

Validation of dynamic heart models obtained using non-linear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries

Marcin Wierzbicki ^{a,*}, Maria Drangova ^{a,b}, Gerard Guiraudon ^{a,c}, Terry Peters ^{a,b}

^a *Imaging Research Laboratories, Robarts Research Institute, 100 Perth Drive, P.O. Box 5015, London, ON, Canada N6A 5K8*

^b *Department of Medical Biophysics, The University of Western Ontario, London, ON, Canada N6A 5C1*

^c *Canadian Surgical Technologies and Advanced Robotics (CSTAR), 339 Windermere Road, London, ON, Canada N6A 5A5*

Available online 17 July 2004

Abstract

Current minimally invasive techniques for beating heart surgery are associated with three major limitations: the shortage of realistic and safe training methods, the process of selecting port locations for optimal target coverage from X-rays and angiograms, and the sole use of the endoscope for instrument navigation in a dynamic and confined 3D environment. To supplement the current surgery training, planning and guidance methods, we continue to develop our Virtual Cardiac Surgery Planning environment (VCSP) – a virtual reality, patient-specific, thoracic cavity model derived from 3D pre-procedural images. In this work, we create and validate dynamic models of the heart and its components. A static model is first generated by segmenting one of the image frames in a given 4D data set. The dynamics of this model are then extracted from the remaining image frames using a non-linear, intensity-based registration algorithm with a choice of six different similarity metrics. The algorithm is validated on an artificial CT image set created using an excised porcine heart, on CT images of canine subjects, and on MR images of human volunteers. We found that with the appropriate choice of similarity metric, our algorithm extracts the motion of the epicardial surface in CT images, or of the myocardium, right atrium, right ventricle, aorta, left atrium, pulmonary arteries, vena cava and epicardial surface in MR images, with a root mean square error in the 1 mm range. These results indicate that our method of modeling the motion of the heart is easily adaptable and sufficiently accurate to meet the requirements for reliable cardiac surgery training, planning, and guidance. © 2004 Elsevier B.V. All rights reserved.

Keywords: Deformable models; Non-linear image registration; Registration validation; Pre-interventional planning; Image-guided surgery

1. Introduction

Coronary artery disease is the most common cause of death in the developed world. Although not as deadly, atrial fibrillation is also a dangerous condition, accounting for approximately a five to sevenfold increase in stroke risk and a doubling of mortality from cardiovascular disease (Gillinov et al., 2002). For more serious cases of these and similar heart diseases, cardiac surgery remains the only option for the patient. For coronary artery disease, the surgery aims to bypass blockages in the coronary

arteries with the aid of grafts, while for atrial fibrillation, the electrical pathways causing the condition are isolated and neutralized by strategic scaring of heart tissue. Access to the heart muscle is therefore required, and is traditionally achieved through a median sternotomy, arresting the heart, and placing it on cardiopulmonary bypass. These highly invasive methods result in additional patient trauma, leading to longer hospital stays and costs compared to less invasive approaches (King et al., 1997). Due to the aging population and the resulting burden on the health care system, it is becoming increasingly important to find treatments for heart disease that are less invasive and more successful than the traditional approaches.

Recent research in endoscopically-aided, port-access, bypass surgeries is showing promising results (Stevens

* Corresponding author. Tel.: +1-519-663-5777x34136; fax: +1-519-663-3900.

E-mail address: mwierzb@imaging.robarts.ca (M. Wierzbicki).

et al., 1996). In these minimally invasive procedures, small incisions (ports) are made in the chest wall to access the heart without requiring a sternotomy. The heart can then be treated manually with long-handled instruments, or with the aid of robotic systems, designed to reduce hand tremor and allow smaller and more appropriate manipulations (Boyd et al., 2000b; Loulmet et al., 2003). Theoretically, this innovative approach to cardiac surgery is a tremendous improvement over the open chest technique, since the surgical target can be treated effectively with less patient trauma. In practice however, minimally invasive cardiac surgery (MICS) is a challenging procedure that requires the surgeon to develop a new and complex set of skills.

Difficulties in maneuvering surgical tools aside, the main challenge in improving the effectiveness of MICS lies in proper visualization of the three-dimensional (3D) operating environment. In spite of the sophistication of the robotic systems used in these procedures, critical tasks such as planning and guidance are often performed entirely with 2D images. The intrinsically 3D port locations are often determined based on 2D angiograms and chest X-rays, while the entire procedure is guided solely through a small field of view video endoscope (Herzog et al., 2003). Even if stereoscopic endoscopy, combined with wide-angle panoramic views are used for navigation, problems can occur due to vision obstruction from water condensation, blood, and tissues. The lack of 3D surgery planning and guidance can lead to improper patient selection, sub-optimal port placement, longer procedures, and increased risks to patients (Boyd et al., 2000a; Herzog et al., 2003).

To address the visual limitations of MICS, we have previously developed a prototype Virtual Cardiac Surgery Planning platform (VCSP) (Chiu et al., 2000). Although currently in the initial stages of development, this project has already yielded promising results. In VCSP, pre-operative computed tomography (CT) and magnetic resonance (MR) images are segmented and stereoscopically rendered in 3D. Models of the patient's ribs and chest wall are created from CT images using the Marching Cubes algorithm (Lorenson and Cline, 1987). A dynamic, 3D, coronary artery model derived from static 3D images and dynamic angiograms (Lehmann et al., 2002) has also been incorporated, and work continues on developing methods to extract patient-specific coronary artery anatomy. Initial models of the patient-specific epicardial surface have been extracted from canine CT images (Wierzbicki and Peters, 2003). We can also overlay the real video endoscopic image with the dynamic virtual environment for the purpose of enlarging the field of view of the endoscope and to allow visualization "through" the tissue (Szpala et al., 2003). Finally, the system can update the position and orientation of virtual surgical tools in VCSP by optically tracking the real instruments. This aspect of VCSP is especially important in

surgery planning, where the dependence of inter-costal port locations on the geometry of the simulated surgical tools can be observed and optimized.

One area for VCSP improvement, and the focus of this work, is the development of a method to generate patient-specific, dynamic models of the heart and its components. The addition of such models will make the current prototype VCSP clinically useful for surgery training, planning and navigation. The method must be easily adaptable to different hearts, and be largely automated. The resulting models should also be readily adjustable during a procedure to correct for any changes between the pre- and intra-operative situations. Finally, the resulting models describing the shape and dynamics of the heart must be accurate to within 1 mm root mean square (RMS) error so that it is possible to accurately merge them with models of other structures of interest, such as the relatively small coronary arteries ($4 \pm 2 \text{ mm}^2$ cross-section in women (Kucher et al., 2001)).

Previous heart modeling research has concentrated primarily on the left ventricle because of its importance in the assessment of function (Frangi et al., 2001). A notable exception is the work by Coste-Manière et al. (2003), describing a visualization system similar to VCSP. In their approach, pre-operative CT image segmentation along with 3D reconstruction are used to create static models of the patient's ribs, chest wall and the epicardial surface. These models are then used to automatically calculate the optimal port locations, depending on the target and entry chosen by the surgeon. However, to the best of our knowledge, the significant effect of heart motion is not considered in the current implementation of the epicardial surface model. Automatic segmentation alone can be used to study or display heart dynamics when combined with 4D data acquisition using and ECG-gated protocol. Mitchell et al. (2002) for example, performed frame-by-frame segmentation of cardiac image data to generate models of the left ventricle at different time-points in the cardiac cycle. Although effective on single 3D images, the segmentation approach fails to take advantage of the correspondence of adjacent time-points in a 4D data set. In addition, such methods are usually specific to a particular area of the heart due to their dependence on image intensity. Finally, each image set is usually treated separately so that previous results are not employed to make these methods more adaptable.

Heart parameterization over time and space can be a more adaptable modeling method than the segmentation technique. Essentially, the geometry of the heart (most often the left ventricle) and its evolution over time is assumed to be of a basic shape defined by a series of parameters and equations. Declerck et al. (1998) for example, simplified the geometry of the left ventricle to minimize the number of parameters describing its motion over the cardiac cycle. Such starting models could then be "customized" to fit different patients. These

models are suitable for simpler structures such as the left ventricle, but they are difficult to apply accurately in the entire heart, due to the complex motion and geometry.

Assumptions about the shape of the heart are not made when voxel-based, non-linear image registration is used to derive the 3D motion of the heart between two image frames. This motion information can then be used to propagate a static model (created using a single segmentation) through the cardiac cycle to generate a dynamic heart model. An example of such work is given by Lorenzo-Valdés et al. (2002), where the myocardium, left ventricle, and right ventricle are segmented from 4D MR images. Such an approach satisfies the criteria necessary for a dynamic model generation scheme for planning and guidance of MICS. Outside of the creation of the static model, the animation phase is entirely automatic, and the method can be easily adapted to model any part of the heart (depending on the static model used). In addition, the resulting models can be readily adjusted during a procedure with the same type of non-linear registration method used to animate them.

In this manuscript, we present a technique to generate easily adaptable models of the heart using a non-linear registration algorithm. First, a patient-specific, static model of the heart is created by segmenting the end-diastolic (ED) image of a 4D data set acquired with ECG-gating. The model is then animated using motion information obtained by registering the ED image to all remaining time-frames in the 4D data set. The registration is based on a free-form deformation (FFD) framework, where a mesh of deformation points is manipulated to optimize a similarity metric (SM), subject to a smoothness constraint. The main purpose of this work, is a comprehensive validation of the methodology and the resulting model dynamics, in addition to identifying the optimal SM for the registration algorithm. We approach the difficult validation problem with a series of experiments, beginning with registration of a specialized data set obtained from an excised porcine heart. This experiment is important not only because a gold standard is available, but also because it confirms the approach used later to validate the method on in vivo canine and human data sets.

2. Cardiac models

In the following subsections we discuss how a dynamic surface model of the heart epicardium or endocardium is created. The process begins with an ECG-gated, 4D image acquisition. The resulting data set consists of several 3D images, each depicting the heart at different time-points in the cardiac cycle. The motion information is extracted from the 4D data using a non-linear registration algorithm described in detail in Section 3. The static geometry of the heart is obtained from one of the 3D image frames using segmentation.

2.1. Static

The first step in generating a dynamic heart model from pre-procedural CT or MR data is the creation of the static anatomy model. CT offers higher resolution images and shorter acquisition times, at the cost of radiation exposure, and the requirement for contrast agent injections to enhance internal heart anatomy. MR on the other hand, can produce images with good soft tissue contrast without the disadvantages of CT. Unfortunately, the acquisition of high-resolution 4D MR data is a difficult and time-consuming process.

Regardless of the imaging modality used, the ED image from the complete 4D data set was chosen as the reference. Manual segmentation of this image was performed slice-by-slice using a “paintbrush” technique to outline the anatomy of interest. The ED image was chosen to create the static models because heart motion is minimal during this stage; therefore, this frame is the least susceptible to motion artefacts and was the easiest to segment. The binary data resulting from the segmentation were filtered with a spherical Gaussian filter to smooth uneven edges. The smoothed results were then passed through the Marching Cubes algorithm to create the static surface model, consisting of a series of triangles defined by 3D vertices. Finally, the resulting surfaces were decimated to reduce the number of triangles and increase rendering speed (Schroeder et al., 1992). Fig. 1 shows some examples of static surface models obtained from canine CT, and human MR images. 2D surfaces were used to represent the 3D heart anatomy due to improved displaying efficiency and easier interaction versus volumetric models. Stereoscopic display was used to further enhance the third dimension.

For the validation experiments discussed in this manuscript only the epicardial surface was outlined in CT images due to poor contrast inside the heart in our

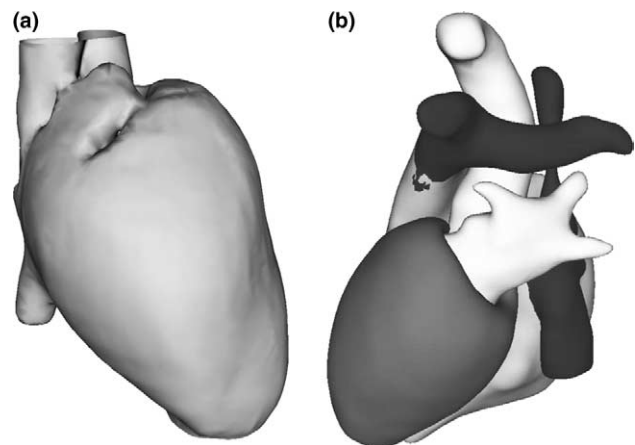


Fig. 1. (a) Epicardial surface model of a canine heart obtained from a CT image at ED. (b) Myocardium, right atrium and right ventricle, aorta and left atrium, vena cava and pulmonary arteries, and epicardial surface models of the human heart obtained from an MR image at ED.

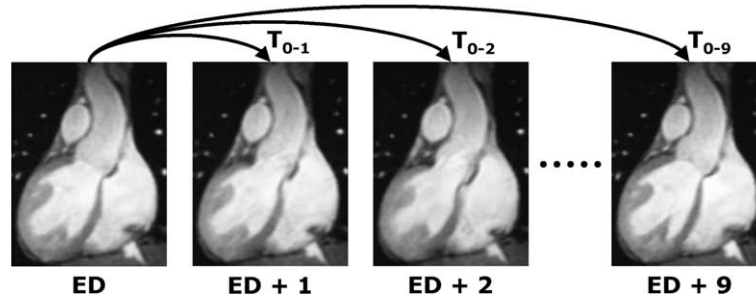


Fig. 2. Registration scheme for extraction of the transformations between the ED time phase, and all the remaining phases in a 4D data set. ED is the end-diastole image, and ED + N is the N th image in the cardiac cycle. T_{0-N} represents the non-linear transformation required to map the ED image to the N th image.

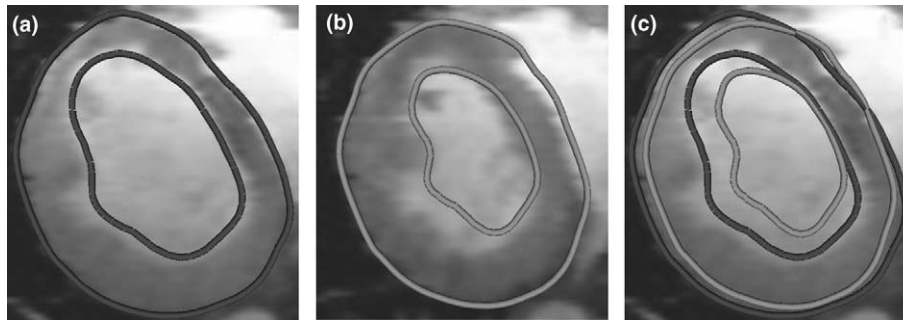


Fig. 3. Cross-sections through the myocardial model at the ED and ES time-points in the cardiac cycle, overlaid on their corresponding images: (a) ED image and model; (b) ES image and model; (c) both ED and ES models on the ED image.

images. For MR data, the heart was divided into five separate models: the myocardium, right ventricle and right atrium, aorta and left atrium, vena cava and pulmonary arteries, and the epicardial surface.

2.2. Dynamic

The static model obtained using image segmentation represents only the ED time point in the cardiac cycle. This model was animated by non-linearly registering the ED image to each of the remaining images in the 4D data set (Fig. 2). Each registration resulted in a discrete, 3D, displacement vector field (T_{0-N}) – the transformation required to deform the ED (0) model to match it to one of the remaining images (N). Deformation of the ED model was achieved by translating all of the triangle vertices according to a continuous version of T_{0-N} , obtained through interpolation. The resulting group of $N + 1$ surfaces was then rendered sequentially in 3D to portray the dynamics of the heart. As an example, Fig. 3 shows cross-sections through the myocardial surface model at the ED and end-systole (ES) time-points in the cardiac cycle.

3. Non-linear registration algorithm

This section describes the registration algorithm used to extract the motion of the heart between two image

frames of a 4D data set. This description is broken into several subsections, beginning with a discussion of the non-linear transformation model, then a description of the solution method, and finally a description of the fundamental image properties and mathematics used in obtaining the solution.

3.1. Transformation model

Non-linear registration is required to track the motion of the heart since global registration methods (e.g., linear or affine) are insufficient to fully characterize heart dynamics. It is possible to parameterize the required non-linear transformations, but this leads to over-simplifications, resulting in lower accuracy and applicability of the motion extraction algorithm. We therefore use a more general, FFD approach to calculate the motion of the heart. In this framework, a 3D grid of points (nodes) is overlaid on the image to be registered (source image). Each node is then assigned a displacement vector estimating the deformation required in the surrounding region to match this part of the source image to the target image. Therefore, image registration in the FFD framework is a problem of estimating a total of $3M$ degrees of freedom (for M nodes and 3D vectors). Theoretically, increasing M results in a higher accuracy transformation, at the cost of computational time. Fig. 4 shows an example of an overlay of a source image with

the corresponding FFD grid of vectors. The actual directions and lengths of the vectors are determined as described in Section 3.2.

Two interpolation steps are required to deform a source image once the grid of vectors has been estimated. First, because the FFD grid has fewer nodes than the source image has voxels, the grid must be interpolated so that each voxel in the source image can be assigned a translation vector. For this purpose, we investigated nearest neighbor, linear, cubic convolution, and cubic b-spline interpolation. Cubic interpolation has the advantage over b-spline interpolation in that displacement vectors estimated by the algorithm at the nodes themselves are not altered. This is very important since these estimates are the only “true” deformations. On the other hand, cubic convolution interpolation is not as smooth as b-spline interpolation, especially around quick changes in the data where uncontrolled “ringing” can occur. We found that the higher order methods were not the optimal choice for our algorithm, and that linear interpolation of the grid gave the best results. Linear interpolation does not alter the data at the nodes, nor does it cause ringing artefacts. In addition, linear interpolation is computationally inexpensive compared with the higher order methods. One concern with linear interpolation is its ability to represent non-linear deformations accurately, especially with a low-resolution deformation grid. Theoretically, investing more time into calculating denser deformation grids can eliminate this effect.

Once the translation of each voxel in the source image is known, a separate interpolation must be performed on the image intensity data. It is important to choose a computationally efficient method since it occurs in the “inner-most loop” of the registration algorithm. As a result, we again chose linear interpolation, due to its computational efficiency, and due to the factors mentioned above for grid interpolation.

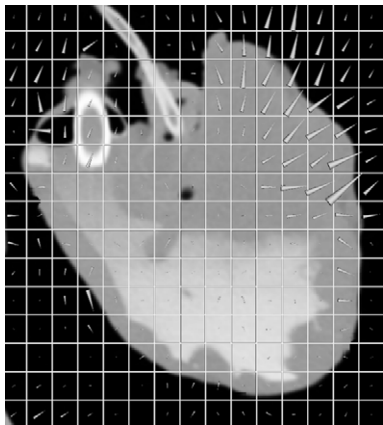


Fig. 4. A slice from a helical CT image of an excised porcine heart phantom overlaid with the corresponding slice of a vector field calculated by the registration algorithm. Vectors are represented by cones, and have been lengthened by a factor of two for better visibility.

3.2. Optimization

The estimation of the complete deformation field is an optimization problem, where vectors in the FFD grid are determined so that the source and target images are more similar after the transformation. In our implementation, each vector is estimated independently based on a series of factors. The primary requirement is that the vector must improve the local alignment between the source and target images. This is determined by calculating a SM, which compares the relationship of voxel intensities in the newly deformed source image and the target image (see Section 3.3 on SMs). Theoretically, the more aligned the images are, the higher the value of the SM. This is not the case in specific situations, where the SM can be maximized even if the registration is not correct (as in two rotated, homogenous spheres). Also, image noise, artefacts and other imaging imperfections can make it difficult to align two images accurately using the SM alone. Additional regularization terms (see Section 3.4) are therefore required in the registration function so that imaging errors do not significantly influence the image alignment process. For example, a long displacement vector in the FFD grid may increase the SM significantly, but may be determined to be highly unlikely by the regularization term. The process of finding the vector displacement is therefore an optimization of a function that balances the SM and the transformation probability terms. For each node in the FFD grid we find the displacement vector by minimizing the following function:

$$C(x, y, z) = -\text{SM}(V, I_S(x, y, z), I_T) + \alpha \cdot \text{BE}(x, y, z), \quad (1)$$

where C is the cost associated with the 3D displacement vector (x, y, z) , SM is the similarity of the source image intensities I_S translated by (x, y, z) and target image intensities I_T calculated over the sub-volume V centered on the current FFD node, and BE is the value of the regularization term, weighed by the constant α .

The overall optimization itself is performed using the downhill simplex method (Nelder and Mead, 1965). A simplex is a geometric figure represented by $R + 1$ points in R -dimensional space – a tetrahedron in our case, since there are three parameters for each displacement vector to be estimated. During minimization, the simplex is iteratively deformed in the parameter space according to the function values at each of the points. Convergence is reached when all points of the simplex are within a specified tolerance of each other. We chose the simplex over the other optimization methods because it considers all parameters simultaneously, does not require the computation of gradients, and is simple to implement. The simplex method was shown to be as accurate, slightly less precise, but significantly more computationally efficient than other optimization algorithms in rigid-body registration using normalized mutual information (Maes et al., 1999).

The overall optimization process is divided into three main stages, each with an increasingly dense grid of vectors. At each stage, the vector grid is refined and applied to the source image iteratively, allowing the source to slowly progress towards the optimal registration. After a given iteration, the improvement in global image alignment is measured using the SM and the average vector displacement magnitude is calculated. If the average displacement is less than a specified constant, the algorithm proceeds to the next stage where a higher density grid is used. Also, if image alignment was reduced, the algorithm “undoes” the last iteration and proceeds to the next stage with the previous result. Between stages the FFD grid is subdivided using linear interpolation. Therefore, if the initial grid in the first stage has $M_x \cdot M_y \cdot M_z$ nodes, the final FFD grid, in the third stage will contain $4M_x \cdot 4M_y \cdot 4M_z$ nodes. This multi-grid approach reduces the probability of incorrect optimization by giving the algorithm roughly aligned images to deform in subsequent iterations and stages. In order to maintain higher accuracy, source and target image resolution is not changed, eliminating the need for an extra interpolation step.

3.3. Similarity metrics

Several voxel-based similarity measures have been previously proposed in the literature to give a numerical estimate of image alignment (Hill et al., 2001). In our algorithm we implemented a total of six SMs for optimization in Eq. (1). These metrics can be subdivided into direct intensity-based and information theoretic groups. The direct-intensity metrics are: mean absolute difference (MAD), mean squared difference (MSD) and normalized cross correlation (zero-mean definition) (NCC). The information theoretic metrics are: mutual information (MI), normalized mutual information (NMI), and entropy correlation coefficient (ECC). These six SMs can be expressed mathematically as:

$$\text{MAD} = \frac{1}{V} \sum_V |I_S - I_T|, \quad (2)$$

$$\text{MSD} = \frac{1}{V} \sum_V [I_S - I_T]^2, \quad (3)$$

$$\text{NCC} = \frac{\sum_V [I_S \cdot I_T]}{\sqrt{\sum_V I_S^2 \cdot \sum_V I_T^2}}, \quad (4)$$

$$\text{MI} = E_S(V) + E_T(V) - E_{ST}(V), \quad (5)$$

$$\text{NMI} = \frac{E_S(V) + E_T(V)}{2 \cdot E_{ST}(V)}, \quad (6)$$

$$\text{ECC} = \sqrt{2 \left(1 - \frac{2E_{ST}(V)}{E_S(V) + E_T(V)} \right)}, \quad (7)$$

where the intensities in the source and target images are I_S and I_T , respectively, the marginal Shannon entropies of the source and target images are E_S and E_T , the joint entropy is E_{ST} , and the sub-volume of image data under consideration is V .

The MAD measure is well suited to intra-subject, intra-modality registration, since images of the same subject obtained at different times are very similar already. Any changes in the object’s shape or location can be deduced by looking at the intensity differences. Strictly speaking, MAD is not a *similarity* metric; rather it measures the *difference* in alignment of two images. Its sign is therefore reversed in Eq. (1).

Like MAD, MSD measures image difference rather than similarity. It is the ideal metric to use, if after registration, the images differ only by Gaussian noise (Wells III et al., 1996). Unfortunately, this is not the case in MR magnitude images, where the noise is Rician distributed (Sijbers et al., 1998). The MSD metric can be biased by image areas with a small number of voxels that differ in intensity by a large amount. This can make MSD more prone to error from imaging artifacts than the MAD measure (Hill et al., 2001).

The NCC metric is optimal if the relationship between intensities in the two images is linear, although this is rarely the case in CT and MR imaging (Hill et al., 2001). NCC is different from MAD or MSD in that it measures image similarity. The NCC value is always between 0 and 1, which makes it less dependent on the difference in intensities than MAD or MSD. This makes NCC easier to use with a variety of images, especially when regularization terms are calculated and scaled to match the SM.

The benefits of direct intensity-based metrics are easy implementation, computational efficiency, and satisfactory results in intra-subject, intra-modality registration. These metrics, however, cannot properly align significantly different images. This might be the case when two images have completely different contrast for the same anatomy. In these situations it is more appropriate to use the information theoretic metrics such as MI. Qualitatively, the MI of two images measures how well one of the images explains the other, and so MI is maximized when the two images are perfectly registered. Other work has shown that the MI measure can be disproportionately influenced by the amount of overlap between the source and target image (Studholme et al., 1999). Specifically, this can occur when the images have a large background area (e.g., lung tissue in cardiac images). The NMI metric has been proposed as a solution to this problem. As defined in Eq. (6), NMI ranges from 0.5 to 1.0, with 0.5 indicating that the source and target are completely misaligned, and 1.0 indicating perfect registration. Very similar to NMI is the ECC similarity metric. While the square root is not necessary, it magnifies variations near zero and smoothes variations near unity.

As a result, different images appear even less similar with ECC than NMI, and almost the same images more similar with ECC than NMI (Astola and Virtanen, 1982). Therefore, ECC can be an efficient metric in image registration, minimizing the computational effort of improving an already acceptable registration.

3.4. Regularization

As shown in Eq. (1), our algorithm uses a regularization term to avoid local minima in the registration function. This term originates from the 2D bending energy of a thin metal plate, and is controlled by the coefficient α . If a thin metal plate is defined from 0 to X and 0 to Y in the xy plane, and undergoes a small deformation z in the perpendicular direction, the bending energy is (Bookstein, 1989)

$$\text{BE}(z) \propto \int_0^X \int_0^Y \left[\left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 \right] dx dy. \quad (8)$$

In 3D, the counterpart of the thin-plate is the FFD grid (defined from 0 to X , 0 to Y and 0 to Z). The small deformation z now becomes a short vector \vec{r} . The “bending energy” in 3D is then

$$\begin{aligned} \text{BE}(\vec{r}) \propto & \int_0^X \int_0^Y \int_0^Z \left[\left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial x^2} \right)^2 \right. \\ & + \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial z^2} \right)^2 \\ & + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial x \partial y} \right)^2 + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial x \partial z} \right)^2 \\ & \left. + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial y \partial z} \right)^2 \right] dx dy dz, \quad (9) \end{aligned}$$

where $\text{FFD}(\vec{r})$ is the FFD grid with vector \vec{r} assigned to the currently considered node.

The 3D bending energy is maximized when vectors inside the FFD are long and point in opposing directions, and minimized (equal to 0) if all the vectors are identical. In other words, a FFD with smooth variations in vectors at adjacent nodes has a smaller BE than a FFD with sharp variations. Therefore, by minimizing a registration function that includes the BE term, we ensure that vectors chosen by the optimization not only increase the SM, but also generate a smooth transformation.

Eq. (9) shows that the BE is calculated by inserting a choice of \vec{r} into the currently optimized node of the FFD grid. The remaining vectors, also used in the calculation, are set to the results obtained in the *previous* iteration (for the first iteration, the previous FFD is composed entirely of 0 vectors). As a result, the current vector will

be considered more favorably (will give lower BE) if it points in a similar direction to the surrounding vectors from the previous iteration. This can be beneficial if these vectors were accurate, but can cause serious problems if they were not. In this case, the algorithm will find it difficult to change the direction of the vector field once it has been initialized. To avoid this problem, we implement the BE term in our registration function by always setting the vectors not being optimized to zero. This not only removes the dependence of the current vector choice on the previous iterations, but also is much less computationally expensive, since only the nearest neighbors of the current node have to be considered. The downside is that the final FFD grid is not as globally smooth. This is not necessarily significant because we are registering one time-frame of a 4D data set to all the other frames (see Fig. 2), instead of registering the images serially frame-to-frame. So, whereas the transformation between two adjacent frames of the heart is expected to be smooth (real heart motion is smooth), the transformation between non-adjacent time-frames need not be.

4. Validation methods

Validation of non-linear registration methods is difficult since a gold standard is not easily attainable. Researchers are forced to use a variety of surrogates, from simplified computer simulated images, to difficult experiments involving implantation of fiducials into the subject (Woods et al., 1998a,b). The heart poses additional validation problems due to a lack of easily identifiable homologous points (Mäkelä et al., 2002). Yet validation is arguably the most important aspect of designing algorithms for direct clinical use. In this section, we focus on three experiments designed to validate our method of animating a static heart model. In the first experiment, we validate the algorithm on the epicardial surface of an excised porcine heart. In the remaining experiments, we use a measure of consistency to estimate the accuracy of the algorithm in modeling the dynamic epicardial surface in CT images, and various components of the heart in MR images.

4.1. Excised porcine heart

For phantom validation studies it is important that the phantom closely represents the real anatomy and function of the system being studied. Our “phantom” is a porcine heart, excised from an animal that was sacrificed in the ED phase of the cardiac cycle. The heart was cannulated through the ascending aorta, and the pulmonary trunk while all remaining vessels were sutured closed. Two IV bags were attached to these openings, each containing iodine contrast agent diluted

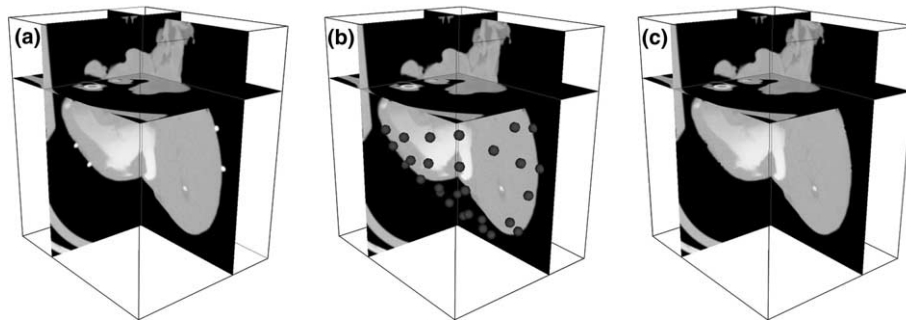


Fig. 5. Three orthogonal planes through a helical CT volume of the excised porcine heart phantom at ED ($N = 5$): (a) original image (notice the high intensity signal from the Teflon spheres); (b) image with overlaid spheres at the fiducial locations; (c) image after fiducial removal.

to 7.5 mg/ml. Cyanoacrylate (superglue) adhesive was used to attach 58 Teflon spheres (0.32 cm diameter) to the heart surface with approximately uniform spacing. The heart was then placed in a plastic container and scanned six times using a GE LightSpeed helical CT scanner (18.6 cm DFOV, 80 kVp, 80 mA). Between each of the six scans, the IV bags were elevated for a short period to force contrast agent into the heart. All images were reconstructed with a voxel dimension of $0.363 \times 0.363 \times 0.63$ mm. The resulting set of six volumes represents the excised porcine heart at different fill levels (image $N = 0$ is minimal fill, image $N = 5$ is maximal fill), simulating the motion of the heart from ED ($N = 5$) to ES ($N = 0$). This “motion” is non-linear, and can be characterized frame to frame due to the presence of fiducial markers. Panel (a) in Fig. 5 shows the CT volume obtained at maximal fill.

The images were prepared for processing as follows: first, the bright voxels from the contrast agent that accumulated at the bottom of the plastic container, and was touching the heart in images 4 and 5, were manually removed through segmentation. This was done to ensure that the appearance of the general environment around the heart is similar between image frames. Second, the locations of the Teflon fiducial markers were determined in each of the six volumes. Due to the high attenuation coefficient of Teflon, simple thresholding at 876 HU resulted in images with only several voxels per fiducial marker (the heart itself, including the contrast agent, was completely removed). The location of a given fiducial was then determined by calculating the centroid of each voxel cluster. All 58 fiducials were located using this approach in each of the CT volumes. Fiducial locations in the ED image can be seen in panel (b) of Fig. 5.

The final preparation step was to remove the image fiducials to prevent the registration algorithm from being biased by these well-defined areas. This was achieved by replacing voxels in a radius of 0.16 cm (radius of the Teflon sphere marker) around a given fiducial with voxels randomly chosen from the background. Residual

marker-related voxels were removed by eroding and dilating a small volume around each marker location. An example of an image after fiducial marker removal is shown in panel (c) of Fig. 5.

In order to validate our algorithm, we registered the prepared image 5 to one of the other images and then compared the locations of the landmarks. The first step is a linear registration using a simple, NMI based, downhill simplex algorithm with six degrees of freedom. Although not necessary in vivo, this pre-registration step is required in our experiment to correct for the global “sagging” of the heart caused by filling it with contrast agent. The resulting image was then registered to the target once again, this time using the non-linear algorithm. We determined the accuracy of the registration by calculating the RMS Euclidian distance between the landmark locations in the source image ($N = 5$) after linear transformation and non-linear deformation, and the landmark locations in the target image ($N = 0, 1, 2, 3, 4$). In this experiment then, registration from image 5 to 0 represents the registration of image at ED–ES in an in vivo data set.

This process was repeated with down-sampled versions of the images to study the effect of lower image resolution on registration accuracy. Two slices from each of the prepared images were averaged together to define the intensities for a thicker slice. As a result the image voxel size was increased from $0.35 \times 0.35 \times 0.63$ to $0.35 \times 0.35 \times 1.25$ mm³, and the signal-to-noise ratio (SNR) was increased by $\sim\sqrt{2}$ (two averages). The new voxel dimensions are the same as those in the canine data described in the following section.

4.2. Canine heart

Images of normal canine subjects, obtained for a separate study approved by the University of Western Ontario animal ethics board, were chosen to validate the algorithm’s ability to model the epicardial surface. Images were acquired using an eight-slice GE LightSpeed helical CT scanner with retrospective ECG gating and 120 kVp, 200 mA X-rays. The animals were

paralyzed and artificially ventilated to eliminate breathing motion artefacts during the 30 s scan time. Immediately before scanning, a bolus of iodine contrast agent (300 mg/ml) was injected intra-venously to provide additional contrast inside the heart. Each of the resulting 4D data sets consist of ten 3D CT images of the thorax at equally spaced time-points in the cardiac cycle reconstructed with a voxel size of $0.35 \times 0.35 \times 1.25 \text{ mm}^3$.

Because the true cardiac motion was unavailable, we used measurements of consistency (Holden et al., 2000) to validate our algorithm. Fig. 6 shows three closed registration loops, each consisting of three transformations (T), for a given 4D data set. For each loop, all three transformations are applied in turn to a starting surface S_N to obtain S'_N , the same surface after a complete loop of transformations. Ideally, S_N and S'_N should correspond exactly. The error in the complete loop can therefore be calculated as the RMS of the Euclidian distance between corresponding points on the two surfaces. No extra calculation is required to determine which points correspond, since both surfaces are composed of exactly the same triangles and vertices. Assuming that the error is normally distributed and uncorrelated, the error associated with each registration is $\text{RMS}/\sqrt{3}$, since there are three transformations per loop. We chose to register images approximately three phases apart to best approximate the method outlined in Fig. 3.

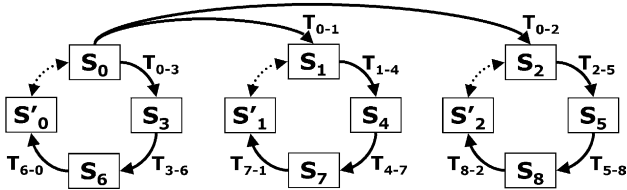


Fig. 6. Registration loops for a given 4D CT canine data set. One of the images is not used in the loops since only nine of the ten images are required. S_N is the epicardial surface at time phase N , while S'_N is the same surface after transformations through a complete loop.

The starting surface for the first loop is the ED surface (S_0), created using the manual segmentation approach outlined earlier. The starting surface for the second and third circuits is S_0 , deformed by T_{0-1} , or T_{0-2} , respectively. These transformations were obtained using our algorithm with the MAD similarity metric. Because the consistency error estimate is only affected by cardiac motion, it is not crucial that the starting surfaces be exceptionally accurate representations of the epicardial surface at that particular time point in the cardiac cycle. As long as the epicardial surface is well represented, the consistency estimated error will correspond well with the epicardial surface.

4.3. Human heart

The ability of the algorithm to model the internal structures of the human heart was also determined through the consistency measure discussed above. For this study, images of two volunteers were acquired in the coronal plane using a 1.5 T GE CVi scanner. Both scans were performed using the fast cine SPGR pulse sequence with an image matrix of 256×128 , eight views per segment, and a 20° flip angle. The image set for the first volunteer (D1) was acquired with a breath hold duration of ≈ 20 s and two signal averages (NEX), while the image set for the second volunteer (D2) was acquired with a breath hold of 42 s and four signal averages. Longer breath holds permitted the creation of 20 image volumes over the cardiac cycle with isotropic voxels ($\sim 1.5 \times 1.5 \times 1.5 \text{ mm}^3$) (Moore et al., 2003). Shorter breath hold duration resulted in more clinically feasible, 20-phase volumes, with $1.5 \times 1.5 \text{ mm}^2$ in-plane resolution, and $\sim 6 \text{ mm}$ slice thickness.

The myocardium, right atrium and right ventricle, aorta and left atrium, pulmonary arteries and vena cava, and epicardial surface were segmented in the ED image frame of the D1 and D2 data sets. Registration loops shown in Fig. 7 were used to estimate the error using the consistency measurements described above. Because

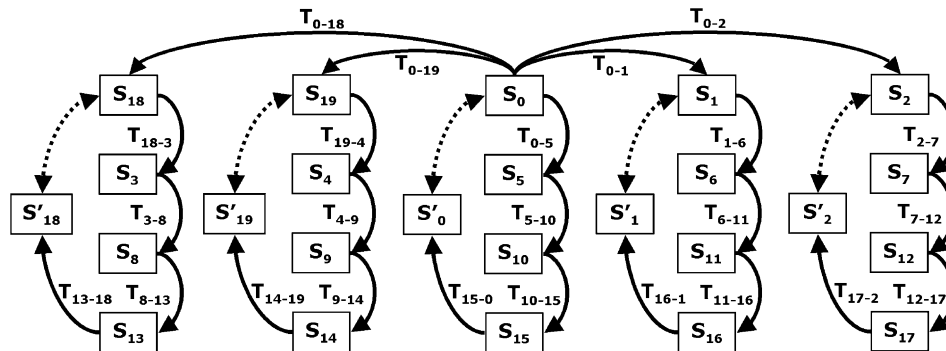


Fig. 7. Registration loops for a given 4D MR data set. S_N is the surface of the heart component of choice at time N , while S'_N is the surface of the same component after transformations through a complete loop.

four transformations were used per registration loop, the error per single registration was estimated as the RMS of the Euclidian distance between two corresponding models divided by $\sqrt{4}$.

The process was repeated with down-sampled versions of the D2 images to facilitate a more direct comparison with the D1 data. Four slices from any original D2 image were averaged to create one thick slice. As a result the voxel dimension in the new D3 data set was increased from $1.5 \times 1.5 \times 1.5$ to $1.5 \times 1.5 \times 6$ mm³, and the SNR was increased by $\sqrt{4}$ (four averages).

5. Results

Original and down-sampled images of the excised porcine heart were registered using the approach discussed above. All registrations were repeated with each of the SMs introduced in Section 3.3. The error is expressed as the RMS of the Euclidian distance between

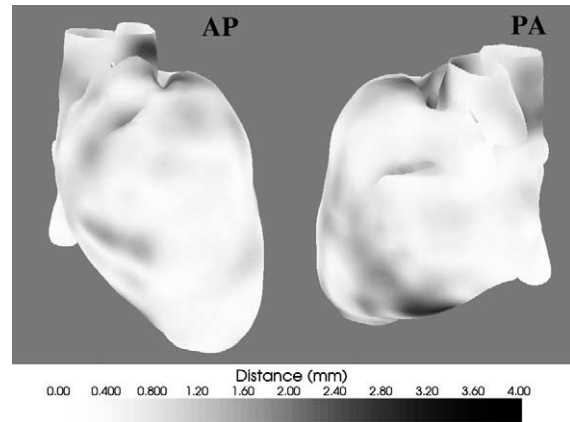


Fig. 8. Canine average Euclidean distance maps between corresponding points on the starting surface and the surface obtained after applying a closed loop of transformations. Both maps were calculated from the S_0 loop of the first canine subject, by averaging distances obtained with the six SMs. Two views of each surface are shown, AP is approximately in the anterior–posterior direction, and PA is approximately in the posterior–anterior direction.

Table 1

Excised porcine heart experiment: RMS of the Euclidean distance between the landmark positions in the target image and the landmark positions in the source image after registration

Registration	MAD	MSD	NCC	NMI	MI	ECC
<i>RMS error (mm) for a registration with the given SM (original data)</i>						
5 to 0	1.49	1.58	2.22	1.20	0.89	1.27
5 to 1	0.91	1.00	1.36	0.83	0.67	0.79
5 to 2	0.88	1.01	1.33	0.68	0.62	0.72
5 to 3	0.75	0.88	1.14	0.65	0.62	0.70
5 to 4	0.64	0.79	1.05	0.60	0.50	0.68
Mean ± SD	0.93 ± 0.33	1.05 ± 0.31	1.42 ± 0.46	0.79 ± 0.24	0.66 ± 0.14	0.83 ± 0.25
<i>RMS error (mm) for a registration with the given SM (down-sampled data)</i>						
5 to 0	1.47	1.60	2.09	1.43	0.96	1.26
5 to 1	0.96	1.03	1.49	1.20	0.75	0.99
5 to 2	0.94	1.08	1.50	0.92	0.65	0.86
5 to 3	0.84	0.92	1.18	0.86	0.66	0.79
5 to 4	0.71	0.83	1.03	0.77	0.63	0.85
Mean ± SD	0.98 ± 0.29	1.09 ± 0.30	1.46 ± 0.41	1.04 ± 0.27	0.73 ± 0.14	0.95 ± 0.19

Five registrations were performed for each data set using six different SMs. Original data is $0.35 \times 0.35 \times 0.63$ mm³, down-sampled data is $0.35 \times 0.35 \times 1.25$ mm³.

Table 2

Canine experiment: RMS (divided by $\sqrt{3}$) of the Euclidean distance between corresponding points on the starting surface and the surface obtained after applying a closed loop of transformations

Subject	MAD	MSD	NCC	NMI	MI	ECC
<i>RMS error (mm) for the given SM</i>						
1	0.65	0.60	0.66	0.82	0.79	0.76
2	0.76	0.71	0.74	0.50	0.52	0.49
3a	0.59	0.53	0.54	0.66	0.74	0.68
3b	0.51	0.51	0.50	0.55	0.72	0.57
Mean ± SD	0.63 ± 0.11	0.59 ± 0.09	0.61 ± 0.11	0.63 ± 0.14	0.69 ± 0.12	0.63 ± 0.12

Transformations were calculated using the six different SMs. The RMS was averaged over three possible loops per subject. Subject 3 was imaged on two different days.

Table 3

Human experiment: RMS (divided by $\sqrt{4}$) of the Euclidean distance between corresponding points on the starting surface and the surface obtained after applying a closed loop of transformations

Structure	MAD	MSD	NCC	NMI	MI	ECC
<i>RMS error (mm) for the given SM (D1 data set)</i>						
M	0.81	0.76	0.81	1.01	1.06	1.02
RA + RV	0.87	0.88	0.84	1.05	1.03	1.09
A + LA	0.87	0.77	0.83	1.16	1.18	1.26
PS + VC	1.00	0.97	0.90	1.16	1.13	1.23
E	0.73	0.70	0.76	1.01	0.98	1.05
Mean ± SD	0.85 ± 0.10	0.82 ± 0.11	0.83 ± 0.05	1.08 ± 0.08	1.08 ± 0.08	1.13 ± 0.11
<i>RMS error (mm) for the given SM (D2 data set)</i>						
M	1.14	0.85	1.14	1.32	1.33	1.32
RA + RV	1.08	1.00	1.08	1.16	1.13	1.15
A + LA	1.10	0.89	0.96	1.30	1.29	1.28
PS + VC	1.09	0.91	0.94	1.21	1.14	1.20
E	0.81	0.72	0.87	1.02	0.96	0.99
Mean ± SD	1.04 ± 0.13	0.87 ± 0.10	1.00 ± 0.11	1.20 ± 0.12	1.17 ± 0.15	1.19 ± 0.13
<i>RMS error (mm) for the given SM (D3 data set)</i>						
M	1.24	0.78	0.65	1.44	1.12	1.32
RA + RV	1.09	0.89	0.77	1.23	1.19	1.20
A + LA	1.17	0.75	0.65	1.43	1.18	1.36
PS + VC	1.01	0.80	0.69	1.29	1.18	1.25
E	0.87	0.62	0.60	1.16	0.97	1.08
Mean ± SD	1.08 ± 0.14	0.77 ± 0.10	0.56 ± 0.06	1.31 ± 0.12	1.13 ± 0.09	1.24 ± 0.11

Transformations were calculated using the six different SMs. The RMS was averaged over four possible loops per subject. Three image sets were used, D1 with a voxel dimension of $1.5 \times 1.5 \times 6 \text{ mm}^3$, D2 with a voxel dimension of $1.5 \times 1.5 \times 1.5 \text{ mm}^3$, and D3 (down-sampled version of D2) with a voxel dimension of $1.5 \times 1.5 \times 6 \text{ mm}^3$. Each heart was divided into five regions – the myocardium (M), the right atrium and right ventricle (RA + RV), the aorta and left atrium (A + LA), the pulmonary arteries and vena cava (PS + VC), and the epicardial surface (E).

the landmark positions in the source image after registration, and the landmark positions in the target image. The results are presented in Table 1.

For the canine data sets we used the consistency measure to estimate error in extracted motion of the epicardial surface. The results are shown in Table 2.

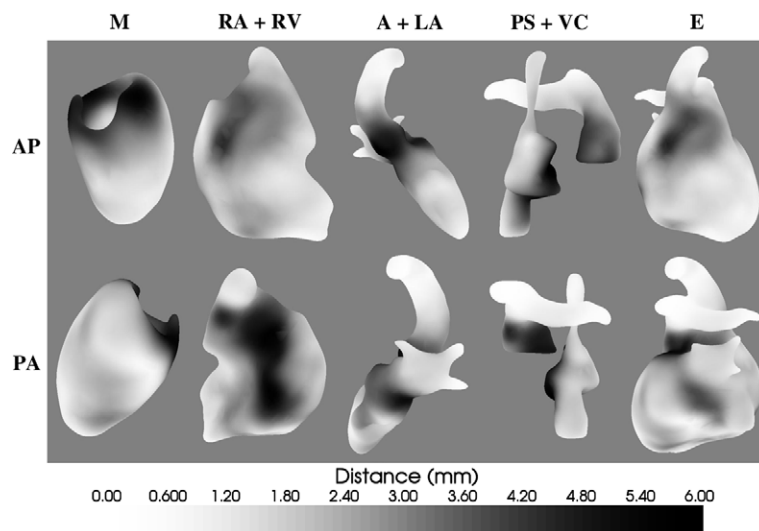


Fig. 9. Human data calculated maps of the Euclidean distance between corresponding points on the starting surfaces and the surfaces obtained after applying a closed loop of transformations. These maps were calculated from the S_0 registration loop of the D2 data set, for each of the sub-heart structures: myocardium (M), the right atrium and right ventricle (RA + RV), the aorta and left atrium (A + LA), the pulmonary arteries and vena cava (PS + VC), and the epicardial surface (E). Two views of each surface are shown, AP is approximately in the anterior–posterior direction, and PA is approximately in the posterior–anterior direction.

In order to explore the spatial distribution of the error over the epicardial surface we also calculated Euclidian distance maps between corresponding points on the S_N and S'_N surfaces. Rather than showing six different sets of maps in Fig. 8, we averaged the Euclidian distances between points obtained with the six SMs to produce a single map. Hence, any dark shades of gray indicate areas on the epicardial surface where the registration is consistently having difficulties recovering the original shape after a full loop of transformations, regardless of the SM used. Light areas indicate parts of the epicardial surface where the algorithm is able to return the starting surface to the original shape.

In the last experiment we validated models of the human heart using the consistency error estimate. Model sets were created from clinically feasible MR data (D1), a specialized set of high-resolution data (D2), and a down-sampled version of the high-resolution data (D3). Table 3 shows the error as a function of the SM, for each of the model components.

We also studied the spatial distribution of the error by calculating distance maps between the starting surfaces and the surfaces obtained after the complete loop of transformations. Again, only one set of maps was produced by averaging the Euclidian distances over the six SMs. Examples of these maps for the D2 data set can be seen in Fig. 9.

6. Discussion

In this manuscript, we extracted motion from 4D heart images using a registration algorithm with several different SMs. The accuracy of the resulting motion was then validated yielding several interesting results that are worth noting. First, the accuracy of the registration is significantly better if the input images are initially more similar. For example, registration 5 to 0 in Table 1 is more challenging than registration 5 to 4. Similar patterns were observed in the in vivo experiments before averaging for a particular subject. For example, the registration loop 0 (ED) to 5 to 10 (ES) to 15 to 0 (ED) of the MR data sets produced the largest error, since these registrations included the ED and ES time points. As a result, the algorithm was used to extract the maximum amount of motion over the cardiac cycle, consequently resulting in more errors. Since errors are more likely when registering less similar images, dynamic models created with our approach are the most accurate around ED, and least accurate around ES.

The optimal choice of SM strongly depends on the images to be registered. For the porcine experiment (Table 1 – original data), information theoretic SMs performed better than the direct intensity-based SMs. The RMS error of 0.66 ± 0.14 mm obtained with MI

was significantly lower than the errors obtained with the other metrics. The only exception was NMI, where p between the two errors was determined to be 0.0568 using a two-tailed t -test for correlated samples. Although there was no statistical difference between the errors obtained with NMI and ECC, only NMI was statistically better than MAD ($p < 0.05$). Finally, MAD was better than MSD, and MSD was better than NCC ($p < 0.05$).

The registration was most accurate with information theoretic metrics mainly due to the heart filling procedure between image acquisitions. The heart chambers were initially full of water, making them mostly indistinguishable from the heart walls. As contrast agent was forced into the heart, the chambers became brighter, resulting in different contrast for the same heart anatomy in two different images. Due to their direct dependence on image intensity, MAD, MSD and NCC were not able to properly align these areas. This is especially evident for the 5 to 0 registration, where we see the largest discrepancy in the error between registrations performed with different SMs. The information theoretic metrics on the other hand, are more successful in registering these types of images since they compare the structure and statistical dependence of image areas instead of directly comparing intensity values.

The results with the canine experiments in Table 2 were quite different. The MSD metric produced the least amount of error, estimated at 0.59 ± 0.09 mm. However, this error is only significantly different from the error obtained with MAD. The errors obtained with NMI were significantly lower from the errors obtained with MI, and MI produced significantly less errors than ECC.

The directly intensity-based metrics perform better in this experiment since the images are much more similar prior to registration (same intensities for the same anatomy). The performance of information theoretic metrics, however, is degraded due to the decreased resolution of the canine images ($0.35 \times 0.35 \times 1.25$ mm³ voxels) with respect to the porcine data set ($0.35 \times 0.35 \times 0.63$ mm³ voxels). This is because MI-based metrics are statistical in nature (the metrics are computed from estimates of probability densities), and hence are more heavily dependent on the number of samples. This phenomenon is well demonstrated in Table 1, when comparing porcine heart results obtained with the original and the down-sampled image data. The reduced resolution decreases registration accuracy for all metrics, but the effect is the largest for NMI and ECC. Another interesting result is the discrepancy in registration error calculated with the canine and porcine data sets. Errors obtained from the canine data are significantly lower due to the porcine heart filling problems discussed above, and due to the presence of artefacts. Artefacts are problematic for the consistency

method since the error is no longer uncorrelated as was first assumed, resulting in an underestimate of the actual error.

In the human heart experiments, we found that the registration is more accurate on the epicardial surface than on the endocardial surface (there is no difference in registration results for internal heart components themselves). This is not surprising since the epicardium-lung boundary is more clearly defined than boundaries inside the heart. Variations in accuracy as a function of SM are demonstrated in Table 3. The best results were obtained with the MSD metric, which did better than the NCC metric for data sets D1 and D2. The NCC metric was not significantly better than MAD ($p > 0.05$), but MAD was significantly better than MI. MI was not significantly better than NMI ($p > 0.05$) but NMI was better than ECC.

The effect of the difference in voxel size between MR and CT images is best observed when comparing results from the D2 and D3 data sets in Table 3. As the resolution of an image is reduced, we see an increase in registration error with NMI, MI and ECC, while the error with MAD and NCC actually decreases. This similar pattern is observed when comparing results obtained with the D1 and D2 data sets. Although counter-intuitive, there are two plausible reasons for the decrease in consistency error with increasing voxel size. First, as the voxel size of an image is increased, the SNR of that image is also increased. For example, the SNR of the D1 data set is $\approx\sqrt{2}$ higher than the SNR of the D2 data set, since any given voxel is four times larger in D1 than in D2, and since the NEX used to acquire D1 was half of that used for D2. The additional noise in high-resolution data has a profound effect on the optimization consistently avoiding local minima in the registration function. The second reason for better accuracy with lower resolution data is the partial volume artefact. The large voxels in the D1 data set make small-scale registrations impossible. The registration of the D1 data set is therefore much more global, and easier to perform consistently for the algorithm. The consistency measure can be misleading in this case, incorrectly indicating that the low-resolution data set can be registered more accurately. These effects also occur in registrations with information theoretic metrics. The reason why these errors increase as the voxel size is increased is due to the heavy dependence of the NMI, MI and ECC on the number of samples used to calculate them. Consequently, although the images have higher SNR, and there is less local deformation to recover, the consistency error with these metrics still increases.

Further insights into the accuracy of the extracted motion can be obtained by examining the spatial distribution of the error for the canine data in Fig. 8 and for the human data in Fig. 9. First, it is important to

note that these maps show the distance between corresponding points averaged over the six SMs and not the average registration error at these points. However, since registration error is proportional to the distance between points, these maps can be used to study any localized errors in the registration. Darker areas on the surfaces indicate larger distances between the starting and ending location, hence a larger error in the motion extraction algorithm in that area of the model. From the two figures we see that there are several areas of concern, where the two surfaces are relatively far apart. These errors occur for several reasons, but are mainly due to poor contrast between surrounding tissues that are being modeled. In the canine and human data sets this occurs in the pulmonary trunk, where we are trying to extract the motion of soft tissue imbedded in other soft tissue. This also occurs around the right atrium – ascending aorta division in both data sets, and the area of the heart that attaches to the sternum, spine and diaphragm, all difficult to segment accurately due to poor image contrast. Specific to the canine CT data is the problem of contrast agent pooling and re-circulation. This is mostly prevalent in the posterior area of the vena cava, where unpredictable re-circulation causes a decrease in registration consistency around the immediate area. Finally, increased errors occur in parts of the heart that undergo the most motion, such as the valves in the internal heart models, and the heart apex and superior areas of the epicardial surface models. This is where our algorithm reaches its limits in extracting the highly non-linear deformations, and consequently where it makes the most errors. As expected, the algorithm performs most consistently on the epicardial surface due to the high contrast between the heart and surrounding lung tissue (except attachments to sternum, spine and diaphragm). This is an important area of the heart to model accurately since it is the most crucial in MICS.

Due to the presence of easily identifiable landmarks, the porcine data set was extremely useful in validation. It does, however, have several limitations that are important to discuss. First, the porcine heart was completely surrounded by air – no attachments to other tissues found in vivo were modeled. On the other hand, the ex vivo heart had no surrounding features that could be used as additional alignment cues. Although the epicardial surfaces in the source and target images are easier to register in the phantom data set, it is more difficult to match up the exact region on the source epicardium with the corresponding region on the target epicardium. In addition to this, the contrived motion of the porcine heart was not entirely realistic, since image acquisition was simplified and did not include problems related to ECG-gating. As a result, the accuracy calculated in this experiment is probably a best-case scenario when compared to in vivo data sets.

The limitations of the *in vivo* experiments arise mostly due to the method of calculating registration error without a real gold standard. The consistency error estimate, although accurate in perfected circumstances, can underestimate registration error in the presence of imaging artefacts. This method can also produce misleading results since the error calculated in this manner is minimal when the registration is accurate and therefore consistent, and/or the registration transformation is nearly zero, in which case it will be considered consistent even if it is not accurate. In our experiments, we minimize this effect by ensuring the transformations are of significant displacement before calculating the consistency error.

In terms of the final models, limitations result mostly due to poor contrast between tissues, and the resulting difficulties in segmentation, as well as imaging artefacts, such as contrast agent pooling seen in the vena cava of the canine images. Also, we see the largest errors in our models for large and fast motion, especially around the valves and the heart apex. Finally, due to difficulties in extracting a large amount of deformation between images separated significantly in time (such as ED and ES), models are most accurate around ED, and the least accurate around ES.

7. Conclusions and future work

In this work, dynamic models of the heart were obtained from CT and MR images, through a combination of segmentation, surface rendering, and image registration. The accuracy of the cardiac models was determined with three validation experiments on *ex vivo*, and *in vivo* canine and human data sets. On average, the models satisfied our accuracy requirement of less than 1 mm error when the correct SM was chosen. In the excised heart experiment we showed that MI gives the best accuracy, probably due to the contrast agent inconsistencies within the heart phantom. In the case of the *in vivo* data sets, best accuracy was obtained using the MSD similarity metric. Conventional *in vivo* 4D data sets are of comparatively low resolution and the differences between image frames are not very large. As a result, we conclude that MSD is the best choice for extracting the motion of the heart from 4D data sets.

This work has concentrated on validating the motion of a model extracted from a given 4D data set; no spatial validation of the actual model was performed. In the future, a dynamic model will be created from a high resolution and high SNR MR image. This model will serve as a template that can be matched to a given patient data set using a variation of the registration algorithm, eliminating the segmentation step for each patient. The ability of this method to represent different

hearts will then be studied by comparing results to other segmentation methods.

Another future goal will be to determine the degree of organ deformation between pre-procedural and intra-procedural situations. Currently, VCSP models are based on the pre-procedural images, which often do not represent the situation during surgery. It is important to determine the effects of lung collapse and chest insufflation on the location, and morphology of the heart. Having such information is essential when correcting a dynamic heart model to represent the situation within the thorax during an operation.

Although not currently employed clinically, systems such as VCSP will greatly benefit the training, planning, and navigation stages of MICS. In this article, we have presented our method of modeling the heart and have shown that the motion can be extracted with sub-mm accuracy. We believe that the current results demonstrate the potential of such deformable models in the development of visualization tools for MICS.

Acknowledgements

The authors thank Aaron So, Rhonda Walcarius, and Kathy Parker for help image acquisition, Joy Dunmore-Buyze for phantom preparation, Atamai, Inc. for visualization software, Ravi Gupta for help with code development, Dr. Mark Wachowiak, John Moore and Guy-Anne Turgeon for helpful discussions and assistance in writing this manuscript. This work was supported by grants from the Canadian Institutes of Health Research (MOP 14735), Canadian Heart and Stroke Foundation (NA 4755), Ontario Consortium for Image-guided Surgery and Therapy, National Science and Engineering Research Council of Canada, and the University of Western Ontario.

References

- Astola, J., Virtanen, I., 1982. Entropy correlation coefficient, a measure of statistical dependence for categorized data. In: Proceedings of the University of Vaasa – Discussion Papers 44. University of Vaasa, Vaasa, pp. 1–12.
- Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Int.* 11 (6), 567–585.
- Boyd, W.D., Desai, N.D., Kiaii, B., Rayman, R., Menkis, A.H., McKenzie, F.N., Novick, R.J., 2000a. A comparison of robot-assisted versus manually constructed endoscopic coronary anastomosis. *Ann. Thorac. Surg.* 70, 839–843.
- Boyd, W.D., Rayman, R., Desai, N.D., Menkis, A.H., Dobkowski, W., Ganapathy, S., Kiaii, B., Jablonsky, G., McKenzie, F.N., Novick, R.J., 2000b. Closed-chest coronary artery bypass grafting on the beating heart with the use of a computer-enhanced surgical robotic system. *J. Thorac. Cardiovasc. Surg.* 120 (4), 807–809.

- Chiu, A.M., Dey, D., Drangova, M., Boyd, W.D., Peters, T.M., 2000. 3-D image guidance for minimally invasive robotic coronary artery bypass. *Heart Surg. Forum* 3 (3), 224–231.
- Coste-Maniere, E., Adhami, L., Mourgues, F., Carpentier, A., 2003. Planning, simulation, and augmented reality for robotic cardiac procedures: the STARS system of the ChIR team. *Semin. Thorac. Cardiovasc. Surg.* 15 (2), 141–156.
- Declerck, J., Feldmar, J., Ayache, N., 1998. Definition of a four-dimensional continuous planispheric transformation for the tracking and the analysis of left-ventricle motion. *Med. Image Anal.* 2 (2), 197–213.
- Frangi, A.F., Niessen, W.J., Viergever, M.A., 2001. Three-dimensional modeling of functional analysis of cardiac images: a review. *IEEE Trans. Med. Imaging* 20 (1), 2–25.
- Gillinov, A.M., Blackstone, E.H., McCarthy, P.M., 2002. Atrial fibrillation: current surgical options and their assessment. *Ann. Thorac. Surg.* 74, 2210–2217.
- Herzog, C., Dogan, S., Diebold, T., Khan, M.F., Ackermann, H., Schaller, S., Flohr, T.G., Wimmer-Greinecker, G., Moritz, A., Vogl, T.J., 2003. Multi-detector row CT versus coronary angiography: preoperative evaluation before totally endoscopic coronary artery bypass grafting. *Radiology* 229, 200–208.
- Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J., 2001. Medical image registration. *Phys. Med. Biol.* 46, R1–R45.
- Holden, M., Hill, D.L.G., Denton, E.R.E., Jarosz, J.M., Cox, T.C.S., Rohlfing, T., Goodey, J., Hawkes, D.J., 2000. Voxel similarity measures for 3-D serial MR brain image registration. *IEEE Trans. Med. Imaging* 19 (2), 94–102.
- King, R.C., Reece, T.B., Hurst, J.L., Shockley, K.S., Tribble, C.G., Spontnitz, W.D., Kron, I.L., 1997. Minimally invasive coronary artery bypass grafting decreases hospital stay and cost. *Ann. Surg.* 225 (6), 805–811.
- Kucher, M., Lipp, E., Schwerzmann, M., Zimmerli, M., Allemann, Y., Seiler, C., 2001. Gender differences in coronary artery size per 100 g of left ventricular mass in a population without cardiac disease. *Swiss Med. Wkly.* 131, 610–615.
- Lehmann, G., Habets, D., Holdsworth, D.W., Peters, T.M., Drangova, M., 2002. Simulation of intra-operative 3D coronary angiography for enhanced minimally invasive robotic cardiac intervention. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002*, LNCS 2489. Springer, Berlin, pp. 268–275.
- Lorenson, W.E., Cline, H.E., 1987. Marching cubes: a high resolution 3D surface construction algorithm. *Comput. Graph. (ACM)* 21 (4), 163–169.
- Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Mohiaddin, R., Ruckert, D., 2002. Atlas-based segmentation and tracking of 3D cardiac mr images using non-rigid registration. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002*, LNCS 2488. Springer, Berlin, pp. 642–650.
- Loulmet, D., Carpentier, A., d’Attellis, N., Berrebi, A., Cardon, C., Ponzio, O., Aupécle, B., Relland, J.Y.M., 2003. Endoscopic coronary artery bypass grafting with the aid of robotic assisted instruments. *J. Thorac. Cardiovasc. Surg.* 118, 4–10.
- Maes, F., Vandermeulen, D., Suetens, P., 1999. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Med. Image Anal.* 3 (4), 373–386.
- Mäkelä, T., Clarysse, P., Sipilä, O., Pauna, N., Pham, Q.C., Katila, T., Magnin, I.E., 2002. A review of cardiac image registration methods. *IEEE Trans. Med. Imaging* 21 (9), 1011–1021.
- Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M., 2002. 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE Trans. Med. Imaging* 21 (9), 1167–1178.
- Moore, J., Drangova, M., Wierzbicki, M., Barron, J., Peters, T.M., 2003. A high resolution dynamic heart model based on averaged MRI data. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2003*, LNCS 2878. Springer, Berlin, pp. 549–555.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Comput. J.* 7, 308–313.
- Schroeder, W.J., Zarge, J.A., Lorensen, W.E., 1992. Decimation of triangle meshes. *Computer Graphics – SIGGRAPH 92*, 65–70.
- Sijbers, J., den Dekker, A.J., Van Audekerke, J., Verhoye, M., Van Dyck, D., 1998. Estimation of noise in magnitude MR images. *Magn. Reson. Imaging* 16 (1), 87–90.
- Stevens, J.H., Burdin, T.A., Peters, W.S., Siegel, L.C., Pompili, M.F., Vierra, M.A., St.Goar, F.G., Ribakove, G.H., Mitchell, R.S., Reitz, B.A., 1996. Port-access coronary artery bypass grafting: a proposed surgical method. *J. Thorac. Cardiovasc. Surg.* 111, 567–573.
- Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Patt. Recognit.* 32 (1), 71–86.
- Szpala, S., Guiraudon, G., Peters, T.M., 2003. Cardiac endoscopy enhanced by dynamic organ modeling for minimally-invasive surgery guidance. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2003*, LNCS 2878. Springer, Berlin, pp. 499–506.
- Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* 1 (1), 35–51.
- Wierzbicki, M., Peters, T.M., 2003. Determining epicardial surface motion using elastic registration: towards virtual reality guidance of minimally invasive cardiac interventions. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2003*, LNCS 2878. Springer, Berlin, pp. 722–729.
- Woods, R., Grafton, S., Holmes, C., Cherry, S., Mazziotta, J., 1998a. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* 22 (1), 139–152.
- Woods, R., Grafton, S., Watson, J., Sicotte, N., Mazziotta, J., 1998b. Automated image registration: II. Intersubject validation of linear and nonlinear models. *J. Comput. Assist. Tomogr.* 22 (1), 153–165.