

Registration of a 4D Cardiac Motion Model to Endoscopic Video for Augmented Reality Image Guidance of Robotic Coronary Artery Bypass

Michael Figl^{1,6}, Daniel Rueckert¹, David Hawkes³, Roberto Casula⁴,
Mingxing Hu³, Ose Pedro¹, Dong Ping Zhang¹, Graeme Penney⁵,
Fernando Bello², and Philip Edwards^{1,2}

¹ Department of Computing, Imperial College London, UK

² Department of Biosurgery and Surgical Technology, Imperial College London, UK

³ Centre of Medical Image Computing, University College London, UK

⁴ Cardiothoracic Surgery, St. Mary's Hospital, London, UK

⁵ Division of Imaging Sciences, King's College London, UK

⁶ Center for Biomedical Engineering, Medical University of Vienna, Austria

Abstract. The aim of the work described in this paper is registration of a 4D preoperative motion model of the heart to the video view of the patient through the intraoperative endoscope, in order to overlay the real video sequence with it. As the heart motion is cyclical it can be modelled using multiple reconstructions of cardiac gated coronary CT.

We propose the use of photoconsistency between the two views through the da Vinci endoscope to align to the preoperative heart surface model from CT. We propose averaging of the photoconsistency over the cardiac cycle to improve the registration compared to a single view.

Results are presented for simulated renderings and for real video of a beating heart phantom. We found much smoother behaviour of the test function at the minimum when using multiple phases for the registration, furthermore convergence was found to be better when more phases are used.

1 Introduction

Totally endoscopic coronary artery bypass (TECAB) is a minimally invasive robotic cardiac procedure performed on the beating heart. The view of the patient is provided by a stereo endoscope which gives the surgeon a 3D view of the operative field.

There is a significant proportion of these operations that need to convert to more conventional procedures. This may be due to misidentification of the diseased vessel or difficulty in locating the vessel through fatty deposits on the surface of the myocardium. For this reason there has been some interest in introducing augmented reality (AR) image guidance to TECAB, overlaying these vessels with some preoperative images. Mourgues et al. [6] and Adhami et al [1] proposed such a system where the preoperative model to be overlaid over the real image comes from bilateral x-rays.

In our case the preoperative model comes from coronary CT. This provides a fully 4D model of the heart from which a motion model of both the myocardium and coronary arteries must be extracted.

To overlay the endoscopic view the preoperative model has to be registered to the endoscopic images. If many landmarks are visible, e.g. bifurcations or tubes, we could use a shape-based registration such as that proposed by Jomier [5]. However in our case only parts of the heart surface are seen via the endoscopic camera and there may not be corresponding vessels visible on the heart surface and in the preoperative image.

Since we have a stereo video stream it is possible to use photo-consistency [3] as a measure of registration to the preoperative surface. If we have temporal synchronisation between the video and the motion model, the periodic motion of the heart enables us to use all points of time in the cardiac cycle.

2 Materials & Methods

2.1 Heart Phantom

To allow for reproducible testing the motion model was derived from a 4D CT scan of a pneumatic heart phantom made by The Chamberlain Group, Great Barrington, MA, USA. This phantom allows three different heart frequency settings. The heart rate of the phantom was determined using a video sequence of the beating phantom and comparing the first image of this sequence to the others using cross-correlation as a similarity measure. The frequency was found as a peak in the Fourier transform of this function.

The phantom was scanned by a Philips Brilliance 64 ECG triggered CT, and reconstructed at ten different phases of the cardiac cycle. The ECG signal for the heart phantom was generated by a signal generator from Hameg, Mainhausen, Germany. The surfaces of the ten phases were extracted using the VTK implementation of Marching Cubes (Kitware Inc, NY, USA).

2.2 Video Images

The target of our work is the daVinci robot which is equipped with a stereo endoscope with stereo video output via BNC and S-video. As a development system two Sony DCR-TRV 30E digital video cameras were used. Video grabbing was performed using an LFG4 4 channel frame grabber made by Active Silicon, Uxbridge, UK. The video cameras were calibrated using a calibration toolbox implemented in MatLAB by Jean-Yves Bouguet, [2].

2.3 Temporal Registration

Aligning the phases of the video sequence and the 4D CT is done by visual judgment. There are several time points that can easily be found in both the rendered CT surface and the video sequences and can be used for temporal synchronisation. E.g. the end of the systolic phase, where the heart is in maximal contraction. In practice this could also be achieved using the ECG signal

2.4 Point Based Registration

In a first step a simple and robust point based registration of the surface to the video images is done. Several points on the surface and in both video channels are selected, and the optimal rigid body transformation T is found by minimising the distance of the projected surface points to the selected image points iteratively. This registration software was developed using the OpenGL widget of the C++ GUI framework Qt3/4 (Trolltech, Oslo, Norway). Multidimensional minimisation was done using the Fletcher-Reeves conjugate gradient algorithm implemented in the GNU scientific library.

2.5 Photo-consistency Registration

Point based registration was used as a starting value for photoconsistency, see figure 1.



Fig. 1. Registration using photoconsistency. In the upper image the situation after the point to point registration step is shown. The bottom image shows the effect of application of photoconsistency registration.

The idea behind registration using photo-consistency is to judge whether given projections of a surface are compatible with the lighting model. Assuming a Lambertian lighting model [4] a surface point should have the same colour in all of its projections. Figure 2 gives an example, where three points have similar colour values in different images. The mean standard deviation is 2.3 in an 8 bit colour space, but the standard deviation for the green component in the three points of the second image is 26.2.

For the registration we have to find the transformation from the surface to the common coordinate system of the camera calibrations. After applying this

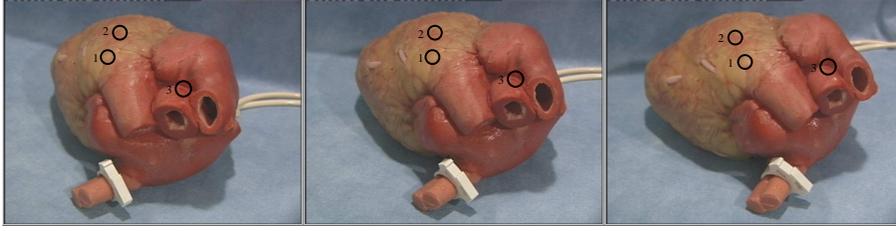


Fig. 2. With Lambert’s light model the colours in the centers of the circles 1, 2, and 3 should be the same in every image. The mean standard deviation of colours values in these points is 2.3 in an 8 bit colour space. On the lower front of the second and third image there is a considerable reflection which does not comply with the Lambertian model and has a very different colour in the first image.

transformation the surface object will be in a position where the projections of the surface points are (photo-)consistent. This consistency can be evaluated using the colour components of the projections of the surface points in different video channels. If the colour diversity in the images is evaluated by their variance, then the desired transformation T can be found by:

$$T = \operatorname{argmin}_x \left(\operatorname{mean}_i (\operatorname{var}_i (P_i T x)) \right) \quad (1)$$

Where P_i denotes the projection matrix for camera calibration i . The comparison for the mean value is done over the surface points that are visible in both images. Figure 3 shows the situation for three cameras. Photo-consistency as a similarity measure for 2D-3D registration was introduced by Clarkson et al. [3].

2.6 Photo-consistency with an image sequence

To take advantage of the fact that there is a whole image sequence available for registration, one can try to use the different phases together to get the registration. Assuming that all the deformation or movement of the object can be found in the surfaces and is completely periodic, the transformation from the preoperative to the intraoperative coordinate system is always the same. In figure 4 the chain of transformations from the preoperative coordinate system to the display is shown for three different phases. Both the transformation T and the camera calibrations are unchanged in this idealised situation. This assumption was also used for 3D-3D registration using the iterative closest point algorithm as proposed by Wilson et al. [8].

3 Results

For camera calibration we used 12 images of a checkerboard pattern with 11×9 squares for every channel. The baselines of the squares was 7.5mm. Typical calibration errors with the endoscope were 0.51 and 0.45 pixels in x resp. y direction

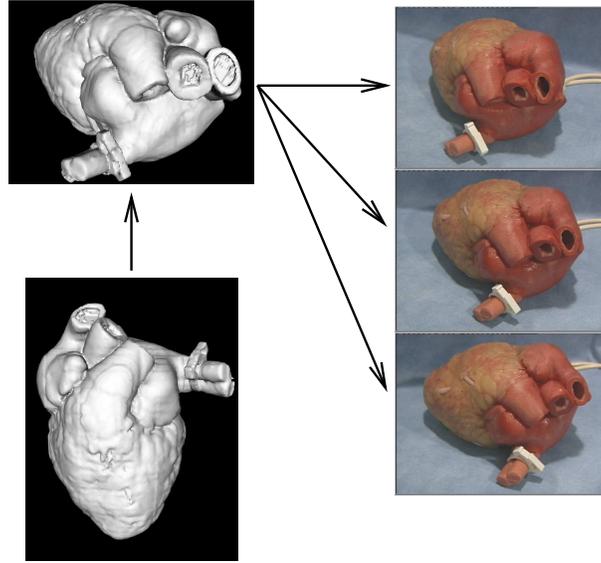


Fig. 3. The transformation T has to move the surface object in a way that the projections of every point of the surface into the video images using the camera calibrations P_i are photo-consistent.

for the separate channels before undistortion and stereo calibration. After undistortion and stereo calibration the distortion coefficients were ten times smaller and the errors in the same magnitude as before, in a concrete example 0.4, 0.33 pixels.

The heart rate could readily be determined using the method described in 2.1, both for endoscopic videos of the beating heart phantom and of a real patient's heart. The frequency measurements on the phantom could be verified by simple counting, the heart rate measured on the patient was identical to the ECG measured heart rate.

3.1 Renderings as Target Images

To evaluate the method without introducing errors by the camera calibration, the temporal alignment or the quality of the video images we did a first registration series from the surface model to renderings of the surfaces. The renderings were made using camera calibration data from the robot's stereo endoscope neglecting the distortion. Unsurprisingly the graph of the photo-consistency surface is smoother when more phases are used. To quantify this, the total variation (normalised by the range) of the graph along the coordinate axes is derived:

$$T(f) := \frac{1}{\max(f) - \min(f)} \int \|f'\| \quad (2)$$

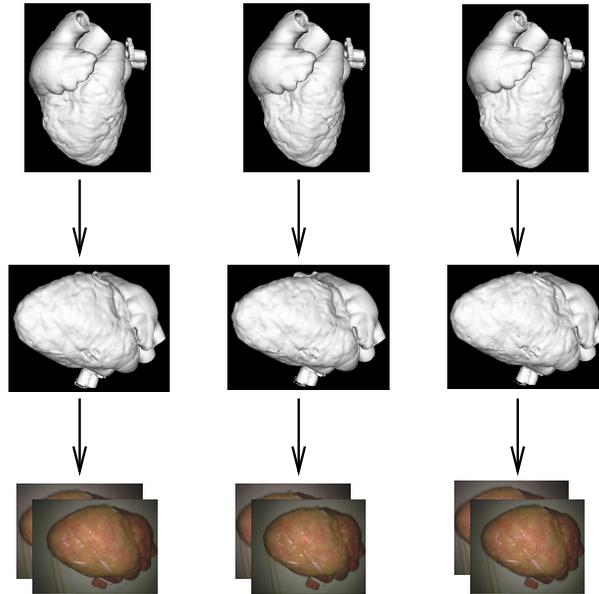


Fig. 4. The transformation between the preoperatively generated surfaces and the intraoperative position is the same for all phases, provided there is no camera movement or respiration.

The corresponding normalised total variations for the function in figure 6 were 1.69, 1.43, and 1.4. The remaining variations can be found in table 5(a).

3.2 Video Cameras

Using the development system consisting of two Sony video cameras we did measurements comparable to that were done with the renderings.

In this test series up to ten phases of the heart cycle were used. The results for translation along the z -axis are shown in figure 6. In figure 6 the situation at the minimum for translation along the x -axis of the world coordinate system is shown when using one, five, or ten phases.

Furthermore we made a test series to compare the convergence, starting at a 5 different positions. The starting positions were found using a minimum point and then changing this with random numbers. The random numbers were generated using the built in function `int random()`. As they are equally distributed in $[0, \text{RAND_MAX}]$ we did $\frac{\text{value}}{\sqrt{6}} \cdot 2 \cdot \left(\frac{\text{random}()}{\text{RAND_MAX}} - 0.5\right)$ with `value=14` as a target distance. The distances to the target point can be found in table 7(a).

Another test series with the same method but the absolute photoconsistency function, a target value of 20 and changes just in the translation components was done. The results can be found in table 7(b).

# phases	t_x	t_y	t_z	r_x	r_y	r_z
1	1.69	2.12	1.57	2.42	2.08	2.22
2	1.66	2.06	1.72	2.33	1.77	2.17
3	1.43	1.96	1.51	1.84	1.9	2.06
4	1.42	1.95	1.65	1.86	1.73	1.91
5	1.4	1.96	1.55	1.97	1.9	2.1
6	1.48	1.93	1.59	1.76	1.82	1.79

# phases	t_x	t_y	t_z	r_x	r_y	r_z
1	1.98	2.21	2.06	2.07	3.7	2.47
3	1.74	2.05	1.82	1.64	2.9	2.1
5	1.73	1.88	1.82	1.61	2.6	1.96
10	1.64	1.76	1.73	1.48	2.07	1.88

(a) phantom renderings

(b) video images

Fig. 5. The normalised total variations for parameter and number of phases for the image sequences.

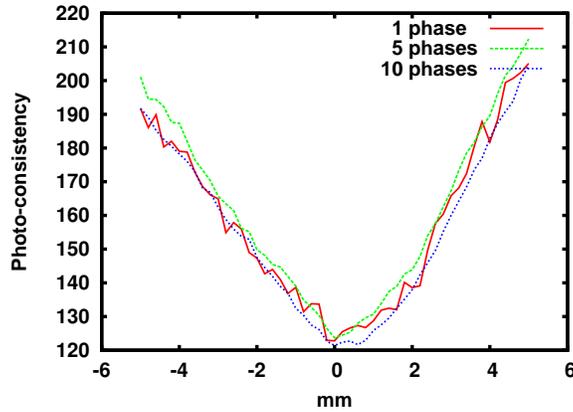


Fig. 6. The image shows translation along the x -axis vs photo-consistency at the minimum point using 1,5, and 10 phases for the video sequence.

before	8.5	4.7	7.8	9.3	9.9
one phase	5.8	2.7	6.4	3.8	8.2
five phases	5.4	2.6	6.4	5.1	7.3

before	8.9	11.6	7.6	2.8	9.3
one phase	5.7	3.9	2.4	2.1	4.6
five phases	4.9	1.0	1.4	1.7	3.5

(a) Rotation and translation were changed.

(b) Just the translation components were changed for the starting value.

Fig. 7. Distances to the minimum points for five sequences.

4 Discussion

From the figures in tables 5(a) and 5(b) we can conclude that the functions are smoother in the meaning of the definition in formula (2). More interesting might be the probable consequences for the convergence which are illustrated in tables 7(a), and 7(b) and show promising behaviour for deviations in the translation component but unfortunately not much improvement if the rotation is included.

An additional complexity in reality is also the heart motion due to breathing. The corresponding movement of the heart can be approximated as rigid. We could therefore hope to be able to separate these two motion components by means of projective geometry.

The confocal light sources for the stereo endoscope of the da Vinci robot are also a significant difficulty, as they are not compatible with our diffuse light model and cause considerable noise. We could simply ignore the brightest parts of the video images as suggested by Stoyanov et al. [7]. Another potential technique would be to calibrate the position of these light sources and to incorporate specular reflections into our lighting model.

Furthermore the endoscopic cameras have a very short baseline of around 5mm. This may make stereo-based registration methods somewhat unstable in the z -direction. This could potentially be overcome using another method, such as external tracking, to estimate the z distance to the surface of the myocardium.

We have presented the notion of using photoconsistency to register intraoperative stereo endoscopic video to a preoperative motion model of the heart surface. Using multiple synchronised frames produces smoother minima in the cost function. We will examine how well this translates into robust 3D registration using both phantom and patient data. This provides a strategy for model-based guidance of TECAB procedures.

References

1. L. Adhami and E. Coste-Maniere. A versatile system for computer integrated minimally-invasive robotic surgery. *MICCAI 2002*, 2488:272–281, 2002.
2. J. Bouguet. Camera Calibration Toolbox for Matlab. www.vision.caltech.edu/bouguetj/calib_doc, 2008.
3. M. J. Clarkson, D. Rueckert, D. L. G. Hill, and D. J. Hawkes. Using photoconsistency to register 2d optical images of the human face to a 3D surface model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:1266–1280, 2001.
4. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics*. Addison Wesley, 2nd edition, 1995.
5. J. Jomier, E. Bullitt, M. van Horn, C. Pathak, and S. Aylward. 3D/2D Model-To-Image Registration Applied to TIPS Surgery. In R. Larsen, M. Nielsen, and J. Sporring, editors, *MICCAI 2006*, LNCS (4191), pages 663–9. Springer-Verlag, 2006.
6. F. Mourgues and E. Coste-Maniere. Flexible calibration of actuated stereoscopic endoscope for overlay in robot assisted surgery. *MICCAI 2002, Pt 1*, 2488:25–34, 2002.
7. D. Stoyanov, G. P. Mylonas, F. Deligianni, A. Darzi, and G. Z. Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. *MICCAI 2005, Pt 2*, 3750:139–146, 2005.
8. K. Wilson, G. Guiraudon, D. Jones, and T. Peters. 4D Shape Registration for Dynamic Electrophysiological Cardiac Mapping. In R. Larsen, M. Nielsen, and J. Sporring, editors, *MICCAI 2006*, LNCS (4191), pages 520–7. Springer-Verlag, 2006.