HAL
archives-ouvertes.fr

PCA-based 3D Pose Modeling for Beating Heart Tracking

Bo Yang, Chao Liu, Wenfeng Zheng

▶ **To cite this version:**

Bo Yang, Chao Liu, Wenfeng Zheng. PCA-based 3D Pose Modeling for Beating Heart Tracking. 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017), Jul 2017, Guilin, China. pp.586-590, 10.1109/FSKD.2017.8393335 . lirmm-03130669

HAL Id: lirmm-03130669
https://hal-lirmm.ccsd.cnrs.fr/lirmm-03130669

Submitted on 3 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PCA-based 3D Pose Modeling for Beating Heart Tracking

Bo Yang
School of Automation Engineering
University of Electronic Science and Technology of China
Chengdu, China

Chao Liu
LIRMM
CNRS-UM
Montpellier, France

Wenfeng Zheng
School of Automation Engineering
University of Electronic Science and Technology of China
Chengdu, China

Abstract—**A statistical pose model is developed for efficient 3D visual tracking of beating heart. The Region of Interest (ROI) on heart surfaces is first pre-tracked with a conventional high-order thin plate spline model. The 3D pose data of the ROI extracted from the pre-tracked results are then used to train a low-order 3D pose model based on the principal component of these pose data. The low-order model is accurate, robust, and efficient for tracking subsequent heart motion as heart beats are quasi-periodic with stable statistics and the redundant degree of freedom for fitting the poses of heart surface is significantly decreased by the principal-component-based dimensionality reduction. The proposed 3D pose modeling is validated on the stereo-endoscopic videos recorded by the da Vinci® surgical system.**

Keywords-3D visual tracking; pose modeling; beating heart tracking; stereo-endoscope

I. INTRODUCTION

In minimally invasive heart surgery, heart motion considerably disturbs the precise execution of surgical procedures, resulting in longer surgery duration and more surgical risks. Active motion compensation techniques were explored to overcome this issue [1]-[3]. By tracking heart motion and actively synchronizing surgical instruments with the motion, it is possible to provide a virtually stable operating environment to surgeons. Passive visual tracking techniques with a stereo-endoscope are very appropriate for the heart motion tracking, because there are no additional instrument ports required besides the endoscope [4].

However, it is very challenging for 3D visual tracking in a highly dynamic MIS setting. Model-based methods have been developed for tracking and fitting the region of interest (ROI) of heart surface. The ROI template designated in the first frame is warped with a 3D deformable model to match regions of subsequent frames based on a similarity measure, and the tracking task is formulated into an optimization problem with respect to the model parameters. Various models were employed for fitting heart surfaces [5]-[11]. Simple models such as affine models are computationally efficient but difficult to handle complex soft-tissue deformations, while sophisticated models generally suffer from heavy computational burden and convergence problem of model parameters. In addition, most existing models used for beating heart tracking are designed for general deformations, without taking into account the statistics of specific heart motion.

In this study, the classical Principle Component Analysis (PCA) [12] is used to extract the principal pose components of a specific heart from its past pose data, on which a low-order pose model is built. The low-order model, trained specially for the specific heart, is efficient when it is incorporated into the model-based scheme for the subsequent tracking task, which is verified on the stereo-endoscopic datasets collected by the da Vinci surgical system.

II. METHODS

Let $\mathbf{m} = \begin{bmatrix} u & v \end{bmatrix}^T$ denote a pixel in a 2D image, $\Re = \{\mathbf{m}_n\}_{n=1}^{N_\Re}$ be a Region of Interest (ROI) on a heart surface, composed of N_R pixels, and $\mathbf{m}_\mathrm{o} \in \Re$ be a Point of Interest (POI), normally the center of the ROI. In our study, we use $\mathbf{p}_\mathrm{o}$, the 3D coordinates associated with $\mathbf{m}_\mathrm{o}$, to denote the 3D position of the whole ROI, and correspondingly the 3D pose of the ROI can be represented by a pose vector

$$\mathbf{s} = \begin{bmatrix} \mathbf{p}'_1 \\ \mathbf{p}'_2 \\ \vdots \\ \mathbf{p}'_{N_\Re} \end{bmatrix} \in \mathbb{R}^N \tag{1}$$

where $N = 3N_\Re$ and $\mathbf{p}'_n = \mathbf{p}_n - \mathbf{p}_\mathrm{o}$ denotes the relative 3D coordinates associated with $\mathbf{m}_n$.

For a specific beating heart, assume that we have a group of past pose vectors. Since the ROI is spatially continuous and the heart motion is quasi-periodic with stable statistics, these pose vectors at various times are correlated and sparse in the space $\mathbb{R}^N$. The PCA technique can be used to extract the principal components of these pose vectors and build a low-order statistical pose model particularly suitable for the specific heart.

Let $\{\mathbf{s}_l\}_{l=1}^L$ be L past pose vectors. A training data matrix of row-wise zero mean can be constructed as

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 - \bar{\mathbf{s}} & \mathbf{s}_2 - \bar{\mathbf{s}} & \cdots & \mathbf{s}_L - \bar{\mathbf{s}} \end{bmatrix} \tag{2}$$

This work was supported by the National Natural Science Foundation of China (Grant No. 61305022), the Science and Technology Planning Project of Sichuan Province (No. 2015HH0022), and the State Key Laboratory Virtual Reality Technology and Systems (BUAA-VR-16KF-11).

with the mean pose

$$\bar{\mathbf{s}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{s}_l \tag{3}$$

The principal components of the $N \times L$ data matrix can be computed by using the thin Singular Value Decomposition (SVD)

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{4}$$

where $\boldsymbol{\Sigma}$ is a $N \times L$ diagonal matrix of singular values $\{\sigma_k\}_{k=1}^{\min\{N,\,L\}}$ with $\sigma_k \geq \sigma_{k+1}$, $\mathbf{U}$ is a $N \times N$ unitary matrix of orthonormal columns $\{\mathbf{u}_k\}_{k=1}^{N}$, and $\mathbf{V}$ is a $L \times L$ unitary matrix. The column $\mathbf{u}_k$ of $\mathbf{U}$, corresponding to the singular value σ_k, is the *k*-th principal component of the data matrix.

According to the Eckart–Young theorem [13], the pose data can be reconstructed optimally with the first few principal components that correspond to the largest singular values. Let $\hat{\mathbf{S}} = \mathbf{U}_{K'}\boldsymbol{\Sigma}_{K'}\mathbf{V}_{K'}^T$ be the reconstructed matrix with the first $K' \leq \min\{N, L\}$ principal components, where $\mathbf{U}_{K'} = [\mathbf{U}]_{K'}$ is formed by the first K' columns of $\mathbf{U}$ and $\boldsymbol{\Sigma}_{K'}\mathbf{V}_{K'}^T = \mathbf{U}_{K'}^T\mathbf{S}$ is the scores of the data matrix on these columns (principal components). The reconstruction error defined with the Frobenius norm can be measured by the sum of square of the smallest singular values being omitted.

$$\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_F^2 = \sum_{k=K'+1}^{\min\{N,L\}}\sigma_k^2 \tag{5}$$

As the principal components corresponding to the largest singular values are reserved, the information loss caused by the dimension-reduced representation is minimal, of all possible K'-D subspace representations in $\mathbb{R}^N$.

With the assumption that the statistics of heart beats are stable, the poses in the future can be represented effectively in the dimension-reduced subspace with the statistical pose model trained on the past pose data

$$\hat{\mathbf{s}} = \mathbf{U}_{K'}\mathbf{w} + \bar{\mathbf{s}} \tag{6}$$

where $\mathbf{w}$ is a new pose coefficient vector.

In order to obtain the pose training data, the classical TPS model [8] may be employed to pre-track the ROI in the first *L* frames of a stereo-endoscopic image sequence. So far we can summarize the complete steps of beating heart tracking based on the proposed pose modeling:

1) The ROI in the first *L* frames is pre-tracked with a high-order TPS model of *K* deformation coefficients, yielding the training data matrix $\mathbf{S}$ and the mean pose $\bar{\mathbf{s}}$.

2) $\mathbf{S}$ is decomposed by the thin SVD, $\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$.

3) The first $K' \leq (K-3)$ columns of $\mathbf{U}$ are used to build a new low-order deformable model of K' pose coefficients and 3 position coefficients, i.e., $\forall\, \mathbf{m} \in \Re$, its corresponding 3D coordinate can be calculated by

$$\mathbf{p} = \mathbf{U}_{K'}(\mathbf{m})\mathbf{w} + \bar{\mathbf{s}}(\mathbf{m}) + \mathbf{p}_{\mathrm{o}}, \tag{7}$$

where $\mathbf{U}_{K'}(\mathbf{m})$ / $\bar{\mathbf{s}}(\mathbf{m})$ is sub-matrix/vector of $\mathbf{U}_{K'}$ / $\bar{\mathbf{s}}$ associated with $\mathbf{m}$, $\mathbf{w}$ is the K'-D pose coefficient vector and $\mathbf{p}_{\mathrm{o}}$ is the 3D position coefficient vector.

4) The subsequent image frames are tracked with the low-order model in (7).

By using the pinhole camera model [11], the model-based tracking involved in the steps 1) and 4) can be finally translated into an optimization problem of model coefficients, i.e., finding the best deformation (pose and position) coefficients that minimize the alignment error of the ROI between the template image and the current stereo images.

For a stereo vision system, the 2D projection of $\mathbf{p}$ onto the left/right image plane can be located with the pinhole model as:

$$\mathbf{m}^{(\mathrm{L/R})} = \rho\left(\mathbf{C}_{\mathrm{L/R}}\begin{bmatrix}\mathbf{p}\\ 1\end{bmatrix}\right) \tag{8}$$

where $\mathbf{C}_{\mathrm{L/R}}$ is the 3×4 left/right camera matrix and $\rho: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ denotes the mapping from homogeneous coordinates to pixel coordinates, i.e.

$$\begin{bmatrix}a/c\\ b/c\end{bmatrix} = \rho\left(\begin{bmatrix}a\\ b\\ c\end{bmatrix}\right) \tag{9}$$

Given the ROI $\Re$ in the template image, the 3D tracking problem can be formalized with the sum of squared pixel differences as

$$\arg\min_{\boldsymbol{\theta}}\left(\begin{Vmatrix}\mathbf{I}_{\mathrm{L}} - \mathbf{I}_{\mathrm{T}}\\ \mathbf{I}_{\mathrm{R}} - \mathbf{I}_{\mathrm{T}}\end{Vmatrix}^2\right) \tag{10}$$

where

$$\mathbf{I}_{\mathrm{T}} = \begin{bmatrix}I_{\mathrm{T}}(\mathbf{m}_1)\\ I_{\mathrm{T}}(\mathbf{m}_2)\\ \vdots\\ I_{\mathrm{T}}(\mathbf{m}_{N_\Re})\end{bmatrix} \text{ and } \mathbf{I}_{\mathrm{L/R}} = \begin{bmatrix}I_{\mathrm{L/R}}(\mathbf{m}_1^{(\mathrm{L/R})})\\ I_{\mathrm{L/R}}(\mathbf{m}_2^{(\mathrm{L/R})})\\ \vdots\\ I_{\mathrm{L/R}}(\mathbf{m}_{N_\Re}^{(\mathrm{L/R})})\end{bmatrix} \tag{11}$$

are $N_\Re$-D column vectors stacked by the pixel values of the ROI in the template image I_{T} and the current stereo images $I_{\mathrm{L/R}}$, and $\boldsymbol{\theta}$ denotes model coefficients, e.g., there is $\boldsymbol{\theta} = (\mathbf{w},\ \mathbf{p}_o)$ for the low-order model in (7).

To solve (11), the Efficient Second-order Minimization (ESM) [14] is employed, which has a faster convergence rate and a larger convergence basin than traditional optimization. The model coefficients are updated with the increment

$$\Delta\boldsymbol{\theta} = -2\begin{bmatrix} \mathbf{J}_{\mathbf{I}_L}(\boldsymbol{\theta}) + \mathbf{J}_{\mathbf{I}_L}(\boldsymbol{\theta}^*) \\ \mathbf{J}_{\mathbf{I}_R}(\boldsymbol{\theta}) + \mathbf{J}_{\mathbf{I}_R}(\boldsymbol{\theta}^*) \end{bmatrix}^{+} \begin{bmatrix} \mathbf{I}_L - \mathbf{I}_T \\ \mathbf{I}_R - \mathbf{I}_T \end{bmatrix} \quad (12)$$

where "+" denotes the matrix pseudo-inverse, $\boldsymbol{\theta}^*$ denotes the ideal coefficients and $\mathbf{J}_{\mathbf{Y}}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{Y}$ at $\mathbf{x}$.

Taking the statistics of heart beats into account, the low-order model trained online can be efficient in computation and robust against the "aperture" problem, when incorporating into the model-based tracking scheme. The order of the pose model (K') is adjustable according to the statistics of past pose data, which can be indicated by the energy distribution of singular values, as shown in Section 3.

III. RESULTS

Two stereo-endoscope video datasets captured by the da Vinci (Intuitive Surgical, CA) surgical platform are used to validate the proposed modelling method. Each video consists of 800 frame stereo-pair images with a frame rate of 25 Hz. Video I records the motion of a phantom heart [15]. Video II is captured from in an in vivo heart in a real TECAB procedure [16], [17].

A. PCA of training data

In order to show the statistics of heart beats, the TPS model of 9 Control Points (CP) is employed to track an 120×120 ROI in each video for 800 frames, obtaining a pose data matrix for each dataset. During the tracking, the ESM algorithm is employed to iteratively compute the 27 coefficients of the 9-CP TPS model at each frame. Since the ESM algorithm cannot follow all 800 frames without interruption in the in vivo dataset, the tracking process is restarted manually when it is interrupted.

Figure 1 shows the squares of 24 non-zero singular values computed from the pose matrices of the two datasets. The energy is concentrated in the several larger singular values for both datasets, showing that the degree of freedom of the 9-CP TPS model is redundant for pose representation. Consequently, it is feasible to reconstruct the poses on several principal components that correspond to the largest singular values.

B. Tracking Tests

In our tracking tests, the first $L = 600$ frames in each datasets, tracked with the 9-CP TPS, serve as the training data, and the subsequent 200 frames serve as the test data, which will be tracked with the low-order model built on the PCA of the training data. We choose $K' = 8 / K' = 12$ for Video I/II, such that the normalized reconstruction error for each dataset satisfies

$$\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_F^2 \Big/ \left\|\mathbf{S}\right\|_F^2 < 0.01$$

As the ground truth is unknown, 3D tracking results are usually evaluated qualitatively by visually inspecting the concordance of the reconstructed surfaces and heart beatings, and by analyzing the motion trajectories of POI on heart surfaces. Figure 2 illustrates some tracked frames with the PCA-based model. The proposed model can efficiently handle soft-tissue deformations on heart surfaces. The reconstructed 3D poses at different instants of heart beating are consistent with those perceived directly by human vision, and the POI (the center of the ROI) is located accurately at different image frames.

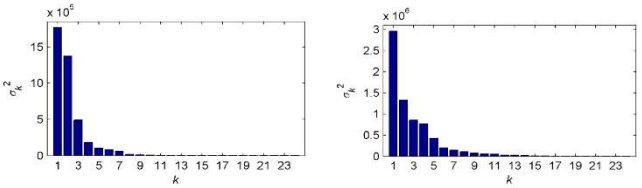

Fig. 1 Energy distributions of singular value computed from: (a) Video I, (b) Video II.

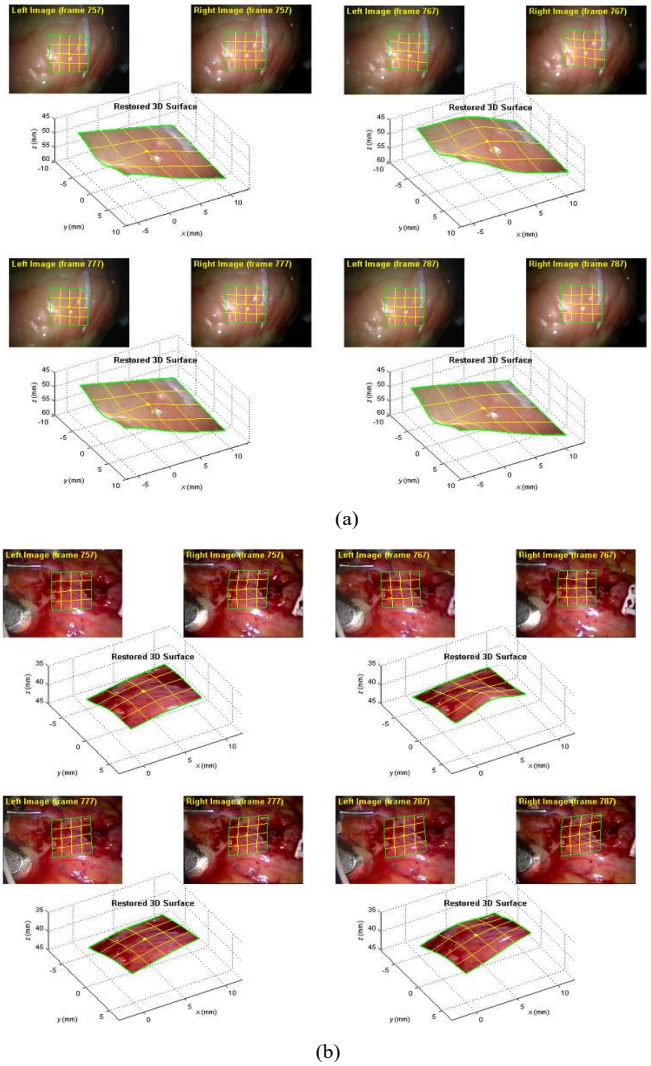

Figure 2: Tracked frames with the PCA-based model from: (a) Video I, (b) Video II.

Figure 3 shows the motion trajectories of POI on Video I tracked by the PCA-based model in x, y, z directions. The Power Spectral Densities (PSDs) of these trajectories are also calculated. Since Video I is captured from a phantom heart, the motion trajectories exhibits highly regular periodicity, and their

energy mainly concentrates at 1.5 Hz and its second (3.0 Hz) and third (4.5 Hz) harmonics. These results are consistent with the given motion rate of the phantom model, showing the accuracy of the proposed tracking algorithm.

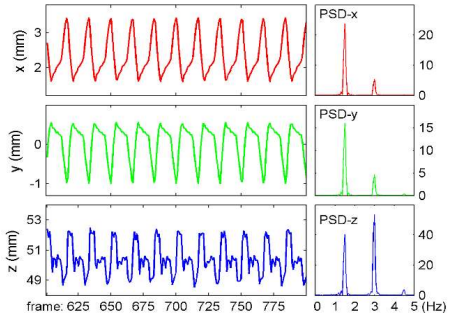

Fig.3 POI motion trajectory and PSD tracked from Video I.

The PCA-based model is compared with the TPS models of various CP in terms of tracking result, robustness and efficiency. The PCA-based model and the 9-CP TPS yield very similar 3D surfaces on both datasets, where the 3D details on heart surfaces are preserved (see Fig. 2). The 3D surfaces tracked by the 4-CP TPS seem over smooth, as shown in Fig. 4, where many 3D details on phantom surfaces are missed, even though the order ($3\times4=12$ coefficients) of the 4-CP TPS is larger than the order ($K'+3=11$ coefficients) of the PCA-based model built for tracking Video I. These comparisons show the effectiveness of the proposed PCA-based pose modeling approach.

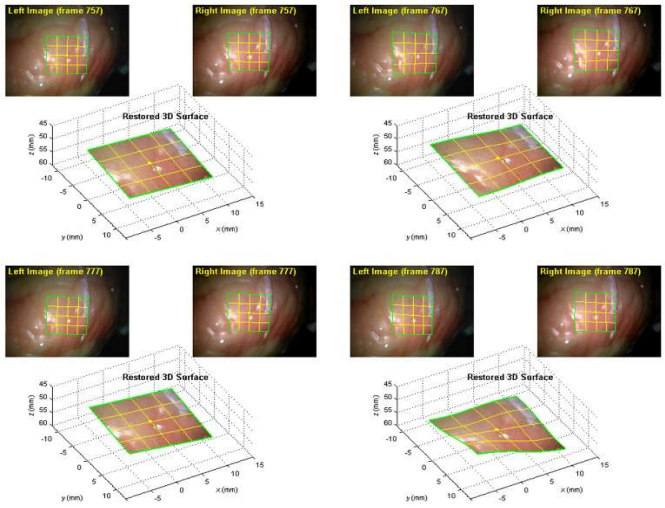

Fig. 4 Tracked frames with the 4-CP TPS model from Video I.

As tracking beating heat with stereo-endoscopic images is very challenging, tracking chain for a continuous image sequence is often interrupted by the dynamic effects in MIS, such as instrument occlusion, specular reflection, and motion blurring. The robustness of tracking method is very important for real MIS settings, which can be evaluated with the average number of continuous frames tracked successfully. The PCA-based model can continuously track all 200 test frames on both videos, performing much more robust than the high-order 9-CP TPS, especially on the in vivo dataset. The low-order models with fewer coefficients, such as the PCA-based, 4-CP and 5-CP TPS models, are easier to find optimal solutions in noisy environments. The PCA-based model also performs slightly better than the 5-CP TPS model on the in vivo dataset, though the two models have the same order (15 coefficients). The comparisons in robustness further show the effectiveness of the proposed modeling approach.

The computational efficiency of the ESM-based tracking algorithms mainly depends on the orders of the models employed. The most time-consuming operation in the ESM is the pseudo-inverse calculation of the Jacobian matrix in (12). Given a *K*-order model on a *N*-pixel ROI, the computational complexity of the pseudo-inverse calculation is $O(NK^2)$. Consequently, for tracking the same ROI, the computational cost of the PCA-based model is about 1/3 of that of the 9-CP TPS, though the tracking results yielded by the two models are very similar and the robustness of the PCA-based model is even much better than that of the 9-CP TPS.

IV. Conclusion

In this study, we propose a PCA-based pose modelling approach for 3D beating heart tracking. The training data are first obtained by using a general high-order TPS model. The principal components of these 3D pose data are then extracted with the PCA. The first few principal components that correspond to the largest singular values are used to build a low-order deformable model online, which is finally incorporated into the ESM-based tracking scheme for efficient beating heart tracking. The experimental results on the phantom and in vivo heart videos captured by the da Vinci surgical robots show the effectiveness of the proposed modeling approach.

References

[1] Y. Nakamur, K. Kishi, and H. Kawakami, "Heartbeat synchronization for robotic cardiac surgery", IEEE ICRA, Seoul, Korea, 2001, pp. 2014-2019.

[2] T. Ortmaier, M. Groger, D. H. Boehm, V. Falk, and G. Hirzinger, "Motion estimation in beating heart surgery," IEEE Trans. Biomed. Eng., vol. 52, pp. 1729-1740, Oct. 2005.

[3] S. G. Yuen, D. T. Kettler, P. M. Novotny, R. D. Plowes, and R. D. Howe, "Robotic motion compensation for beating heart intracardiac surgery," Int. J. Robot. Res., vol. 28, pp. 1355-1372, Oct. 2009.

[4] P. Mountney, D. Stoyanov, and G. Z. Yang, "Three-dimensional tissue deformation recovery and tracking," IEEE Signal Process. Mag. Vol. 27, pp. 14-24, Jun. 2010.

[5] W. Lau, N. Ramey, J. Corso, N. Thakor, and G. Hager, "Stereo-based endoscopic tracking of cardiac surface deformation," MICCAI, Saint-Malo, France, 2004, pp. 494–501.

[6] D. Stoyanov, A. Darzi, and G. Z. Yang, "A practical approach towards accurate dense 3-D depth recovery for robotic laparoscopic surgery," Comput. Aid. Surg., vol. 10, pp. 199-208, Jul. 2005.

[7] T. Bader, A. Wiedermann, K. Roberts, et al., "Model-based motion estimation of elastic surfaces for minimally invasive cardiac surgery," IEEE ICRA, Roma, Italy, 2007, pp. 2261-2266.

[8] R. Richa, P. Poignet, and C. Liu, "Three-dimensional motion tracking for beating heart surgery using a thin-plate spline deformable model," Int. J. Robot. Res., vol. 29, pp. 218-230, Feb. 2010.

[9] E. Bogatyrenko, P. Pompey, and U. D. Hanebeck, "Efficient physics-based tracking of heart surface motion for beating heart surgery robotic systems," Int. J. Comput. Assist. Radiol. Surg, vol. 6, pp. 387-399, May 2011.

[10] B. Yang, W. K. Wong, C. Liu, and P. Poignet, "3D soft-tissue tracking using spatial-color joint probability distribution and thin-plate spline model," Pattern Recognition, vol. 47, pp. 2962-2973, Sep. 2014.

[11] B. Yang, C. Liu, K. Huang, and W. Zheng, "A triangular radial cubic spline deformation model for efficient 3D beating heart tracking," Signal Image and Video Processing, Apr. 2017. (pulished online, doi: 10.1007/s11760-017-1090-y)

[12] I. Jolliffe, "Principal component analysis," John Wiley & Sons, Hoboken, 2002, pp. 433-459.

[13] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika, vol. 1, pp. 211-218, Sep. 1936.

[14] S. Benhimane and E. Malis, "Homography-based 2D visual tracking and servoing," Int. J. Robot. Res, vol. 26, pp. 661-676, July. 2007.

[15] D. Stoyanov, G. Mylonas, F. Deligianni, A. Darzi, and G. Z. Yang, "Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures," MICCAI, Palm Springs, CA, USA, 2005, pp. 139-146.

[16] D. Stoyanov, M. Visentini-Scarzanella, P. Pratt, and G. Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," MICCAI, Beijing, China, 2010, pp. 275-82.

[17] P. Pratt, D. Stoyanov, M. Visentini-Scarzanella, and G. Z. Yang, "Dynamic guidance for robotic surgery using image-constrained biomechanical models," MICCAI, Beijing, China, 2010, pp. 77-85.

View publication stats