearance depends on the angle of incidence of the ultrasound beam on the region of interest.² Finally, while imaging at acceptable spatiotemporal resolution, the 3-D field of view remains limited. The latter is particularly true for intraoperative imaging devices such as transesophageal (TEE) and intracardiac echocardiography (ICE) due to their small aperture size.³ Having a wide 3-D field of view can be of clinical relevance for interventional guidance as well as for diagnostic purposes. For example, left-right heart interactions might more easily be studied, which could be of diagnostic value in pulmonary hypertensive patients or patients with tetralogy of Fallot.

In an attempt to overcome the aforementioned limitations, several postprocessing methodologies have been proposed.

These approaches have typically focused on compounding several acquisitions, which involves both registration^{3–6} and fusion.^{7–10} The former is performed to bring all recordings in a common coordinate system while the latter combines the excess of data in the overlapping area for optimal visualization. In this article, the focus is set on the registration problem.

The registration of echocardiographic recordings requires registration in the spatial as well as temporal dimension. For temporal registration, several methods have been developed. All methodologies used a temporal characteristic function that can uniquely represent the cardiac cycle, e.g., the ECG. As an alternative to the ECG, several methodologies used a temporal characteristic function that is based on a temporal normalized cross correlation (NCC), which is the NCC function between the first frame with the rest of the frames in the sequence.^{11,12} For the latter, the method relied solely on the image content without the need of any error prone external measurements. The temporal transformation was then computed by optimizing a cost function between the temporal characteristic sequences. Perperidis et al.¹³ and Zhang et al.¹¹ used a one-dimensional (1-D) cubic B-spline to perform the temporal transformation based on the temporal NCC sequences from cardiac MR and 3-D echo sequences. Alternatively, Perissinotto et al.¹² used a dynamic time warping method that transforms the sequences through computing the path of least resistance.

*Address all correspondence to: Adriyana Danudibroto, E-mail: adriyana.danudibroto@kuleuven.be

Several approaches toward spatial registration have been proposed. On one hand, approaches using electromagnetic or optical tracking devices have been studied.^{6,9} However, for electromagnetic tracking, there is a risk of disturbance in the electromagnetic field when ferromagnetic material is present and in the case of using an optical tracker, a clear line of sight from the optical marker to the sensor has to be kept. On the other hand, several image-based registration methods have been proposed. These can be grouped into voxel-based and feature-based methods. Most of the voxel-based registration methods rely on optimization of an objective function that is computed directly from the voxel intensity values, e.g., NCC and mutual information.^{7,8,14,15} Alternatively, feature-based methods extract features from the original images providing a framework that limits the amount of data processing by dealing only with representative informations from the image. Few examples can be found in the work of Ni et. al.⁴ and Schneider et. al.⁵

In this article, a temporal registration method based on a 1-D cubic B-spline^{11,13} was adapted and applied prior to the spatial registration in order to ensure registration of corresponding frames in the sequence, which thus justifies the rigid transformation applied in the spatial registration, as no physiological deformation was expected between the registered frames. The image-based temporal registration was chosen because it utilizes the information contained in the image itself without requiring additional measurements. In addition, a spatial registration method that is based on image decomposition into a polynomial basis using Farneback optic flow^{16,17} is proposed as it offers several advantages. First, it is robust to variations in the image in comparison to other classical optic flow methods and in comparison to the classic Lukas Kanade optic flow, the Farneback method achieves higher accuracy.^{18,19} Second, it attracts the floating image to the optimum position and orientation that matches the reference image based on the analytical solution of the optic flow problem, in contrast to the voxel-based methods that optimize a global cost function by applying an extensive series of transformations. Third, given reasonable overlap between volumes, it is independent to pose initialization. Finally, in comparison to the aforementioned feature-based

methods, it does not rely on feature detection. The method was implemented in an iterative multiscale framework, allowing for more efficient registration of images with very small frame-to-frame transformations as well as larger transformations.

This article is an extension of our previous report²⁰ in which the multiscale iterative registration method based on Farneback optic flow was introduced and its accuracy validated using a set of simulated ultrasound volumes²¹ and estimated using an *in vivo* data set. The addition presented in this article consists of (1) proposing an automatic method for determining the number of levels and iterations in the multiscale iterative scheme, (2) validating the spatial registration accuracy quantitatively using *in vivo* recordings, (3) including temporal registration in the proposed frame work, and (4) validating the accuracy of the temporal registration quantitatively using *in vivo* recordings.

2 Methodology

The aim of the temporal registration is to map each event in the cardiac cycle of the reference sequence $f_R(\mathbf{x}, t_r)$ to corresponding event in the floating sequence $f_F(\mathbf{x}, t_f)$ [Fig. 1(a)]. On the other hand, for spatial registration, the aim is to compute the optimum transformation that can map each point in the floating volume $f_F(\mathbf{x})$ to a point in the reference volume $f_R(\mathbf{x})$. In the proposed method, this was obtained through three steps: (1) preprocessing, (2) motion field generation using 3-D Farneback optic flow, and (3) rigid transform regularization with Procrustes analysis. This process was run iteratively in a multiscale scheme [Fig. 1(b)].

2.1 Temporal Registration

2.1.1 Temporal characteristic curve

The temporal registration was performed on a temporal characteristic curve that compactly represents the events in the cardiac cycle. The chosen characteristic curve was computed using temporal NCC function that has previously been shown to have a shape closely matching the left ventricular (LV) volume curve.^{11,12} The temporal NCC function was computed as follows:

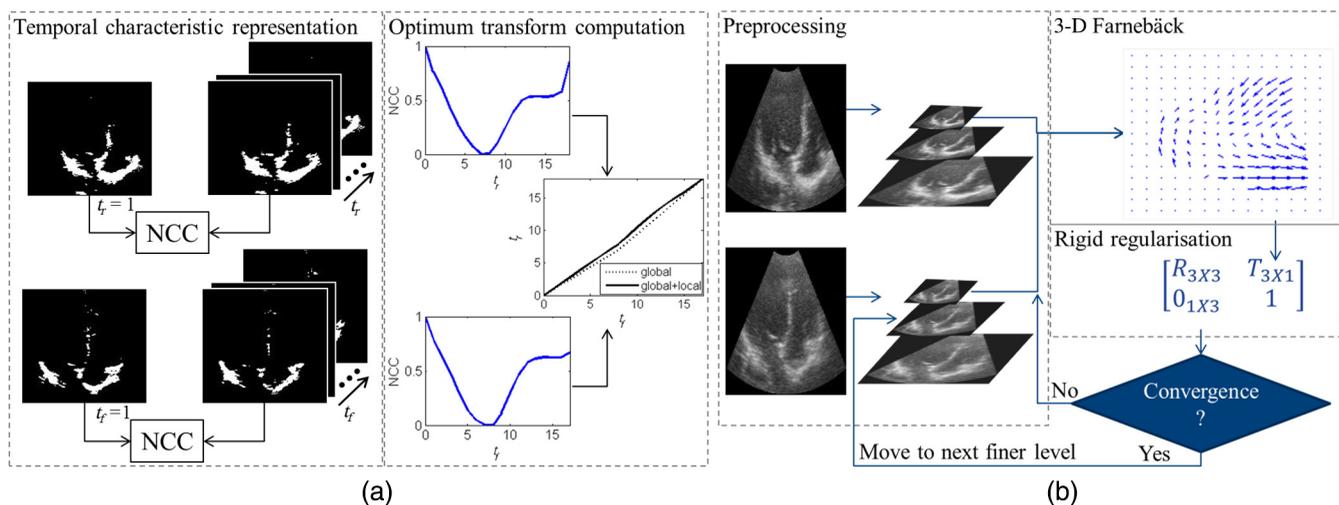


Fig. 1 Schematic overview of the proposed (a) temporal and (b) spatial registration method.

$$r(t) = \frac{\sum_x^{M_x} \sum_y^{M_y} \sum_z^{M_z} [f(\mathbf{x}, t) - \bar{f}(t)][f(\mathbf{x}, 0) - \bar{f}(0)]}{\sqrt{\sum_x^{M_x} \sum_y^{M_y} \sum_z^{M_z} [f(\mathbf{x}, t) - \bar{f}(t)]^2} \sqrt{\sum_x^{M_x} \sum_y^{M_y} \sum_z^{M_z} [f(\mathbf{x}, 0) - \bar{f}(0)]^2}}, \quad (1)$$

where $\mathbf{x} = [x, y, z]$, \bar{f} is the mean intensity of the volume at a given time, and M_x, M_y, M_z are the volume dimensions in x, y, z directions, respectively. For speeding up the computation, the volume was binarized using a thresholding value set to half of the intensity range. The resulting curve was scaled to the range of 0 to 1 in order to put both characteristic curves on the same scale.

2.1.2 Optimum temporal transformation computation

The proposed temporal transformation was composed of a global and a local term that were both divided in a piecewise function for the systolic and the diastolic phases. The transformation was thus modeled as follows:

$$T(t) = T_{\text{global_systolic}}(t) + T_{\text{local_systolic}}(t), \quad t \in [0, ES]$$

$$T_{\text{global_diastolic}}(t) + T_{\text{local_diastolic}}(t), \quad t \in (ES, \text{end}]. \quad (2)$$

The end systolic (ES) point was determined as the minimum of the temporal NCC function $r(t)$ based on its demonstrated resemblance to the LV volume curve.¹² The choice of the piecewise function for the systolic and diastolic phases was based on the knowledge that changes in heart rate primarily impact the duration of the diastolic phase.

The global term of the transformation consisted of linear scaling to match the length of the sequences. The local term takes the finer deformations into account. This term was modeled using a 1-D cubic B-spline as follows:

$$T_{\text{local}}(t) = \sum_{i=1}^{N_t} b_i \left(\frac{t}{T_s} \right) \tau_i, \quad (3)$$

where T_s is the length of the $r(t)$ curve, b_i is the B-spline basis function, N_t is the number of spline knots, and τ_i is the control time displacement. The computation of the local term was done by minimizing the following cost function:

$$\text{argmin}_r \|r_R(t_r) - r_F[T(t_f)]\|^2, \quad (4)$$

where r_R, r_F are the reference and the floating characteristic curve, respectively, and t_r, t_f are the time index for reference and floating curve, respectively.

2.2 Spatial Registration

2.2.1 Preprocessing

Prior to spatial registration, three preprocessing steps were applied to both volumes. First, the tip of the apex was masked out for as much as 20% of the sector depth to exclude near field artifacts. The volumes were resampled so that the voxel dimensions were isometric. Trilinear interpolation was used for the resampling of all volumes to the largest voxel dimension in the dataset, which was $1.0 \times 1.0 \times 1.0$ mm. This uniform sampling was done to ensure an isotropic grid and thus avoid bias in a particular direction.

Second, the volumes were convolved with a Gaussian kernel and subsampled n times, where n is the number of levels in the Gaussian pyramid, determined through the inequality

$$\frac{\min(M_x, M_y, M_z)}{2^{n-1}} > k, \{n | n \in \mathbb{Z}^+\}, \quad (5)$$

with M_x, M_y, M_z the number of voxels in the x, y, z direction of the volume and k the window size in which a single displacement was computed using the Farneback optic flow. In this study, k was empirically determined to be 9 so that it was sufficient to cover a large range of transformations but still allowed for fast computation.

Lastly, thresholding based on a typical myocardial intensity value in echocardiography was done to ensure that the displacement estimation was based on the dominant tissue structures. Only the voxels with an intensity value higher than the cutoff intensity were retained. For 8-bit data (range: 0 to 255) in this case, the cutoff intensity value was empirically set to 70. Please note that this also ensured that shadow regions and the shape of the sector were ignored.

2.2.2 Three-dimensional Farneback optic flow

The Farneback optic flow method models the volumes as a composite of polynomial functions that can subsequently be used to analytically find the displacement between corresponding points in the different volumes.¹⁷

First, the volumes were decomposed into polynomial basis functions. The volume f can be approximated by polynomials as

$$f(\mathbf{x}) \approx \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c, \quad (6)$$

where \mathbf{A}, \mathbf{b} , and c are the coefficients of the polynomial and $\mathbf{x} = [x, y, z]^T$. With expansion and factorization, the polynomial can be expressed as a summation of basis functions \mathbf{B} modulated by their respective coefficients \mathbf{r}

$$f(\mathbf{x}) \approx \mathbf{B} \mathbf{r}, \quad (7)$$

with

$$\mathbf{B} = [1, x, y, z, x^2, y^2, z^2, xy, xz, yz]$$

$$\mathbf{r} = [r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, r_{10}]^T. \quad (8)$$

Given that f and \mathbf{B} are known, the coefficients \mathbf{r} can readily be computed. The polynomial coefficients \mathbf{A}, \mathbf{b} , and c can be expressed as a function of the coefficients \mathbf{r} as

$$\mathbf{A} = \begin{bmatrix} r_5 & \frac{r_8}{2} & \frac{r_9}{2} \\ \frac{r_8}{2} & r_6 & \frac{r_{10}}{2} \\ \frac{r_9}{2} & \frac{r_{10}}{2} & r_7 \end{bmatrix} \quad \mathbf{b} = [r_2, r_3, r_4]^T \quad c = r_1. \quad (9)$$

The complete derivation can be found in Appendix A.

For robustness, the computation was done over the k^3 neighborhood instead of point-by-point and solved in the weighted least square sense.¹⁶

Second, the displacement was estimated as a function of the polynomial coefficients. Similar to optic flow methods in general, Farnebäck optic flow assumes brightness constancy in the sense that an exact intensity pattern exists in both the floating volume f_F and the reference volume f_R but displaced by \mathbf{d} ,

$$f_F(\mathbf{x}) = f_R(\mathbf{x} - \mathbf{d}). \quad (10)$$

Deriving from the assumption, the displacement \mathbf{d} can be expressed as

$$\mathbf{d} = \mathbf{A}_d^{-1} \Delta \mathbf{b}, \quad (11)$$

where

$$\mathbf{A}_d = \frac{\mathbf{A}_R + \mathbf{A}_F}{2} \quad \Delta \mathbf{b} = -\frac{1}{2}(\mathbf{b}_F - \mathbf{b}_R), \quad (12)$$

and \mathbf{A}_R , \mathbf{b}_R and \mathbf{A}_F , \mathbf{b}_F are the polynomial coefficients for reference and floating volume, respectively. The extended derivation can be found in Appendix B.

It should be noted that the constant term of the polynomial c [Eq. (6)] is not taken into account in the displacement computation. As a result, local overall brightness variations do not affect the accuracy of the displacement, which could be advantageous in ultrasound image registration.

To improve robustness, Farnebäck suggested the computation of the displacement over a neighborhood of voxels.¹⁶ As such, as mentioned in Sec. 2.2.1, a single displacement was computed over a k^3 neighborhood. Hereto, weighting w was applied over the neighborhood $\Delta \mathbf{x}$, where $-\frac{k}{2} < \Delta \mathbf{x} < \frac{k}{2}$, as

$$\sum_{\Delta \mathbf{x}} w(\Delta \mathbf{x}) \|\mathbf{A}_d(\mathbf{x} + \Delta \mathbf{x}) \mathbf{d}(\mathbf{x}) - \Delta \mathbf{b}(\mathbf{x} + \Delta \mathbf{x})\|^2. \quad (13)$$

By minimizing the above expression for $\mathbf{d}(\mathbf{x})$, the averaged displacement estimate can be determined. This was done by equating its derivative to zero, resulting in

$$\mathbf{d}(\mathbf{x}) = \left[\sum_{\Delta \mathbf{x}} w(\Delta \mathbf{x}) \mathbf{A}_d^T \mathbf{A}_d \right]^{-1} \left[\sum_{\Delta \mathbf{x}} w(\Delta \mathbf{x}) \mathbf{A}_d^T \Delta \mathbf{b} \right]. \quad (14)$$

The notations for $\mathbf{A}_d(\mathbf{x} + \Delta \mathbf{x})$ and $\mathbf{b}_d(\mathbf{x} + \Delta \mathbf{x})$ were shortened for simplicity.

The computation of the coefficients \mathbf{r} of the polynomial basis functions \mathbf{B} [Eq. (7)] was performed similarly over the k^3 neighborhood and solved in the weighted least square sense as mentioned earlier in this section.

The weighting function should ensure that the importance is emphasized on the center voxel for which the displacement is computed. Moreover, this function should be isotropic so that bias toward any specific direction is avoided. These two factors led to the selection of a Gaussian weighting function.

The analysis was performed at the end diastolic frame of both sequences and started from the top of the Gaussian pyramid where the volume is at the coarsest level.

2.2.3 Rigid regularization with Procrustes analysis

As only spatial registration of ultrasound volumes taken at corresponding time points is considered in this study, the condition of rigid transform can be imposed. The Farnebäck optic flow

provides the point-to-point correspondences between the tissue voxels among the volumes. To obtain the optimum rigid transformation T that can map all the voxels in the floating volume f_F to the reference volume f_R coordinate system, Procrustes analysis was used by optimization of the objective function

$$\min_T \sum_{\mathbf{x}} \|\mathbf{x}_R - T[\mathbf{x}_R - \mathbf{d}(\mathbf{x}_R)]\|^2, \quad (15)$$

where \mathbf{x}_R is the spatial coordinates of point correspondences in the reference volume coordinate system and \mathbf{d} is the previously computed displacement. The computed optimum rigid transform T had 6 deg of freedom containing three rotations and three translations.

2.2.4 Convergence criteria

The motivation behind the proposed multiscale iterative implementation was twofold: (1) to increase the range of motion that the method would allow to detect and (2) for computation efficiency. An effective stopping criteria ensured that iterations were run only as required. Following the expression in Eq. (13) that would ideally be zero for optimum \mathbf{d} in the absence of noise, an expression for an indication of displacement error e could be derived. Indeed by substituting the expression for optimum \mathbf{d} back into Eq. (13), the error can be expressed as

$$e(\mathbf{x}) = \left[\sum_{\Delta \mathbf{x}} w(\Delta \mathbf{x}) \Delta \mathbf{b}^T \Delta \mathbf{b} \right] - \mathbf{d}(\mathbf{x})^T \sum_{\Delta \mathbf{x}} w(\Delta \mathbf{x}) \mathbf{A}_d^T \Delta \mathbf{b}. \quad (16)$$

Since the floating volume f_F is pulled toward the analytical optimal solution at each iteration, the error was assumed to decrease monotonically while iterating. When the last two gradient values of the error hit a plateau or changed sign then a minimum is determined. In case no convergence was found, a maximum number of iterations was empirically set to 20.

2.2.5 Implementation details

Several parameters of the Farnebäck method (Sec. 2.2.2) such as the window size k and the σ of the Gaussian weighting function could be tuned to balance out the accuracy with the motion range. Ideally, both were to be set at a comparable magnitude to the motion to be detected. Window size k determined the amount of data to be processed per voxel, hence the computation time. As for Gaussian σ , its size was restricted by the window size k as to maintain the isotropy of the weighting function. Larger σ with comparable size to k would result in a severely truncated Gaussian weighting function that behaved differently in the diagonal direction. Running the disparity computation iteratively allowed for larger motion range detection with the same internal parameters setting. Incorporation of multiscale level further enhanced the range as well as improved the computation efficiency by distributing the computation burden into coarser image scale and reducing the computation requirement on the finer scale.

In this study, the window size k was tuned for the optimum results and was set to 9, while the σ of the Gaussian kernel was empirically set to 1.5 voxels. The relative size of σ to k implied that 3σ was contained in the k^3 neighborhood implying that 99.7% of the weight was contained within the $k/2$ radius around the center voxel.

After setting the value for k and σ , the intensity threshold (Sec. 2.2.1) was also tuned for the best results and was set to 70. Further investigation revealed that the set threshold is just above the average typical intensity value of the blood pool in the data sets. Thus, it adequately included the dominant tissue structures in the registration algorithm.

3 Experiments

The proposed method was implemented in MATLAB® (The MathWorks, Inc., Natick, Massachusetts) and tested on *in vivo* data from 10 healthy subjects (six females and four males) using a Vivid E9 (GE Healthcare) ultrasound scanner with a 4V 3-D probe. All subjects gave their informed consent prior to the acquisition. The acquisition protocol was similar to the one described in Ref. 7. All images were acquired with multibeat acquisition over four heart cycles using ECG gating and a sector width of 65 deg. They were also acquired during one breath hold in order to minimize stitching artifact. Images were taken from several positions as shown in Fig. 2: (1) from the conventional position for imaging the LV in an apical view, (2) aimed toward the interventricular septum, (3) aimed at the right ventricle (RV), (4) toward the lateral wall of the LV, (5) further lateral wall of the LV, (6) and (7) from one intercostal space above and below the first position. Since not all of the volunteers had a good acquisition window for all of the aforementioned positions, more than one image was taken at the same position but with slight tilt in the acquisition angle to differentiate the image content in some cases. The number of acquisitions per subject therefore ranged from 4 to 8 volumes.

In most cases, the volumes that were taken from position 1 were used as the reference volume because of its superior image quality in comparison to volumes taken from other nonconventional views. From a total of 50 volume pairs, 12 were chosen at random to assess the accuracy of the temporal registration while 20 were chosen to assess the spatial accuracy.

For validation of the temporal accuracy, the mitral valve opening (MVO) time was manually annotated in each of the 12 volume pairs. The temporal proximity of the landmarks after temporal registration with the proposed method was contrasted with the proximity obtained through linear scaling. In addition, as an alternative quantitative measure of registration accuracy, the mean NCC between corresponding frames across the sequence was computed.

For assessment of the spatial accuracy, the endocardial border of the LV was manually contoured in the first frame of each of the randomly chosen volumes. The contouring was performed

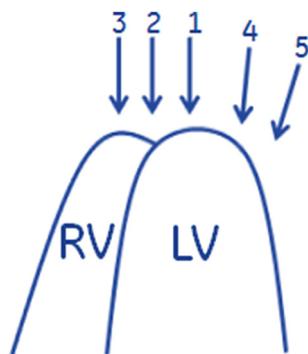


Fig. 2 Illustration of probe positions used in the acquisition protocol. Positions 6 and 7 are one intercostal space above and below position 1, respectively. LV = left ventricle, RV = right ventricle.

by the first author and reviewed by a trained clinical expert. Two distance measures, Hausdorff distance (HD) and mean absolute distance (MAD), were computed between the contours before and after registration. HD measures the maximum distance and MAD the MAD between both contours. For a fair comparison, only the LV contour that was in the overlap area was used to assess the distances, as some acquisition views did not contain the full LV border. For benchmarking, HD and MAD were also computed for alignment using Procrustes analysis as mentioned in Sec. 2.2.3 using the point correspondences from the manual contours.

Additionally, a voxel-based registration method using the `imregtform` function of the MATLAB® Image Processing Toolbox™ was used as a comparison. The MATLAB® image registration framework is similar to the open-source Insight Toolkit (ITK)²² registration framework with fewer options for similarity metrics and optimizers. The chosen voxel-based method used the mean squared distance between the reference and floating volumes as the cost function and a gradient descent algorithm as optimizer. Moreover, the method was set to solve rigid transformations, similar to the proposed method. The voxel-based method was also performed in a multiscale scheme using three levels as set by default. Two parameters, i.e., the maximum number of iterations and the maximum step size were tuned to ensure that enough number of iterations was allowed to reach a minimum. These parameters were tuned to give the lowest HD and MAD for the current data sets and were set to 100 and 0.01, respectively.

Furthermore, to demonstrate the capability of the proposed spatial registration method in widening the field of view, the percentage change of field of view from prior to after the registration was computed using a previously reported measurement by Rajpoot et al.⁷ as follows:

$$\Delta(\text{fov}) = \left[\frac{\text{FOV}_f}{\frac{1}{N} \sum_{i=1}^N \text{FOV}_i} - 1 \right] * 100, \quad (17)$$

where

$$\text{FOV} = \sum_{v=1}^n d(v), \quad (18)$$

$d(v) = 1$ for voxels v belonging to the 3-D echo sector and 0 for voxels v belonging to the background, FOV_f is the FOV computation for the fused volume, N is the number of volumes that were being fused and n is the total number of voxels in the assessed volume.

To test for statistical significance of changes due to the registration process, a paired t -test for two tailed distributions was performed.

4 Results

The summary of the results for temporal registration can be seen in Table 1. The shown MVO errors are in absolute value. In comparison to the linear method, the proposed method resulted in less MVO error and higher NCC ($p < 0.05$). The average interframe time period for all sequences was 47 ± 4 ms which is more than twice the MVO error from the proposed method.

Figure 3 shows an example of the spatial alignment between two volumes before and after registration with the proposed

Table 1 Temporal accuracy validation in terms of MVO time error and NCC.

Method	MVO error (ms)	NCC
Linear	70.4 ± 84.6	0.855 ± 0.058
Proposed	19.6 ± 13.6	0.856 ± 0.058

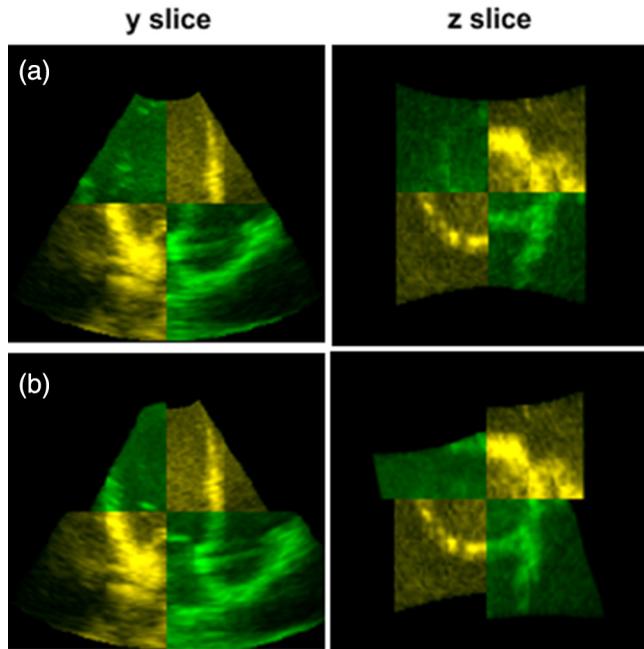


Fig. 3 Example of registration result using the proposed method. Mid volume slices (a) before registration and (b) after registration.

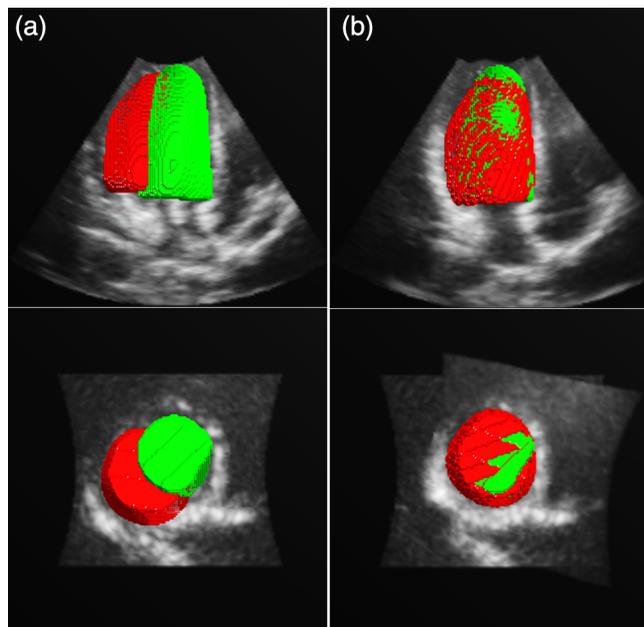


Fig. 4 Left ventricle endocardial surface alignment (a) before and (b) after proposed registration superimposed on mid-volume slices of fused volume. Surface of reference volume is shown in green, surfaces of floating and registered volume are shown in red.

method. The mid-volume slices before and after registration are shown.

An example of the endocardial border of the LV that was used to evaluate the spatial accuracy is shown in Fig. 4. The surfaces were superimposed on the mid-volume slices of the fused volume without and with registration. The reference surface is shown in green and the floating and registered surface are shown in red.

The HD and MAD before and after registration for all cases are given in Figs. 5 and 6, respectively. The registration results from the proposed, the voxel-based method and Procrustes analysis using the manual contours are shown. The line of equality is shown as a solid black line and the area under it refers to improved alignment after registration. The average number of iterations of the proposed method for all scale levels in all cases were 3.6 ± 4.9 iterations.

A summary of the average distances before and after registration is given in Table 2. The proposed method, the voxel-based method and the manual contour registration significantly reduced the misalignment in terms of HD and MAD ($p < 0.01$).

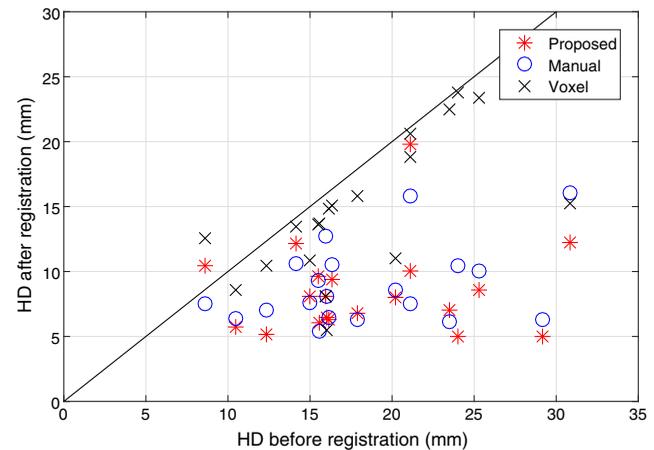


Fig. 5 HD between the contours before and after registration with proposed, manual, and voxel-based method for all volume pairs. Line of equality is shown as a solid black line.

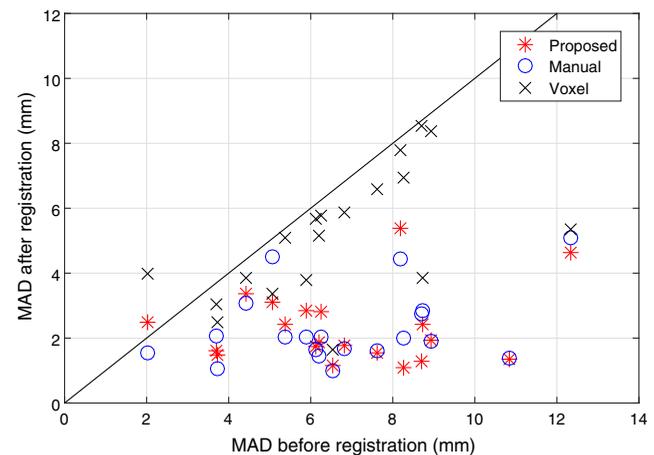


Fig. 6 MAD between the contours before and after registration with proposed, manual and voxel-based method for all volume pairs. Line of equality is shown as a solid black line.

Table 2 HD and MAD prior and after alignment with proposed method and manual contour.

Method	HD (mm)	MAD (mm)
Before	18.47 ± 5.88^b	6.79 ± 2.49^b
Voxel	15.28 ± 5.36^{ab}	5.42 ± 2.24^{ab}
Proposed	8.51 ± 3.45^a	2.31 ± 1.14^a
Manual	8.94 ± 3.06^a	2.31 ± 1.16^a

^a $p < 0.01$ against before registration.

^b $p < 0.01$ against proposed registration.

Moreover, the proposed method resulted in significantly lower HD and MAD in comparison to the voxel-based method ($p < 0.01$), while no statistically significant difference was found between the proposed method and the manual one ($p > 0.01$).

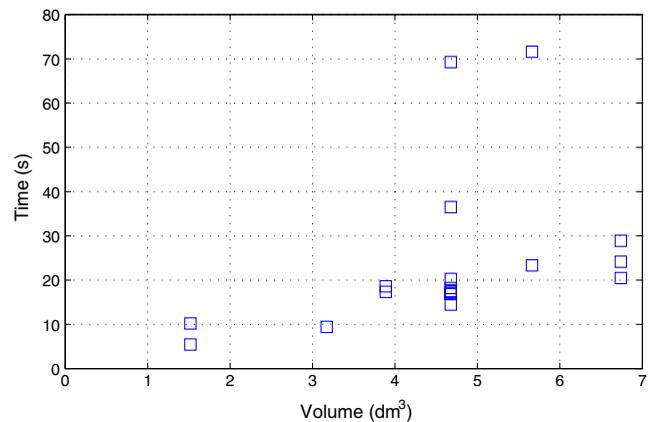
The percentage change in field of view was computed for the fused volume after spatial registration. Among the 20 randomly chosen cases, 20 fused volumes were formed using two contributing volumes and seven fused volumes were formed using three contributing volumes. The summarized results for the field of view change of all fused volumes is given in Table 3. It can be seen that all fused volumes had an increased field of view in comparison to their individual contributing volumes.

An example of spatiotemporal registration result can be seen in Fig. 7. The mid-volume slices of the fused volume across all corresponding frames are shown in the video. The border of each individual volume is highlighted.

The validation was run on a laptop with an Intel CORE i7-3740QM CPU at 2.7 GHz under the 64-bit MATLAB[®] development environment. Figure 8 shows the computation time for each volume in reference to its metric volume. The average computation time for all volumes was 23.7 ± 17.0 s whereas the dimension of the volumes ranged from $118 \times 118 \times 109$ to $194 \times 193 \times 170$ voxels.

Table 3 Percentage change in field of view $\Delta(\text{fov})$ after spatial registration.

Nr. volumes	2	3
% $\Delta(\text{fov})$	15.7 ± 3.62	25.4 ± 4.87

**Fig. 7** Still image of the spatiotemporal registration result video (Video 1, MP4, 0.4 MB) [URL: <http://dx.doi.org/10.1117/1.JMI.3.3.037001.1>].**Fig. 8** Computation time in relation to registered echocardiographic image volume.

5 Discussion

The proposed spatiotemporal registration method based on a multiscale iterative Farneback optic flow and piecewise 1-D cubic B-spline interpolation is a competitive approach for accurate image registration.

As an extension to our previous work,²⁰ temporal registration was incorporated. Regarding the temporal accuracy results (Table 1), a reduction in MVO error in comparison to the linear method has been shown. The resulting MVO error is also less than half of the average temporal resolution of the current data sets. In addition, the slight increase of NCC indicated higher similarity between corresponding frames after the proposed temporal registration in comparison to the linear alignment.

For the spatial registration, the average number of iterations from all scale levels (3.6 ± 4.9 iterations) in all cases was well below the maximum (20 iterations). This demonstrated the effectiveness of the added convergence criteria in comparison to our previous work²⁰ where all numbers of iterations were set empirically.

Table 2 shows that the proposed voxel-based registration method and registration using manual contours significantly reduced the misalignment of the volumes. The results also show no statistically significant difference between the results from the proposed method and registration using manual contouring. This is further supported by Figs. 5 and 6, where it can be seen that the proposed method was able to perform registration with comparable accuracy as the registration using manual contours.

In comparison to the voxel-based method, the proposed method resulted in statistically significantly lower MAD and HD. Moreover, as seen in Figs. 5 and 6, the HD and MAD from the voxel-based method often appeared just under the line of equality, which means that the computed solutions were close by the initial positions. These results indicated that the voxel-based method was likely to converge to a local solution. In contrast, the proposed method managed to converge to a global solution that is close to the manual registration even though it is also an iterative method and was given the same initial position. A comparison with a feature-based registration using 3-D SIFT²³ in combination with RANSAC for outliers removal²⁴ was also considered; however, a robust feature-based registration cannot be achieved for our data sets due to too few good feature matches in comparison to the degree of

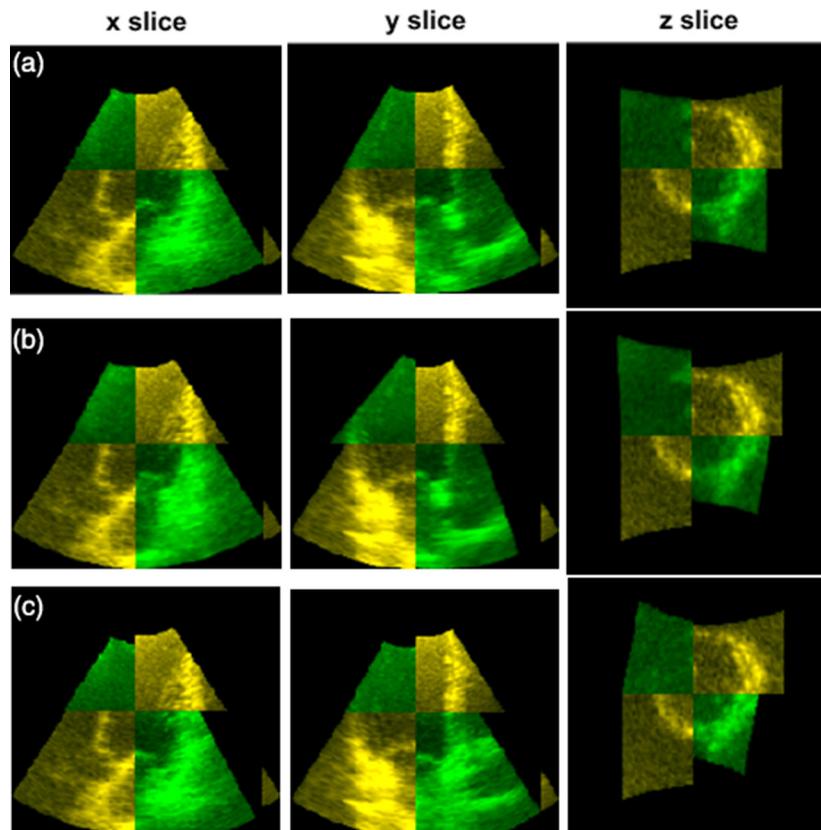


Fig. 9 Example of mid-volume slices (a) before and (b) after registration using proposed and (c) manual method for the one case where HD and MAD are higher after registration with proposed method.

freedom in the registration, thus the registration was often under determined.

Although the manual contouring was used as a benchmark in this accuracy evaluation, this approach has limitations. The quality of manual contouring is user and acquisition dependent, as one structure might appear better in one acquisition than in another. However, *in vivo* data presented the most realistic challenges for the method which motivated the current validation. In the one case where the proposed method produced greater misalignment in terms of HD and MAD, it could be observed that the initial misalignment was relatively small. Due to the variability of the manual contouring, the increased misalignment in this case could be superficial. Examining closely in Fig. 9, where the mid-volume slices of the one case before and after registration with the proposed and manual methods are shown, it is undetermined by visual inspection whether the proposed registration worsens the alignment. Nevertheless, the work of Mulder et al.¹⁴ showed that the interobserver variation of LV annotation in three anatomical landmarks at ED was in the range of 4.1 mm for the median and 8.4 mm for the maximum value. These values fall within the range between the average MAD and HD of the proposed registration, which further supports the claim of the proposed method's ability to register as well as the manual registration.

Another limitation of the validation approach is the limited variation of cases in the data sets. All images were acquired by one expert clinician and the images for each subject were taken in one examination. This approach resulted in approximately consistent image quality across the whole data sets, which

limited the possibility of performance evaluation with varying image quality.

In regard to the widening of the field of view, as expected, fused volumes using either two or three contributing volumes resulted in an increase of the field of view. Nevertheless, although the trend suggests that the more contributing volumes produce higher increase in the field of view, the effect will strictly depend on the amount of overlap between the contributing volumes.

The registration computation time varied from case to case. There seems to be a positive correlation between the computation time and the size of the volume, thereupon the amount of data to be processed. However, it can also be noted that there is variability between the computation time of same sized volumes. This is a consequence of other factors such as initial misalignment as well as image content, which also affect the required number of iterations and hence the computation time.

The method presented here allows combination of several overlapping volumes, resulting in a single volume with larger field of view, which has applications in both transthoracic imaging (e.g., for simultaneous assessment of the left and right heart) and for transesophageal or intracardiac imaging where the field of view is limited. In this paper, validation of the method was performed using transthoracic images. For potential application on transesophageal and intracardiac imaging, the initial field of view is typically smaller, which could lead to less overlap between different acquisitions and less tissue structure contents to align. One way to avoid the problem is to set an acquisition protocol that will ensure enough context is captured

in the image. In combination with the proposed registration method, this will potentially help overcoming the limited field of view challenge that is inherent to TEE and ICE probes that are used in interventional monitoring.

Furthermore, to be used in the interventional setting, the computation time is a critical issue. The current computation time was achieved through implementation in MATLAB[®], which allowed for fast prototyping but had not been extensively optimized. Thus, porting the algorithm into C/C++ or GPU could be considered an option to enhance the performance and make it more clinically relevant for interventional monitoring.

The heart is a dynamic object that deforms over time. However, the rigid spatial transform in this implementation was applied under the assumption that the registered volumes were acquired at the same point in the cardiac cycle. Hence physiological cardiac deformation between registered volumes was not expected. Moreover, the rigid regularization restricted the computed transformation to be the result of relative displacement between the probe and the subject instead of other stochastic nonphysiological deformation information in the volume such as noise. Therefore, incorporation of temporal alignment ensures that the spatial registration is performed on temporally corresponding frames in the cardiac cycle. Thus, the presented methodology offered a solution for spatiotemporal registration between 3-D echocardiographic recordings as a part of the data compounding framework.

6 Conclusion

An image-based spatiotemporal registration method without manual initialization for echo recordings is proposed. Hereto, the temporal registration and spatial registration are considered separately where the former is done using a piecewise 1-D cubic B-spline interpolation and the latter through a multiscale iterative Farneback optic flow method. The temporal accuracy was validated by manual annotation of MVO event showing that the achieved temporal accuracy is in the range of less than half of the echo sequences time resolution. The spatial accuracy was validated by manual contouring of LV endocardium showing that the misalignment in terms of HD and MAD is significantly reduced. Comparable registration accuracy with manual contouring was also achieved. The fusion of the spatially registered volumes has resulted in an increase of field of view.

The proposed registration method fits into the first step of data compounding, which would ensure that the combined information from several echocardiography recordings can be used to produce higher quality images.

Acknowledgments

This work was supported by Marie Curie Initial Training Network (USART-project, Grant ID: PITN-GA-2012-317132). The funding institution had no involvement in any part of the study. The authors would also like to acknowledge J. Pedrosa for the help in editing the article.

Appendix A: Polynomial Coefficient Derivation

Using the expression for \mathbf{B} and \mathbf{r} from Eq. (8), the compact polynomial expression in Eq. (7) can be expanded to

$$f(\mathbf{x}) \approx r_1 + r_2x + r_3y + r_4z + r_5x^2 + r_6y^2 + r_7z^2 + r_8xy + r_9xz + r_{10}yz, \quad (19)$$

which can then be factorized to

$$f(\mathbf{x}) \approx r_1 + [r_2 \ r_3 \ r_4] \begin{bmatrix} x \\ y \\ z \end{bmatrix} + [x \ y \ z] \begin{bmatrix} r_5 & r_8 & r_9 \\ r_8 & r_6 & r_{10} \\ r_9 & r_{10} & r_7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (20)$$

By equating the coefficients in Eq. (20) to the polynomial expression in Eq. (6), the expression for \mathbf{A} , \mathbf{b} , and c can be expressed as in Eq. (9).

Appendix B: Optic Flow Displacement Derivation

Following up from the optic flow assumption in Eq. (10), the expression for the displacement \mathbf{d} can be derived as a function of the polynomial coefficients. By expansion and equating coefficients, the polynomial coefficients of the floating volume \mathbf{A}_F , \mathbf{b}_F , and c_F can be expressed in terms of the polynomial coefficients of the reference volume \mathbf{A}_R , \mathbf{b}_R , and c_R as follows

$$\begin{aligned} f_F(\mathbf{x}) &= (\mathbf{x} - \mathbf{d})^T \mathbf{A}_R (\mathbf{x} - \mathbf{d}) + \mathbf{b}_R^T (\mathbf{x} - \mathbf{d}) + c_R \\ f_F(\mathbf{x}) &= \mathbf{x}^T \mathbf{A}_F \mathbf{x} + \mathbf{b}_F^T \mathbf{x} + c_F \\ \mathbf{A}_F &= \mathbf{A}_R \\ \mathbf{b}_F &= \mathbf{b}_R - 2\mathbf{A}_R \mathbf{d} \\ c_F &= \mathbf{d}^T \mathbf{A}_R \mathbf{d} - \mathbf{b}_R^T \mathbf{d} + c_R. \end{aligned} \quad (21)$$

Thus, from the expression for \mathbf{b}_F , the displacement \mathbf{d} can thus be expressed as

$$\mathbf{d} = -\frac{1}{2} \mathbf{A}_R^{-1} (\mathbf{b}_F - \mathbf{b}_R). \quad (22)$$

According to the expressions in Eq. (21), \mathbf{A}_d is theoretically equal to \mathbf{A}_R and \mathbf{A}_F . However, as the value between \mathbf{A}_R and \mathbf{A}_F could differ due to noise, the use of the mean of \mathbf{A}_R and \mathbf{A}_F was suggested by Farneback¹⁶ as a practical solution. The final expression for displacement \mathbf{d} is stated in Eq. (11).

References

1. P. Gripari et al., "Real-time three-dimensional transoesophageal echocardiography: a new intraoperative feasible and useful technology in cardiac surgery," *Int. J. Cardiovasc. Imaging* **26**(6), 651–660 (2010).
2. V. Grau, H. Becher, and J. A. Noble, "Phase-based registration of multi-view real-time three-dimensional echocardiographic sequences," *Lect. Notes Comput. Sci.* **4190**, 612–619 (2006).
3. H. Mulder et al., "Simultaneous pairwise registration for image mosaicing of TEE data," in *IEEE 10th Int. Symp. on Biomedical Imaging (ISBI '13)*, pp. 242–245 (2013).
4. D. Ni et al., "Volumetric ultrasound panorama based on 3D SIFT," in *Medical Image Computing and Computer-Assisted Intervention—(MICCAI '08)*, pp. 52–60, Springer (2008).
5. R. J. Schneider et al., "Real-time image-based rigid registration of three-dimensional ultrasound," *Med. Image Anal.* **16**(2), 402–414 (2012).
6. L. J. Brattain and R. D. Howe, "Real-time 4D ultrasound mosaicing and visualization," *Lect. Notes Comput. Sci.* **6891**, 105–112 (2011).
7. K. Rajpoot et al., "Multiview fusion 3-D echocardiography: improving the information and quality of real-time 3-D echocardiography," *Ultrasound Med. Biol.* **37**(7), 1056–1072 (2011).

8. M. J. Gooding et al., "Investigation into the fusion of multiple 4-D fetal echocardiography images to improve image quality," *Ultrasound Med. Biol.* **36**(6), 957–966 (2010).
9. C. Yao et al., "Spatial compounding of large sets of 3D echocardiography images," *Proc. SPIE* **7265**, 726515 (2009).
10. V. Grau and J. A. Noble, "Adaptive multiscale ultrasound compounding using phase information," *Lect. Notes Comput. Sci.* **3749**, 589–596 (2005).
11. W. Zhang, J. Noble, and J. Brady, "Spatio-temporal registration of real time 3D ultrasound to cardiovascular MR sequences," *Lect. Notes Comput. Sci.* **4791**, 343–350 (2007).
12. A. Perissinotto et al., "Robust temporal alignment of multimodal cardiac sequences," *Proc. SPIE* **9413**, 94131K (2015).
13. D. Perperidis, R. H. Mohiaddin, and D. Rueckert, "Spatio-temporal free-form registration of cardiac MR image sequences," *Med. Image Anal.* **9**(5), 441–456 (2005).
14. H. W. Mulder et al., "Multiframe registration of real-time three-dimensional echocardiography time series," *J. Med. Imaging* **1**(1), 014004 (2014).
15. C. Wachinger, W. Wein, and N. Navab, "Three-dimensional ultrasound mosaicing," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, pp. 327–335, Springer (2007).
16. G. Farneback, "Disparity estimation from local polynomial expansion," in *Proc. of the SSAB Symp. on Image Analysis*, Norrköping, pp. 77–80 (2001).
17. G. Farneback, *Polynomial Expansion for Orientation and Motion Estimation*, PhD Thesis, Dissertation No. 790, Linköping University, Linköping, Sweden (2002).
18. G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proc. of the 13th Scandinavian Conf. on Image Analysis (SCIA '03)*, pp. 363–370, Springer-Verlag, Berlin, Heidelberg (2003).
19. A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods," *Int. J. Comput. Vision* **61**(3), 211–231 (2005).
20. A. Danudibroto et al., "3D Farneback optic flow for extended field of view of echocardiography," *Lect. Notes Comput. Sci.* **9126**, 129–136 (2015).
21. M. Alessandrini et al., "Simulation of realistic echocardiographic sequences for ground-truth validation of motion estimation," in *19th IEEE Int. Conf. on Image Processing*, pp. 2329–2332 (2012).
22. L. Ibanez et al., *The ITK Software Guide*, 1st ed., Kitware, Inc., Clifton Park, New York (2003).
23. B. Rister et al., "Scale- and orientation-invariant keypoints in higher-dimensional data," in *IEEE Int. Conf. on Image Processing (ICIP '15)*, pp. 3490–3494 (2015).
24. M. Zulfiani, "Ransac toolbox for MATLAB," (2008), <http://www.mathworks.com/matlabcentral/fileexchange/18555> (3 May 2016).

Adriyana Danudibroto is an industrial PhD student at the University of Leuven, Belgium, and GE Vingmed Ultrasound, Norway. She received her BEng (Hons) in electrical and electronics engineering at the University of Auckland, New Zealand, in 2010 and MSc degree in computer vision and robotics from an Erasmus Mundus Master course held by Heriot-Watt University, UK, University of Girona, Spain, and University of Burgundy, France, in 2012. Her research interests include medical image processing and machine learning.

Jørn Bersvendsen received his MSc degree in engineering cybernetics from the Norwegian University of Science and Technology in 2011 and is currently a PhD research fellow at GE Vingmed Ultrasound and the University of Oslo, studying segmentation of cardiac chambers in three-dimensional echocardiography.

Olivier Gérard is currently holding the position of program manager and lead feature designer at GE Cardiovascular Ultrasound. He has earned an MSc degree of computer science from Florida Tech and an industrial PhD from Paris 6 University. He has been active with research and product development with x-ray and ultrasound images for the last 18 years. He is co-author of 16 issued patents and numerous articles in both technical in clinical aspects of echocardiography.

Oana Mirea is an assistant professor at the University of Medicine of Craiova since 2015. She received her MD and PhD degrees in 2009 and 2013 from the University of Medicine of Craiova and is currently training in cardiology. She received a research grant from the European Association of Cardiovascular Imaging in 2015, and the project was carried out in collaboration with KU Leuven University, Belgium. Her current research is focused on cardiac imaging.

Jan D'hooge is a professor at the Cardiovascular Sciences Department at the University of Leuven. He received his MSc and PhD degrees in physics at the University of Leuven, Belgium, in 1994 and 1999, respectively. He then worked as a postdoctoral researcher in the department of electrical engineering in collaboration with the departments of cardiology and acoustics at the same university. His current research interests include myocardial tissue characterization and strain imaging in combination with cardiac pathophysiology.

Eigil Samset received his MSc degree in engineering cybernetics NTNU in 1997 and his PhD from the University of Oslo in 2003. His postdoctoral was performed at Brigham and Women's Hospital. He is currently a research project coordinator at GE Cardiovascular Ultrasound coordinating the Center for Cardiological Innovation. Eigil Samset is also an adjunct professor at the University of Oslo. His 17 years of experience span academic and industrial R&D and leadership within medical imaging.