

Spatio-temporal Registration of Real Time 3D Ultrasound to Cardiovascular MR Sequences

Weiwei Zhang, J. Alison Noble, and J. Michael Brady

Wolfson Medical Vision Lab, Department of Engineering Science, University of Oxford, Parks Road OX1 3PJ, Oxford, UK
{weiwei,noble,jmb}@robots.ox.ac.uk.

Abstract. We extend our static multimodal nonrigid registration [1] to a spatio-temporal (2D+T) co-registration of a real-time 3D ultrasound and a cardiovascular MR sequence. The motivation for our research is to assist a clinician to automatically fuse the information from multiple imaging modalities for the early diagnosis and therapy of cardiac disease. The deformation field between both sequences is decoupled into spatial and temporal components. Temporal alignment is firstly performed to re-slice both sequences using a differential registration method. Spatial alignment is then carried out between the frames corresponding to the same temporal position. The spatial deformation is modeled by the polyaffine transformation whose anchor points (or control points) are automatically detected and refined by calculating a local mis-match measure based on phase mutual information. The spatial alignment is built in an adaptive multi-scale framework to maximize the phase-based similarity measure by optimizing the parameters of the polyaffine transformation. Results demonstrate that this novel method can yield an accurate registration to particular cardiac regions.

1 Introduction

Cardiovascular disease is one of the world's leading causes of death. The comparison and fusion of information from multiple cardiac image modalities, e.g. cardiovascular magnetic resonance imaging (CMR) and real-time 3D echocardiography (RT3DUS), is of increasing interest in the medical community for physiological understanding and diagnostic purposes.

Most of the current cardiac image registration methods focus on establishing the spatial correspondence between image sequences, while ignoring the temporal correspondence [2]. However, incorporating temporal information into the alignment of multi-modal cardiovascular images can contribute to the study of dynamic properties or motion patterns of the heart. For example, Huang et al. have presented a method for rapid spatio-temporal registration of RT3DUS and CMR images [3]. They first calibrate and synchronize both sequences by integrating electrocardiogram (ECG) and a spatial tracking system to a RT3US machine. These augmented sequences are then manually co-registered to generate a temporal transformation matrix on which the real-time registration step is based. But the dependence of temporal alignment on the ECG signal might be

error-prone due to the sampling errors in the ECG signal. In addition, the limited number of degree of freedoms (DOFs) assumed in that work can not be sufficient to represent the heart as a deformable model in the real case. In other work, Perperidis et al. report two algorithms for spatio-temporal registration of CMR sequences [4]. The transformation model is separated into spatial and temporal components. The first algorithm optimizes the spatial and temporal components simultaneously while the second algorithm optimizes the temporal component before optimizing the spatial component. However, the first algorithm is computationally expensive and the second algorithm can only align limited temporal positions in the cardiac cycle. To our knowledge, spatio-temporal (2D+T) non-rigid registration of RT3DUS and CMR sequences has not been reported before.

In this paper, we extend our static (gated) multi-modal registration method to multi-frame RT3DUS and CMR registration to enable correlation of 3D information between these two modalities [1]. The novelty is to perform the temporal alignment by re-slicing both sequences to contain the same number of frames in order to make them correspond to the same temporal position by using a differential registration technique. The spatial alignment is achieved by a polyaffine transformation framework whose control points are automatically identified by a local mis-match measure using phase information. Therefore the optimization of the proposed method is computationally efficient.

2 Method

We describe the 2D+T version of the method, although it extends naturally to 3D+T. An image sequence $I(x, y, t)$ can be represented as a series of 2D images, where $(x, y) \in \Omega_I$ denotes a point in the 2D image domain and $t \in [0, T - 1]$ denotes the temporal position of a scene and T usually represents the total number of the frames recorded in a cardiac cycle. The goal of our image sequence registration is to find a deformation field that maps each point in the floating image sequence to its corresponding point in the reference image sequence. As in the work of Perperidis et al [4], the spatio-temporal registration problem is solved by decoupling the deformation field \mathbf{D} into the spatial component and the temporal component respectively:

$$\mathbf{D}_{temporal}(t) = t'(t) \quad \text{and} \quad \mathbf{D}_{spatial} = (x'(x, y), y'(x, y)) \quad (1)$$

More precisely, the temporal alignment is performed before the spatial alignment to ensure that each temporal frame in the floating image sequence is registered with the corresponding temporal frame in the reference image sequence. This prevents different image scenes in the same sequence being deformed in different temporal directions. After temporal alignment is complete, the spatial alignment can be taken between both frames that correspond to the same position in the cardiac cycle.

One assumption of the proposed method is that both modality images are in the same spatial coordinate view, e.g. in the short axis (SA) or long axis (LA) view. We use the orientation information acquired during the CMR examination

to bring the RT3D dataset into the same spatial coordinate system as the CMR dataset using commercial software. In the experimental section, every frame in the cardiac sequence is taken from the mid slice of the SA and LA views.

2.1 Spatial Alignment

The heart is undergoing a spatially varying motion during the cardiac cycle. Even at the same phase of the cardiac cycle, multi-modal cardiac images exhibit local differences due to the motion between the examinations and the different ways to acquire and slice the datasets, which are not easily modeled only by a global rigid transformation. We assume the deformation of the ventricles in multi-modal cardiac images is composed of a locally affine transformation. The polyaffine frame work is employed to obtain a globally non-rigid transformation to model the deformation that presents several local behaviours [5]. The idea behind the polyaffine transformation is to weight the sum of local displacements according to a weight function for each arbitrary image region:

$$\mathbf{D}_{spatial}(x, y) = \frac{\sum_i w_i(x, y) \mathbf{D}_i(x, y)}{\sum_i w_i(x, y)} \quad (2)$$

where w denotes the weight function for image region i within a cardiac scene and \mathbf{D} denotes the local displacement in each image region. The weight function is modeled by a Gaussian function that acts as a pre-defined shape for each region and models its influence in the image space. The local displacement modeled by an affine transformation is obtained via an ODE to guarantee invertibility, since all transformations induced by an ODE are reversible [6]. Therefore it can avoid the 'folding' effect that a traditional spline-based transformation tends to produce at high image resolution [7,8].

A local mis-match measure is employed to automatically identify the anchor point (the centre of a local misaligned region) of a polyaffine transformation by removing superfluous degrees of freedom and avoiding the regular grid of control points that increases the computation. Therefore, the polyaffine transformation can effectively adjust these control points to align and correct the spatial difference using less transformation parameters. Adopting the local phase as the image descriptor has been shown to increase the accuracy of non-rigid registration in terms of the information theory based similarity measure [9,10]. Furthermore as the mutual information of local phase can avoid the intensity changes across multi-modalities and makes no assumption of a like-with-like correspondence, it is well-suited for cardiovascular image analysis [11,10]. A local mis-match measure based on the local mutual phase information in a small spatial neighbourhood is used to find the most misaligned region between the US and MR images [12].

2.2 Temporal Alignment - General Principles

Due to the variations in the temporal sampling rate, RT3DUS and CMR image sequences of the same subject are not generally synchronized. This implies that

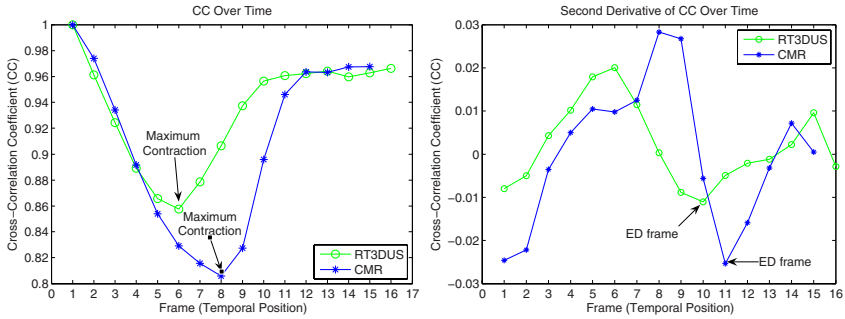


Fig. 1. Cross-correlation coefficient (CC) for MR and US sequences over time and their second derivatives. Left: CC of both MR and US sequences shown together. Right: Using the second derivative (central differences method) of CC to identify the end-diastole frame.

the spatial alignment of the corresponding frames between image sequences may lead to the comparison of frames at different positions in the cardiac cycle. One approach to accommodate this could combine spatial and temporal alignment in one optimization step. However, it has been reported that this combined optimization demands expensive computation and in practice, the floating image scene can be deformed in different temporal directions [4].

In our temporal alignment, we prefer to re-slice the multi-modal cardiac sequences in time to make each pair of frames correspond to the same time position. To this end, we identify key temporal positions in the cardiac cycle based on temporal cross-correlation. For each imaging modality sequence and for each identified segment (phase) of the cardiac cycle, we then estimate the total motion field by summation of the consecutive pairwise estimations of the motion between frames. Each sequence is then re-sliced to contain the same number of frames.

Detection of Key Cardiac Phases. Figure 1 shows how the cross-correlation coefficient (CC) is used to derive temporal information from a cardiac image sequence. This is achieved by calculating the cross-correlation coefficient of each subsequent frame with respect to the initial frame, which is proposed by Perperidis et al [4]. The initial frame is arbitrarily assumed to be the beginning of the cardiac cycle (end-diastole). The maximum contraction frame can be easily identified by finding the minimum value of the cross-correlation coefficient since it has the lowest degree of similarity with the initial frame (frames 6 and 8 respectively for RT3DUS and CMR). Similarly, the end-diastolic frame should have a high degree of similarity with the initial frame, which can be identified by finding the minimum value of the second derivative of the cross-correlation coefficient after the location of maximum contraction (frame 10 and 11 respectively for RT3DUS and CMR).

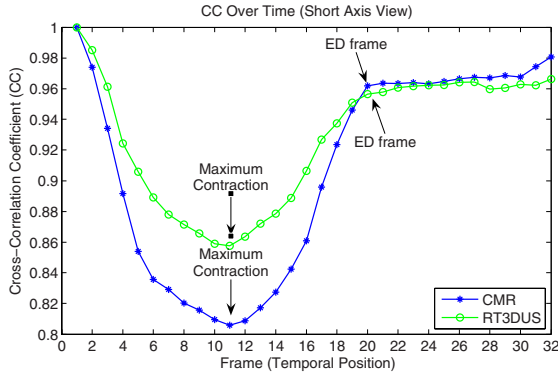


Fig. 2. The result of re-slicing multi-modal cardiac sequences to contain the same number of frames (32 frames) in a cardiac cycle

Re-slicing Cardiac Sequence. Using temporal cross-correlation, the cardiac cycle is now defined into four phases. But as observed in figure 1, the equivalent frame numbers are not aligned to the same point in the cardiac cycle. To align them, the dense estimated motion field is firstly re-sliced between these positions and the local transformation flow is then used to non-linearly interpolate and generate a new frame. Figure 2 shows the CC versus frame number curves of RT3D and CMR sequences sliced to 32 frames across the cardiac cycle.

Key to doing re-slicing is the estimation of the motion field. The differential local affine transformation is employed for estimating the motion field between two consecutive frames [13]. $I(x, y, t)$ and $I(x', y', t-1)$ are denoted as the consecutive frames. The motion field can be modeled by a local affine transformation, assuming the brightness and contrast between frames are constant:

$$I(x, y, t) = I(m_1x + m_2y + m_5, m_3x + m_4y + m_6, t - 1) \quad (3)$$

where m_1, m_2, m_3, m_4 are the linear affine parameters and m_5, m_6 are the translation parameters. A quadratic error function is minimized in order to estimate these transformation parameters. This error function can be approximated using a first order truncated Taylor series expansion. By differentiating with respect to the unknowns, setting the result to zero and solving for \vec{m} , the error function can be estimated analytically:

$$\vec{m} = \left[\sum_{x,y \in \Omega} \vec{c} \vec{c}^T \right]^{-1} \left[\sum_{x,y \in \Omega} \vec{c} k \right] \quad (4)$$

where $\vec{m} = (m_1 \cdots m_6)^T$ and Ω denotes a small spatial neighborhood; and the scalar k and vector \vec{c}^T are given as: $k = I_t + xI_x + yI_y$ and $\vec{c}^T = (xI_x, yI_y, xI_y, \cdots, yI_x, I_x, I_y)$.¹

¹ For notational convenience, the dependencies/parameters of $I_x(\cdot), I_y(\cdot)$ and $I_t(\cdot)$ with respect to t are dropped.

The invertibility of the matrix \vec{m} can be guaranteed by integrating a large enough spatial neighborhood. Moreover, to account for intensity changes and to ensure that model parameters vary smoothly across space, additional parameters that capture a change in contrast and brightness and a smoothness constraint term can be added to equation 3 and the quadratic error function respectively. The algorithm is built in a coarse-to-fine scheme to speed up the computation. Finally, the motion field between the ultrasound frames is estimated by applying the differential technique to the phase image rather than intensity [9]. It should be noted that the registered (or newly-generated) frame is produced by warping the intensity image instead.

When estimating the motion field across the whole cardiac cycle, the total motion field is obtained from the contribution of the partial motion fields of the consecutive pairs over the cardiac phases [14]. Both forward and backward consecutive registrations are performed to reduce registration errors. The total motion field is given by the maximum likelihood of the accumulated motion field from the consecutive pairs, assuming the error distribution of the consecutive registration is identically distributed:

$$\mathbf{M}_{0,t} = \omega_t \mathbf{M}_{0,t}^{forward} + (1 - \omega_t) \mathbf{M}_{0,t}^{backward} \quad (\omega_t = \frac{T-t}{T}) \quad (5)$$

where $\mathbf{M}_{0,t}^{forward} = \mathbf{M}_{0,t-1}^{forward} + \mathbf{M}_{t-1,t}^{forward}$ denotes the accumulated motion field from the forward consecutive pairs and $\mathbf{M}_{0,t}^{backward} = \mathbf{M}_{T,t+1}^{backward} + \mathbf{M}_{t+1,t}^{backward}$ denotes the accumulated motion field from the backward consecutive pairs. For further implementation details see [9].

3 Results

Figure 3 shows the results of the proposed spatio-temporal registration of multi-modal cardiac sequences. Due to space constraints, we show the results at two temporal positions in two different views. Alternatively, the results can be viewed as a movie (<http://www.robots.ox.ac.uk/robots/people/WeiweiZhang.html>).

4 Validation

We follow our previous work to validate the method using expert-identified landmarks. For each pair of the corresponding frames, the anatomical landmarks such as the conjoint of the anterior and anterior septum, the conjoint of the septum and inferior PPM, and the APM, were carefully identified by an experienced clinical expert. The RT3DUS slice identified anatomical landmarks were registered to the CMR image as the reference standard. The measure of the registration accuracy is defined as the root mean square error (RMS) distance of the identified anatomical landmarks between the nonrigidly registered RT3D slice and the reference standard. In this experiment, the registration accuracy were $4.35 \pm 1.25mm$ and $5.23 \pm 1.65mm$ for the SA and LA sequences respectively.

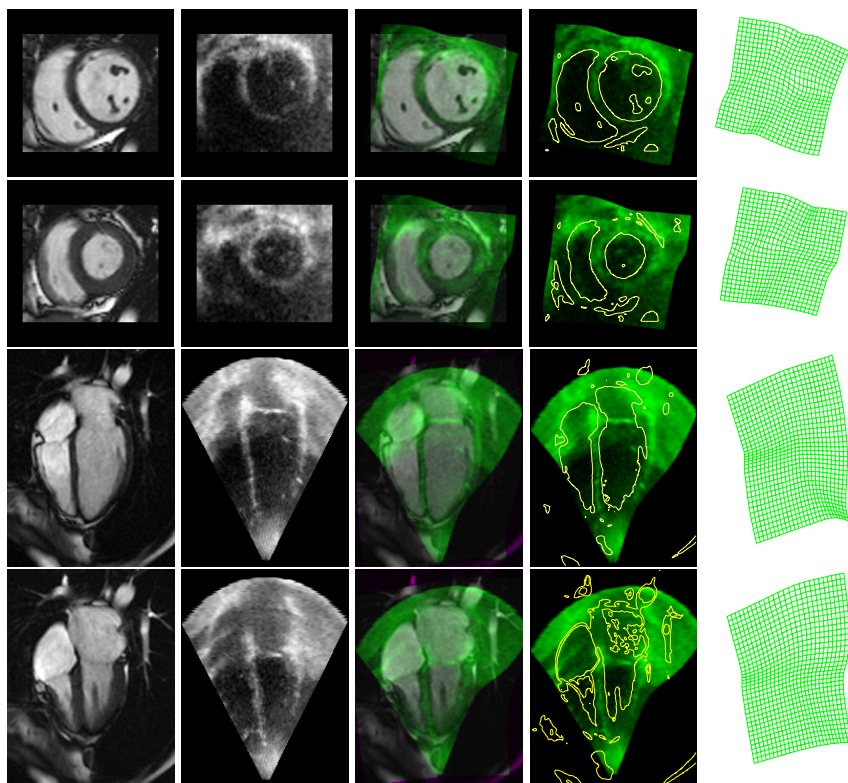


Fig. 3. The result of spatio-temporal registration of RT3DUS and CMR sequences (top two rows in SA view and bottom two rows in LA view). From left to right column: CMR image frames at two temporal positions (the beginning of the cardiac cycle, the peak contraction of the left ventricle); the corresponding RT3DUS frames; the registered RT3DUS frames superimposed to the CMR frame with iso contours; the spatial transformation grids generated by the polyaffine transformation.

5 Discussion

In this paper, a novel method has been presented to solve the problem of non-rigidly registering RT3DUS and MR sequences. The deformation field between both cardiac sequences has been decoupled into spatial and temporal components and each component is optimized separately. The temporal alignment is an important part of our method and is solved by re-slicing both sequences using a differential registration technique. After the temporal alignment, each sequence has the same number of frames and each frame is ensured to correspond to the same temporal position. Therefore our previous static registration method can be extended to multi-frame registration. Since the algorithm only uses the image information, it can reduce the dependence on ECG signal and spatial tracking

system which might be a source of error. A visual comparison from several key cardiac phases has shown that the polyaffine transformation can correct spatial differences in small regions like the myocardium and papillary muscles. In the future, we will investigate the application of RT3DUS/CMR in clinical practice.

Acknowledgements. We thank the Oxford Centre for Clinical MR Research (Dr Matthew Robson) and the Oxford John Radcliffe Hospital Cardiology Department (Dr Harold Becher) for providing the clinical motivation and data.

References

1. Zhang, W.W., Noble, J.A., Brady, J.M.: Adaptive non-rigid registration of real time 3d ultrasound to cardiovascular mr images. In: IPMI 2007. LNCS, vol. 4584, pp. 50–61 (2007)
2. Makela, T.J., et al.: A review of cardiac image registration methods. *IEEE Trans. Med. Imag.* 21, 1001–1021 (2002)
3. Huang, X.S., et al.: Dynamic 3d ultrasound and mr image registration of beating heart. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 171–178. Springer, Heidelberg (2005)
4. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Spatio-temporal free-form registration of cardiac mr image sequences. *Med. Imag. Anal.* 9, 441–456 (2005)
5. Arsigny, V., Pennec, X., Ayache, N.: Polyrigid and polyaffine transformations: A novel geometrical tool to deal with non-rigid deformations - application to the registration of histological slices. *Med. Imag. Anal.* 9, 507–523 (2005)
6. Tenebaum, M., Pollard, H.: *Ordinary differential equations*. Dover (1985)
7. Rueckert, D., et al.: Non-rigid registration using free-form deformation: Application to breast mr images. *IEEE Trans. Med. Imag.* 18, 712–721 (1999)
8. Meyer, C.R., et al.: Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations. *Med. Imag. Anal.* 1, 195–206 (1997)
9. Zhang, W.W.: *Multimodal Cardiovascular Image Analysis Using Phase Information*. PhD thesis, University of Oxford, PhD Thesis forthcoming (2007)
10. Mellor, M., Brady, J.M.: Phase mutual information as a similarity measure for registration. *Med. Imag. Anal.* 9, 330–343 (2005)
11. Felsberg, M., Sommer, G.: A new extension of linear signal processing for estimating local properties and detecting features. In: *Proc. DAGM*, 195–202 (2000)
12. Park, H., Bland, P.H., Brock, K.K., Meyer, C.R.: Adaptive registration using local information measures. *Med. Imag. Anal.* 8, 465–473 (2004)
13. Periaswamy, S., Farid, H.: Elastic registration in the presence of intensity variations. *IEEE Trans. Med. Imag.* 7, 865–874 (2003)
14. Ledesma-Carbayo, M., et al.: Cardiac motion analysis from ultrasound sequences using nonrigid registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 889–896. Springer, Heidelberg (2001)