Robust 3D Visual Tracking for Robotic-Assisted Cardiac Interventions

Rogério Richa, Antônio P.L. Bó, and Philippe Poignet

LIRMM - UMR 5506 CNRS - UM2
161 Rue Ada 34392 Montpellier, France
{rogerio.richa,antonio.bo,philippe.poignet}@lirmm.fr

Abstract. In the context of minimally invasive cardiac surgery, active vision-based motion compensation schemes have been proposed for mitigating problems related to physiological motion. However, robust and accurate visual tracking is a difficult task. The purpose of this paper is to present a hybrid tracker that estimates the heart surface deformation using the outputs of multiple visual tracking techniques. In the proposed method, the failure of an individual technique can be circumvented by the success of others, enabling the robust estimation of the heart surface deformation with increased spatial resolution. In addition, for coping with the absence of visual information due to motion blur or occlusions, a temporal heart motion model is incorporated as an additional support for the visual tracking task. The superior performance of the proposed technique compared to existing techniques individually is demonstrated through experiments conducted on recorded images of an *in vivo* minimally invasive CABG using the DaVinci robotic platform.

Keywords: robotic assisted cardiac surgery, robust visual tracking, stereo.

1 Introduction

Recently, great developments have been made in the field of Minimally Invasive Surgery (MIS). While the benefits of this modality of surgery for the patient are numerous, mobility and visibility difficulties hinder the precise execution of the surgical gestures. For tackling some of these difficulties, surgical robots have been developed for assisting surgeons by improving the ergonomics, visualization and dexterity issues related to the minimally invasive procedure. However, current surgical platforms do not offer solutions for restoring the tactile feedback and physiological motion still needs to be manually compensated by the surgeons.

In this context, active vision-based motion compensation schemes [1] have been proposed for mitigating problems related to physiological motion during surgery. Such systems are particularly useful in beating heart interventions such as the off-pump minimally invasive coronary bypass artery grafting (CABG). Furthermore, the estimation of the beating heart motion using the visual feedback from the endoscope is a practical solution since no additional sensors are required in the MIS workspace.

T. Jiang et al. (Eds.): MICCAI 2010, Part I, LNCS 6361, pp. 267–274, 2010.

In the literature, several techniques for tracking the heart motion using vision have been proposed. Mainly, visual cues such as salient features [2,3], texture [4] or shading [5] have been explored. However, most techniques display poor performance in the presence of the large heart surface deformations, illumination variations and specular reflections. Furthermore, occlusions by surgical tools are not handled in most cases.

The purpose of this paper is to present a hybrid tracker that estimates the 3D temporal and spatial deformation of a selected region of interest on the heart surface using stereo endoscopic images. The method comprises the estimation of a parametric deformable model for representing the heart surface based on the most reliable outputs from multiple visual tracking techniques. In addition, for coping with the absence of visual information due to motion blur or occlusions (e.g. by specular reflections, surgical instruments), a temporal heart motion model based on a time-varying dual Fourier series is incorporated as an additional support for the visual tracking task. Hence, the proposed technique is able to robustly track large regions of interest on the heart with high spatial resolution while naturally handling eventual occlusions. The superior performance of the proposed technique compared to existing techniques individually is demonstrated through *in vivo* experiments conducted on recorded images of a minimally invasive CABG using the DaVinci robotic platform.

2 Methods

2.1 The Hybrid Visual Tracking Method

In the context of MIS, the assumptions on which most tracking algorithms are based are often violated due the presence of large soft-tissue deformations, illumination variations and specular reflections, making continuous and accurate visual tracking with a single technique a difficult task. In this study, the use of multiple visual cues is proposed for increasing tracking robustness and spatial resolution. Using multiple methods, the failure of an individual technique can be circumvented by others, enabling the estimation of the heart surface deformation with superior robustness and spatial resolution. Furthermore, all tracking techniques used in the scheme run in parallel and can be implemented in a computationally efficient manner.

Our objective is the estimation of the deformation of a region of interest on the heart surface, manually chosen by the surgeon from any of the stereo images. For this purpose, rectified images of a calibrated stereo endoscope were used. Three different tracking methods are employed: two feature-based approaches – the modified Lucas-Kanade tracker proposed in Stoyanov et al. [2] and the SIFT [6] – and the region-based iterative registration technique proposed in Richa et al. [4]. Since the proposed framework is modular, additional methods can be easily incorporated for increased tracking quality.

Initially, for choosing the most reliable estimate of the heart surface deformation given the outputs of multiple methods, a quality evaluation step is executed. For this purpose, the alignment error is chosen as a measure of tracking quality

and the normalized cross correlation (NCC) in a subregion around each feature or control point is computed between both left and right stereoscopic images and the reference image. If the image alignment error drops below a defined threshold ($\tau < 0.90$), the 3D coordinates relative to a given estimate is treated as unreliable. The threshold τ was chosen empirically and is set very high for avoiding the need of fine tuning and false matches.

Next, using the set of reliable estimates of the 3D motion of the region of interest, a dense model of the heart surface deformation is computed. Here, the Thin-Plate Spline (TPS) model proposed in [4] is employed since it offers a good approximation of the heart surface shape in comparison with similar parametric deformable models. However, it can also be replaced by models that handle surface topological changes [7] or consider the biomechanical tissue properties [8]. The TPS is an interpolating function $m : \Re^2 \rightarrow \Re$ of a 2D point $\mathbf{x}$, defined by the basis function $B(r) = r^2 log(r^2)$, a $(n+3)$ parameter vector $(w_1, ..., w_n, a_1, b_2, c_3)$ and a set of control points $\mathbf{c} = (\check{x}, \check{y})$:

$$m(\mathbf{x}) = c_1 + a_2 x + b_3 y + \sum_{i=1}^{n} w_i \, B(||\mathbf{c}_i - \mathbf{x}||) \tag{1}$$

The TPS model is computed using the reliable tracking estimates as control points $\mathbf{c}$. By stacking three TPS functions $[m^x(\mathbf{x}) \; m^y(\mathbf{x}) \; m^z(\mathbf{x})]$, the 3D position of any point on the reference image of the region of interest selected by the surgeon can be computed in 3D by back-projection using the parameters of the calibrated stereo cameras.

The 3D deformable model has two functions: it provides a reliable dense model of the heart surface deformation and interpolates the unreliable tracking measurements. The latter consists in one of strong points of the proposed method, enabling tracking to recover from failures (due to large deformations, local minima problems). In our works, we found that no regularization or constraints on the TPS surface deformation were necessary (consequence of the alignment error threshold). This represents an advantage since *ad hoc* assumptions about the heart surface deformations are avoided. Figure 1 summarizes the different steps of the hybrid tracking method.

2.2 Temporal Modeling of the Beating Heart Motion

Due to motion blur, large specular reflections or the motion of the surgical instruments, visual information from the tracked region of interest may not be available for certain periods of time. For circumventing such problems, the quasi-periodic beating heart motion can be modeled and occlusions can be bridged using the motion predicted by the model.

As proposed in [9], the heart motion can be considered as the sum of the respiratory and cardiac motions, which can be represented as a dual non-stationary Fourier series. Given the 3D coordinates $\mathbf{p} = [^x p \; ^y p \; ^z p]$ of a given point on the heart surface, the motion dynamics p of a Cartesian coordinate at a given instant t can be parameterized as the sum of two Fourier series, such that:

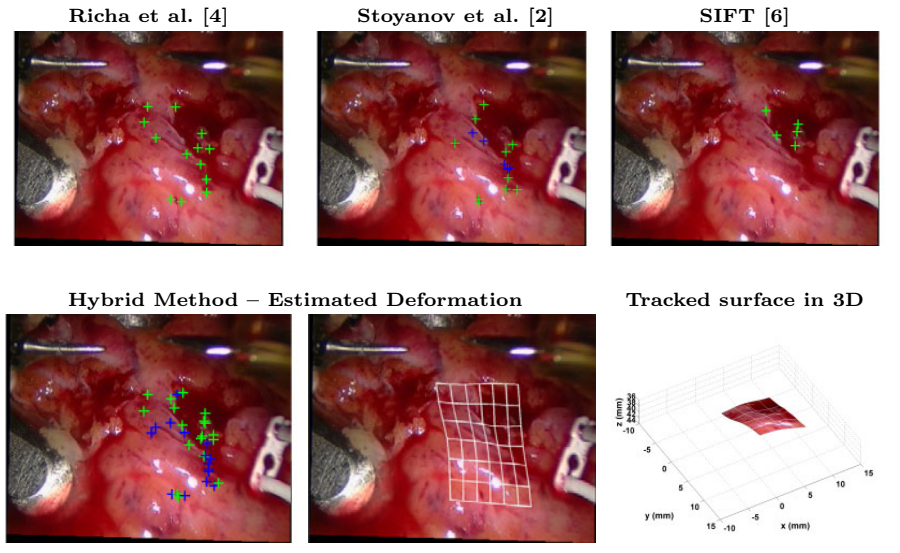


Fig. 1. (Top row) The output of the tracking techniques used in the study from the left camera of the stereo endoscope. Lost features from [2] are marked in blue. Notice that only two reliable SIFT matches were available when large deformations occurred. (Bottom left) The reliable correspondences are marked in green, while interpolated unreliable estimates are marked in blue. (Bottom center) Interpolating TPS surface (Bottom right) Estimated 3D surface.

$$p(t,\mathbf{f}) = \sum_{h=1}^{H_r}\Big[a_h \sin\Big(h\sum_{k=t_0}^{t}\omega_r(k)\Big) + b_h \cos\Big(h\sum_{k=t_0}^{t}\omega_r(k)\Big)\Big] + c_r + \sum_{h=1}^{H_c}\Big[d_h \sin\Big(h\sum_{k=t_0}^{t}\omega_c(k)\Big) + e_h \cos\Big(h\sum_{k=t_0}^{t}\omega_c(k)\Big)\Big], \tag{2}$$

where H_r and H_c are the number of harmonics for modeling the respiratory and cardiac components respectively, ω_r and ω_c are the respiratory and cardiac frequencies, $\sum_{t_0}^{t}\omega$ is the sum of all estimated ω starting from t_0 and $\mathbf{f}$ is the corresponding vector containing the Fourier series parameters:

$$\mathbf{f} = [a_1, ..., a_{H_r}, b_1, ..., b_{H_r}, c_r, d_1, ..., d_{H_c}, e_1, ..., e_{H_c}]^{\mathrm{T}} \tag{3}$$

Consequently, a point of interest (POI) $\mathbf{p}$ can be modeled with $\gamma = 3 \cdot [2 \cdot ({}^{x}H_r + {}^{x}H_c) + 1 + 2 \cdot ({}^{y}H_r + {}^{y}H_c) + 1 + 2 \cdot ({}^{z}H_r + {}^{z}H_c) + 1]$ parameters plus the respiratory and cardiac frequencies, which are shared among all coordinates and points. The number of harmonics H_r and H_c among the xyz directions may vary due to differences in their motion complexity. For recursively estimating the Fourier series parameters, the Extended Kalman Filter (EKF) is employed. The EKF state vector $\mathbf{y}$ for estimating the trajectory of φ POIs $\mathbf{p} = [{}^{x}p\ {}^{y}p\ {}^{z}p]$ is composed of $(\varphi \cdot \gamma + 2)$ parameters, such that:

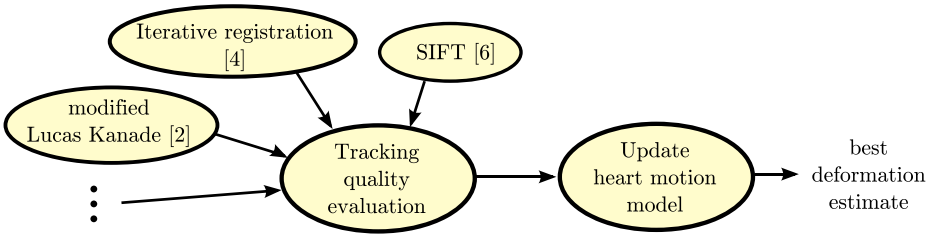


Fig. 2. A schematic overview of the hybrid visual tracking algorithm

$$\mathbf{y} = [{}^1\mathbf{f}_x, {}^1\mathbf{f}_y, {}^1\mathbf{f}_z, {}^2\mathbf{f}_x, {}^2\mathbf{f}_y, {}^2\mathbf{f}_z, ..., {}^\gamma\mathbf{f}_x, {}^\gamma\mathbf{f}_y, {}^\gamma\mathbf{f}_z, \omega_r, \omega_c]^{\mathrm{T}}; \tag{4}$$

where $[{}^i\mathbf{f}_x, {}^i\mathbf{f}_y, {}^i\mathbf{f}_z]$ are the parameter vectors of the i-th estimated POI $\mathbf{p}_i$. In the correction step of the filter, the most reliable tracking outputs are used for updating the temporal heart motion model. For more details on the predictive EKF design, see [9].

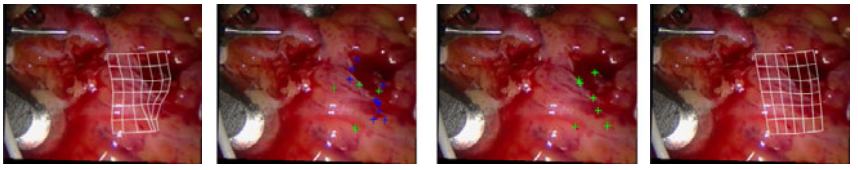
In this framework, the predicted heart motion from a moment preceding a tracking loss can be used for bridging the disturbance. As described in section 2, tracking failures are detected by the tracking quality evaluation step which indicates if the number of tracked features is insufficient for estimating the current shape of the heart. The computation of future position estimates at a given instant t can be done in a straightforward manner using equation (2), considering a stationary system within the prediction horizon. Furthermore, spatially close points can be clustered for reducing the computational effort when estimating a large number of POI.

3 Experiments and Results

The diagram in Figure 2 summarizes all steps involved in the estimation of the heart surface motion. For evaluating the performance of the proposed hybrid tracking concept, two sets of experiments on recorded images of a minimally invasive CABG using the DaVinci robotic platform have been conducted. The first set aims to compare the performance of the hybrid tracker with the individual techniques ([4,2,6]) while the second focuses on the prediction quality of the predictive EKF described in section 2.2.

3.1 Comparative Study

For the method proposed in Richa et al. [4], 14 control points distributed on textured parts of the region of interest are used. The maximal number of iterations for the minimization loop are fixed to 20. Although the method is robust to illumination changes and specular reflections, the image blur caused by large inter-frame motion can induce large errors in the estimation of the heart surface deformation (see figure 3).



(a) (b) (c) (d)

Fig. 3. (a) Tracking errors induced by motion blur and specular reflections in [4]. (b-c) Using the motion of stable features obtained with techniques [2] and [6] marked in green, respectively, a more reliable estimate of the heart surface deformation shown in (d) can be computed.

In Stoyanov et al. [2], tracking loss caused by large specular reflections and tissue deformation is the main performance issue. Figure 4 (top) indicates the significant drop in the number of tracked features while tracking (from the 30 Shi-Tomasi features initially detected within the reference region of interest, only 5 are tracked after 6 seconds of tracking).

Although a very high number of SIFT features are detected during the initialization (384 features within the region of interest), very few matches are available when to large tissue deformations occur (see Figure 4). Furthermore, 'tracking-by-detection' methods such as the SIFT are unsuitable for performing motion compensation since continuous tracking is not possible.

The proposed hybrid tracking concept offers the possibility of overcoming the failure of an individual technique using the output of others. In the example in figure 1, the hybrid tracker uses 14 control points [4], 15 Lucas-Kanade features [2] and 384 SIFT features [6]. Using the interpolating TPS surface, problems due to lost features or convergence errors can be circumvented. In addition, 'tracking-by-detection' methods such as the SIFT can be incorporated for increased tracking quality. Figure 4 shows the number of active features during tracking, demonstrating the superior performance of the proposed technique compared to existing techniques individually.

3.2 Improvements Using the Temporal Heart Motion Model

For evaluating the performance of the temporal heart motion model, the prediction errors for 0.2, 1 and 3 second prediction horizons are evaluated using the recorded tracked coordinates of the POI on the heart highlighted in Figure 5. The error is calculated as the Euclidean distance $||\boldsymbol{d} - \boldsymbol{p}||$ between the predicted $\boldsymbol{d}$ and true $\boldsymbol{p}$ positions of the POI for all xyz coordinates. The obtained root mean square and peak prediction errors at every motion sample were (0.73mm/1mm), (0.86mm/1.55mm) and (1.00mm/2.03mm) for the 0.2, 1 and 3 second prediction horizons, respectively. The low prediction errors attest the capability of the prediction scheme to overcome both short occlusions by specular reflections or motion blur and long occlusions by surgical instruments (it is important to remark that although the predicted motion is accurate enough to restart tracking,

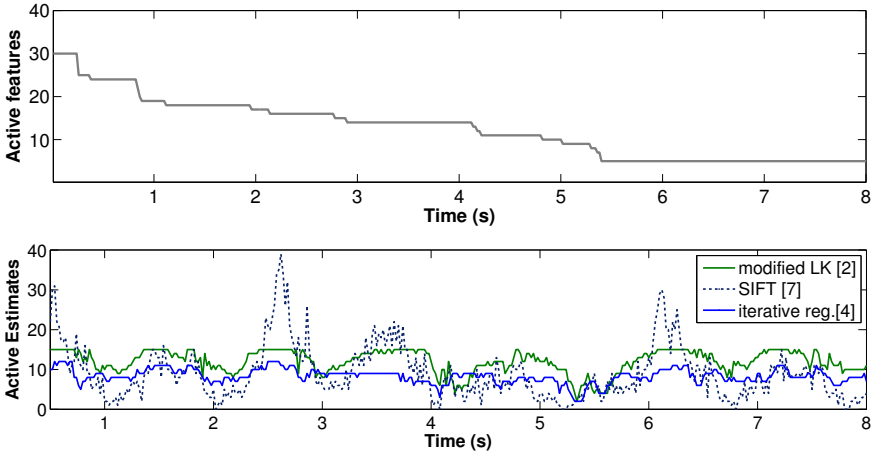


Fig. 4. (Top) The number of active tracked features using the method proposed in [2]. (Bottom) Active motion estimates during tracking using the hybrid method.

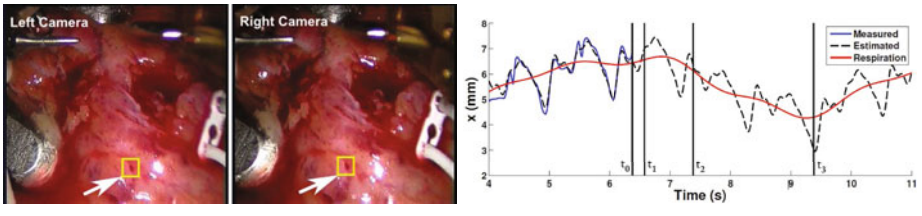


Fig. 5. The estimated dual Fourier series model of the x coordinate of the tracked feature on the heart surface highlighted on the left at $t_0 = 6.4s$. The 0.2s, 1s and 3s prediction horizons (t_1, t_2, t_3 respectively) used for evaluating the quality of the predicted heart motion are defined in the plots.

it cannot be used for motion compensation since the prediction errors can exceed the minimal precision requirements of cardiac surgery [9]).

3.3 Computational Requirements

The computational burden introduced by the tracking quality evaluation and the EKF update is negligible compared to the computational time required by a single tracking technique. In addition, all techniques used in this study [2,4,6] have successfully been implemented in Graphics Processor Units (GPU) and tracking speeds over 100 Hz have been reported. Therefore, it is expected that the even though the proposed method incorporates an additional computational burden, tracking at high speeds is possible. This is a great advantage since the future deployment on a prototype surgical platform is envisaged.

4 Conclusion

In this paper, we presented a hybrid tracker for estimating the 3D motion of the heart surface using stereo endoscopic images. The method uses multiple visual trackers working in parallel for tracking with increased robustness and spatial resolution. For coping with tracking failures and occlusions, the temporal heart motion dynamics have also been incorporated as an additional support for the visual tracking task. Experiments on recorded *in vivo* images of a minimally invasive CABG using the DaVinci robotic platform attest the superior performance of the proposed tracker in comparison with existing techniques individually. The future implementation of the tracking method in a control scheme is envisaged.

References

1. Nakamura, Y.K., Kawakami, H.: Heartbeat synchronization for robotic cardiac surgery. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA 2001), Seoul, Korea, vol. 2, pp. 2014–2019 (May 2001)
2. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
3. Mountney, P., Yang, G.Z.: Soft tissue tracking for minimally invasive surgery: Learning local deformation online. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 364–372. Springer, Heidelberg (2008)
4. Richa, R., Poignet, P., Liu, C.: Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model. The International Journal of Robotics Research (IJRR) – Special Issue on Robot Vision 29(2-3), 218–230 (2010)
5. Visentini-Scarzanella, M., Mylonas, G.P., Stoyanov, D., Yang, G.Z.: *i*-BRUSH: A Gaze-Contigent Virtual Paintbrush for Dense 3D Reconstruction in Robotic Assisted Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 353–360. Springer, Heidelberg (2009)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Stoyanov, D., Yang, G.Z.: Soft-tissue deformation tracking for robotic assisted minimally invasive surgery. In: Proceedings of IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS 2009), Boston, USA, pp. 254–257 (2009)
8. Bogatyrenko, E., Hanebeck, U.D., Szabó, G.: Heart surface motion estimation framework for robotic surgery employing meshless methods. In: Proceedings of IEEE Conference on Intelligent Robots and Systems (IROS 2009), St. Louis, USA, pp. 67–74 (2009)
9. Richa, R., Bó, A.P.L., Poignet, P.: Beating heart motion prediction for robust visual tracking. In: Proceedings of IEEE Conference on Robotics and Automation (ICRA 2010), Anchorage, USA, pp. 4579–4584 (2010)