

# PROCEEDINGS OF SPIE

[SPIEDigitalLibrary.org/conference-proceedings-of-spie](https://SPIEDigitalLibrary.org/conference-proceedings-of-spie)

## Robust spatio-temporal registration of 4D cardiac ultrasound sequences

Jørn Bersvendsen, Matthew Toews, Adriyana Danudibroto, William M. Wells, Stig Urheim, et al.

Jørn Bersvendsen, Matthew Toews, Adriyana Danudibroto, William M. Wells, Stig Urheim, Raúl San José Estépar, Eigil Samset, "Robust spatio-temporal registration of 4D cardiac ultrasound sequences," Proc. SPIE 9790, Medical Imaging 2016: Ultrasonic Imaging and Tomography, 97900F (1 April 2016); doi: 10.1117/12.2217005

**SPIE.**

Event: SPIE Medical Imaging, 2016, San Diego, California, United States

# Robust Spatio-Temporal Registration of 4D Cardiac Ultrasound Sequences

Jørn Bersvendsen<sup>a,b,c</sup>, Matthew Toews<sup>d</sup>, Adriyana Danudibroto<sup>a,e</sup>, William M. Wells III<sup>f</sup>, Stig Urheim<sup>g</sup>, Raúl San José Estépar<sup>f</sup> and Eigil Samset<sup>a,b,c</sup>

<sup>a</sup>GE Vingmed Ultrasound, Horten, Norway;

<sup>b</sup>University of Oslo, Oslo, Norway;

<sup>c</sup>Center for Cardiological Innovation, Oslo, Norway;

<sup>d</sup>École de Technologie Supérieure, Montreal, Canada;

<sup>e</sup>University of Leuven, Leuven, Belgium;

<sup>f</sup>Brigham and Women's Hospital, Harvard Medical School, Boston, USA;

<sup>g</sup>Oslo University Hospital, Oslo, Norway

## ABSTRACT

Registration of multiple 3D ultrasound sectors in order to provide an extended field of view is important for the appreciation of larger anatomical structures at high spatial and temporal resolution. In this paper, we present a method for fully automatic spatio-temporal registration between two partially overlapping 3D ultrasound sequences. The temporal alignment is solved by aligning the normalized cross correlation-over-time curves of the sequences. For the spatial alignment, corresponding 3D Scale Invariant Feature Transform (SIFT) features are extracted from all frames of both sequences independently of the temporal alignment. A rigid transform is then calculated by least squares minimization in combination with random sample consensus. The method is applied to 16 echocardiographic sequences of the left and right ventricles and evaluated against manually annotated temporal events and spatial anatomical landmarks. The mean distances between manually identified landmarks in the left and right ventricles after automatic registration were (mean  $\pm$  SD)  $4.3 \pm 1.2$  mm compared to a reference error of  $2.8 \pm 0.6$  mm with manual registration. For the temporal alignment, the absolute errors in valvular event times were  $14.4 \pm 11.6$  ms for Aortic Valve (AV) opening,  $18.6 \pm 16.0$  ms for AV closing, and  $34.6 \pm 26.4$  ms for mitral valve opening, compared to a mean inter-frame time of 29 ms.

## 1. INTRODUCTION

Ultrasound is the image modality of choice for assessing the heart in clinical routine, offering high frame rate imaging of the beating heart. As 3D ultrasound is being more widely studied, new applications for imaging, visualization and quantification of the heart are emerging.

Registration of 3D ultrasound images has many potential uses, including motion estimation and extended field of view. Several methods have been proposed using elastic registration to study the motion of the heart walls, which can be used to estimate the strain in the myocardium [1, 2, 3, 4]. As 3D ultrasound images are typically acquired in smaller sectors, to maintain adequate spatial and temporal resolution, registration can be used to fuse multiple 3D sectors together. This extends the field of view, allowing the quantification of larger structures while preserving resolution. Methods for spatial registration of 3D ultrasound include optical flow [5], feature-based registration such as Scale Invariant Feature Transform (SIFT) [6], voxel-wise similarity measures such as Normalized Cross-Correlation (NCC) [7], and similarity measures based on local orientation and phase [8].

Ni *et al.* used 3D SIFT features to create a panorama image from several 3D ultrasound images by rigid registration [6]. This work was extended by Schneider *et al.* who used efficient rotation-variant features to provide a transform between consecutive frames in real-time during acquisition [9]. This can be useful for compensating for slight movement of the probe during acquisition, or to cancel movement of anatomy.

Because the heart undergoes a complex contraction that is non-linear with respect to changes in the heart rate, establishing a temporal correspondence between sequences is important for accurate registration. Temporal registration for cardiac images often relies heavily on external recording from ECG. Utilizing image based measurement allows the registration to be independent from this external measurement.

Perperidis *et al.* used Normalized Cross Correlation (NCC) over time for temporal alignment and normalized mutual information for spatial alignment to provide a free-form spatio-temporal registration between MRI sequences [10]. Expanding on this, Zhang *et al.* used a similar approach for 3D ultrasound to MRI registration, using NCC over time for temporal alignment and 3D SIFT features for spatial alignment [11].

There are many aspects of ultrasound image processing that present challenges compared to other medical imaging modalities. Ultrasound often contain artifacts such as acoustic shadows, speckles, and reverberations. This means that several of the intensity-based matching methods used for registering images of other modalities are less appropriate for ultrasound processing. Furthermore, acquisitions can have very different gain settings, temporal and spatial resolutions, and can be acquired from different locations depending on the patient, in order to improve the acoustic window.

In this paper we present a method of registering two partially overlapping 3D cardiac ultrasound sequences in space and time. The main contribution of this work is to solve both the spatial and temporal alignment problems for 4D ultrasound, and to solve the spatial alignment with no user interaction and without any a priori assumptions on either the temporal or spatial alignment of the sequences. The temporal alignment is solved by aligning the NCC-over-time functions of the floating and reference sequences. We then use corresponding 3D SIFT features between all frames, without any assumptions on the temporal alignment, to extract a single rigid transform for the whole cardiac cycle by minimizing a least squares problem using Random Sample Consensus (RANSAC). The method was validated by registering 3D sectors of the left and right ventricles in 16 clinical cases.

## 2. METHODS

Given a reference image sequence  $\mathcal{I}_r(x, y, z, t) = \{I_{r,i}(x, y, z)\}$  and a floating sequence  $\mathcal{I}_f(x, y, z, t) = \{I_{f,i}(x, y, z)\}$ , the objective of the registration method is to find a transform  $\mathcal{T}$  such that any voxel in the transformed sequence  $\mathcal{T}(\mathcal{I}_f)$  coincides with the corresponding voxel in  $\mathcal{I}_r$ .

Following the approach used by others [10, 11], we decouple the registration problem into spatial and temporal domains and solve these separately. We first solve the temporal alignment using the NCC over the cardiac cycle, before extracting a spatial transform from 3D SIFT feature correspondences between all frames.

### 2.1 Temporal Registration

The heart undergoes a complex contraction during the cardiac cycle. Because this contraction is slightly different for each beat and non-linear with changes of heart frequency, different acquisitions will generally not be synchronized, even if performed within a relatively short period of time without external influences.

The NCC over time has been shown to be a characteristic function describing the events of the cardiac cycle in a consistent manner, and has been used for temporal alignment of MRI sequences [10] and between ultrasound and MRI [11]. In both of these studies, the temporal alignment was achieved by first detecting key cardiac events in the NCC-over-time function, which involves calculating its second order derivative.

However, our experiments indicate that the second order derivation is unstable in ultrasound images of high frame rate, because of the inherently low signal-to-noise ratio and the presence of artifacts such as acoustic shadows, speckles and reverberations. This is specially apparent for images of the right ventricle which generally has lower image quality than the left ventricle, due to the challenging acquisition [12]. We therefore propose to solve the temporal registration by aligning the NCC-over-time functions with an optimization problem.

The temporal transform  $\mathcal{T}_{\text{temporal}}$  is modeled as a global affine transform  $\mathcal{T}_{\text{temporal}}^{\text{global}}(t) = \alpha t + \beta$  used to scale the sequences to the same length and compensate for global phase shift, and a local transform  $\mathcal{T}_{\text{temporal}}^{\text{local}}(t)$  used to adjust for non-linear difference in the contraction pattern of the sequences. The local transform is modeled as a 1D B-spline with  $N_t$  knots,

$$\mathcal{T}_{\text{temporal}}^{\text{local}}(t) = \sum_{i=1}^{N_t} b_i \left( \frac{t}{T} \right) \tau_i, \quad (1)$$

where  $T$  is the length of the sequence, and  $b_i$  and  $\tau_i$  are the B-spline basis functions and control time displacements respectively. The combined transform is given by

$$\mathcal{T}_{\text{temporal}}(t) = \mathcal{T}_{\text{temporal}}^{\text{local}}(t) + \mathcal{T}_{\text{temporal}}^{\text{global}}(t). \quad (2)$$

The control time displacements  $\tau_i$  of the local temporal deformation are calculated by solving the minimization problem

$$\vec{\tau} = \arg \min_{\vec{\tau}} |f_r(t) - f_f(\mathcal{T}_{\text{temporal}}(t))|^2, \quad (3)$$

where  $f_r$  and  $f_f$  are the NCC functions of the reference and floating sequences. The optimization problem is solved numerically by sequential quadratic programming. By constraining  $\mathcal{T}_{\text{temporal}}$  to be monotonically increasing, we guarantee that the frame order remains unchanged.

Because 3D cardiac ultrasound is gated by electrocardiography in almost all clinical cases, all image sequences were ordered such that the first frame corresponded to the QRS complex. We therefore assumed a zero phase shift,  $\beta = 0$  in  $\mathcal{T}_{\text{temporal}}^{\text{global}}$ . Figure 1 shows the NCC functions before and after temporal alignment in an example case.

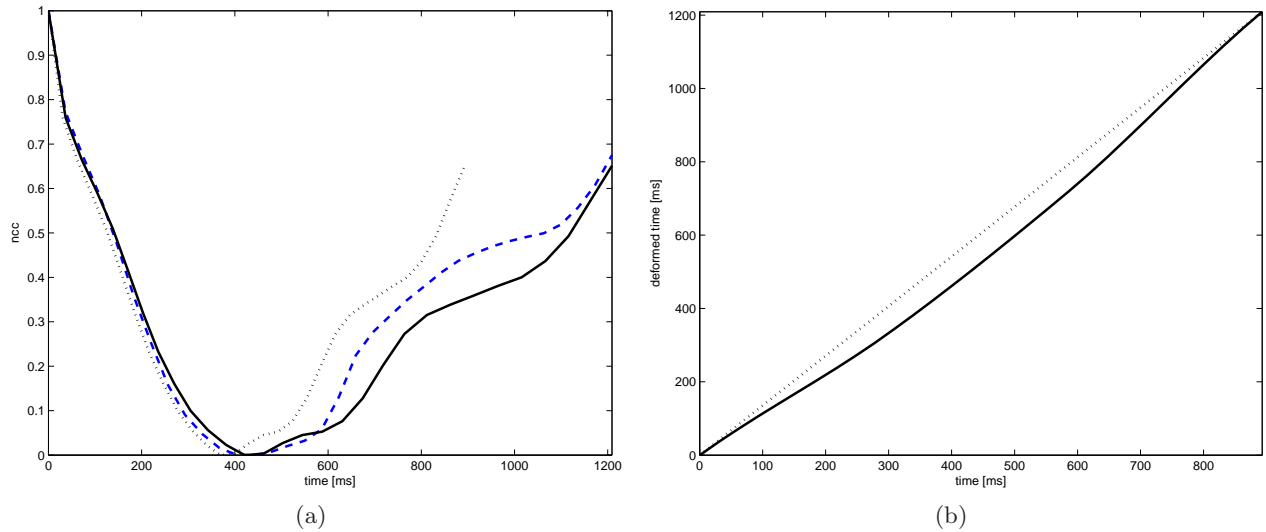


Figure 1. The normalized cross-correlation over time for a reference (dashed) and floating sequence before (dotted) and after (solid) temporal alignment (a), resulting temporal registration (b).

## 2.2 Time-Independent Spatial Registration

When registering sequences of the same patient during the same exam without external influences, it is fair to assume that the true spatial transform between the sectors of two ultrasound sequences is rigid and constant over the cycle. The spatial transform between frames of the reference and floating sequences  $\mathcal{I}_r$  and  $\mathcal{I}_f$  at similar points in the cardiac cycle can thus be assumed to be approximately rigid. This assumption leads a novel time-independent feature-based alignment (FBA) method [13], involving 3D SIFT feature extraction and matching.

First, a set of position-scale pairs  $\{\vec{p}_i, \sigma_i\}$  are extracted in all 3D image frames of each sequence by identifying local maxima and minima of the difference-of-Gaussian function,

$$\{\vec{p}_i, \sigma_i\} = \text{local arg max}_{\vec{p}, \sigma} |f(\vec{p}, \kappa\sigma) - f(\vec{p}, \sigma)| \quad (4)$$

where  $f(\vec{p}, \sigma)$  is the convolution of the image with a Gaussian kernel of variance  $\sigma^2$  and  $\kappa$  the scale sampling rate. Following detection, an orientation is assigned to each feature using local gradient orientation information,

and finally an appearance descriptor is generated from the patch of voxels within the image region defined by  $(\vec{p}_i, \sigma_i)$ . Appearance descriptors are normalized according to image intensity and local geometry, and can thus be used to compute image-to-image correspondences in a manner invariant to monotonic intensity shifts and global similarity transforms.

For time-independent matching, each sequence is modeled as an unordered bag-of-features with spatial but no temporal information. The FBA method [13] is then used to estimate a set of highly probable sequence-to-sequence feature correspondences. Briefly, a set of nearest-neighbor feature correspondences are identified based on the  $\mathcal{L}_2$  distance between appearance descriptors, resulting in a set of corresponding feature positions  $\mathcal{C} = \{\vec{p}_{f,i}, \vec{p}_{r,i}\}$ . We then apply RANSAC to solve the minimization problem

$$\mathcal{T}_{\text{spatial}} = \arg \min_{\mathcal{T}_{\text{spatial}}} \sum_i \|\vec{p}_{r,i} - \mathcal{T}_{\text{spatial}}(p_{f,i})\|^2, \quad (5)$$

while simultaneously rejecting outlier correspondences.

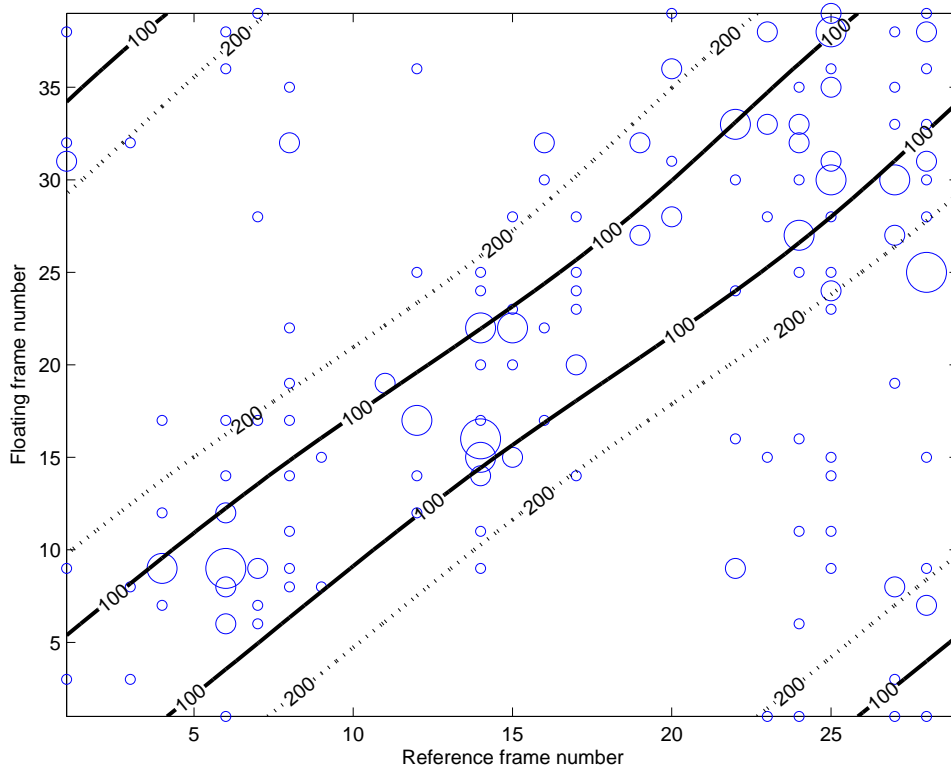


Figure 2. Correspondence density of feature matching between all reference and floating frames in one case. The size of each circle is proportional to the number of correspondences for a single frame pair. The black lines are contour lines of the time displacement, in milliseconds, after temporal alignment, illustrating the deformable temporal alignment.

### 3. VALIDATION

The registration method was evaluated on 3D transthoracic echocardiographic studies of 16 clinical cases of patients with aortic insufficiency. Each case had two acquisitions showing the left and right ventricles in different sectors with varying degree of overlap, both acquired from an apical position. All images were recorded on a Vivid E9 scanner using a 4V-D transducer (GE Vingmed Ultrasound AS, Horten, Norway). Each sequence contained a single heart cycle acquired using multi-beat during breath hold, containing on average 38 frames.

The Mitral Valve (MV) and Aortic Valve (AV) opening and closing events were manually tagged for each recording by an experienced cardiologist. Anatomical landmarks were identified in the MV opening and MV closing frames. These included the AV, MV and Tricuspid Valve (TV) center points, and the AV-MV junction.

## 4. RESULTS

### 4.1 Temporal alignment

The signed temporal alignment error between the key cardiac events were (mean  $\pm$  SD)  $4.9 \pm 18.2$  ms for AV opening,  $-2.0 \pm 24.9$  ms for AV closing and  $-5.9 \pm 44.0$  ms for MV opening. The corresponding absolute errors were  $14.4 \pm 11.6$  ms,  $18.6 \pm 16.0$  ms, and  $34.6 \pm 26.4$  ms. For reference, the inter-frame time was  $29 \pm 5.1$  ms.

### 4.2 Spatio-temporal alignment

The average euclidean distance between corresponding manually identified anatomical landmarks under the automatic registration was  $4.3 \pm 1.2$  mm, compared to an average distance of  $2.9 \pm 0.7$  mm with a Procrustes alignment between all manual landmarks.

Figure 1(a) shows an example of the NCC curves before and after the temporal alignment, as well as the resulting temporal transform. Figure 1(b) illustrates the densities of 3D SIFT correspondences across the cardiac cycle. It is clear that, although no assumption is made on the temporal alignment, correspondences are more frequent in temporally aligned frames. Finally, examples of the resulting spatio-temporal alignment are shown in figure 3.

## 5. DISCUSSION

In this paper, a spatio-temporal registration method for 3D cardiac image sequences has been presented and evaluated on ultrasound images. The method performed close to manual registration in both time and space, while requiring no manual user input.

The temporally aligned valve events were all close to the ground truth, and within clinically acceptable values. The temporal errors were noticeably larger for the MV opening event compared to AV opening and closing, which is expected as MV opening is the last valvular event in the cardiac cycle and thus furthest from the ECG-gated first frame. For the spatial alignment, the aligned anatomical landmarks were close to the ground truth, and comparable to values reported by others for similar feature-based ultrasound to MRI registration [11].

Our experiments indicate that even with near perfect temporal alignment, a rigid transform between 3D SIFT feature correspondences of temporally aligned frame pairs was not robust in all cases. However, by combining all correspondences across the cardiac cycle in a time-independent manner and utilizing the assumption that inter-frame deformations are negligible, the registration was highly robust and accurate in all cases, despite the deformable contraction of the heart. This usage of correspondences can be justified by the following considerations: Firstly, the extracted features inherently contain some temporal information, as feature scale and appearances tend to fade in and out of existence at certain parts of the cardiac cycle. This can be seen in fig. 2, as the majority of the feature correspondences are found on temporally aligned frames. Secondly, although some spurious matches between unrelated points in the cardiac cycle will arise, they will bear no spatial consistency. Furthermore, as the heart's motion is cyclical, feature correspondences between deformed points during contraction will tend to be canceled out when the heart relaxes. Finally, by employing RANSAC and least squares optimization, we are estimating a rigid transform in a very robust manner.

One of the strengths of the presented method is that it requires no prior information on the spatial relationship between the acquisitions. This means that the method could be used for registering different views, such as apical, parasternal or subcostal, allowing the sonographer maximum freedom to find the best acoustic window.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the Research Council of Norway and the Norwegian Research School in Medical Imaging for their research grants, as well as the support of the Center for Cardiological Innovation. Research reported in this publication was also supported by the National Institutes of Health under award numbers P41EB015898 and R01CA138419.

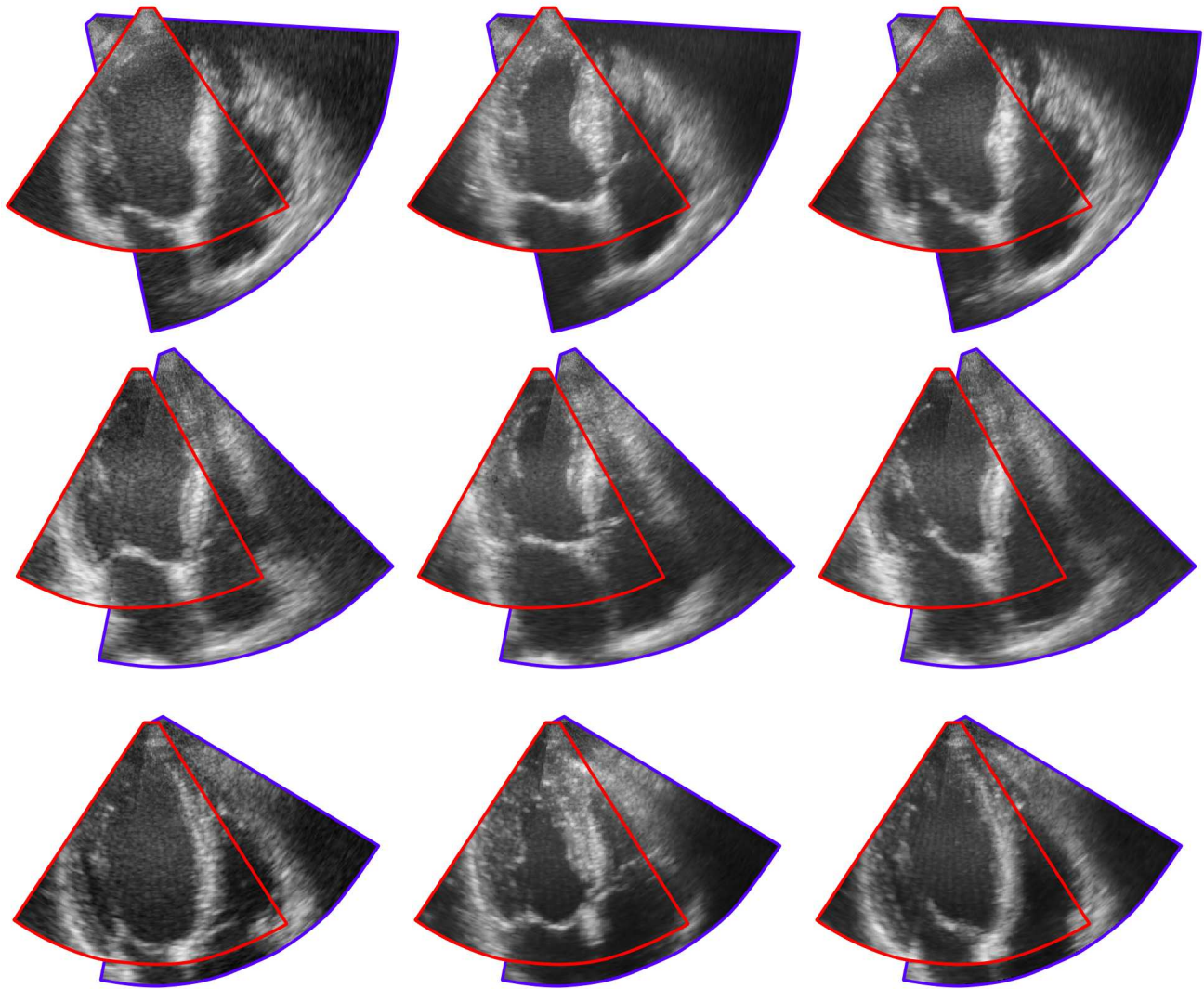


Figure 3. Resulting rigid registration for three sequence pairs across the cardiac cycle. Each row shows three frames from a single case throughout the cardiac cycle.

## References

- [1] Ledesma-Carbayo, M., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., and Unser, M., "Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation," *Medical Imaging, IEEE Transactions on* **24**, 1113–1126 (Sept 2005).
- [2] Elen, A., Choi, H. F., Loeckx, D., Gao, H., Claus, P., Suetens, P., Maes, F., and D'hooge, J., "Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: A feasibility study," *Medical Imaging, IEEE Transactions on* **27**, 1580–1591 (Nov 2008).
- [3] Kiss, G., Barbosa, D., Hristova, K., Crosby, J., Orderud, F., Claus, P., Amundsen, B., Loeckx, D., D'hooge, J., and Torp, H., "Assessment of regional myocardial function using 3d cardiac strain estimation: comparison against conventional echocardiographic assessment," in [*Ultrasonics Symposium (IUS), 2009 IEEE International*], 507–510 (Sept 2009).
- [4] De Craene, M., Piella, G., Camara, O., Duchateau, N., Silva, E., Doltra, A., D'hooge, J., Brugada, J., Sitges, M., and Frangi, A. F., "Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography," *Medical image analysis* **16**, 427–50 (Feb. 2012).
- [5] Danudibroto, A., Gerard, O., Alessandrini, M., Mirea, O., D'hooge, J., and Samset, E., "3d farnebäck optic flow for extended field of view of echocardiography," in [*Functional Imaging and Modeling of the Heart, Lecture Notes in Computer Science* **9126**, 129–136, Springer International Publishing (2015)].
- [6] Ni, D., Qu, Y., Yang, X., Chui, Y., Wong, T.-T., Ho, S., and Heng, P., "Volumetric ultrasound panorama based on 3d sift," in [*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*], Metaxas, D., Axel, L., Fichtinger, G., and Székely, G., eds., *Lecture Notes in Computer Science* **5242**, 52–60, Springer Berlin Heidelberg (2008).
- [7] Mulder, H. W., van Stralen, M., van der Zwaan, H. B., Leung, K. Y. E., Bosch, J. G., and Pluim, J. P. W., "Multiframe registration of real-time three-dimensional echocardiography time series," *Journal of Medical Imaging* **1**(1), 014004 (2014).
- [8] Grau, V., Becher, H., and Noble, J., "Registration of multiview real-time 3-d echocardiographic sequences," *Medical Imaging, IEEE Transactions on* **26**, 1154–1165 (Sept 2007).
- [9] Schneider, R. J., Perrin, D. P., Vasilyev, N. V., Marx, G. R., Del Nido, P. J., and Howe, R. D., "Real-time image-based rigid registration of three-dimensional ultrasound," *Medical image analysis* **16**, 402–14 (Feb. 2012).
- [10] Perperidis, D., Mohiaddin, R. H., and Rueckert, D., "Spatio-temporal free-form registration of cardiac mr image sequences," *Medical Image Analysis* **9**(5), 441 – 456 (2005).
- [11] Zhang, W., Noble, J., and Brady, J., "Spatio-temporal registration of real time 3d ultrasound to cardiovascular mr sequences," in [*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007*], Ayache, N., Ourselin, S., and Maeder, A., eds., *Lecture Notes in Computer Science* **4791**, 343–350, Springer Berlin Heidelberg (2007).
- [12] Valsangiacomo Buechel, E. R. and Mertens, L. L., "Imaging the right heart: the use of integrated multimodality imaging," *European Heart Journal* **33**(8), 949–960 (2012).
- [13] Toews, M. and Wells, W. M., "Efficient and robust model-to-image alignment using 3d scale-invariant features," *Medical image analysis* **17**(3), 271–282 (2013).