Research Article Vol. 9, No. 12 | 1 Dec 2018 | BIOMEDICAL OPTICS EXPRESS 6222

Biomedical Optics EXPRESS

Reconstructing a 3D heart surface with stereo-endoscope by learning eigen-shapes

Bo Yang,[1,*] Chao Liu,[2] Wenfeng Zheng,[1] Shan Liu,[1] and Keli Huang[3]

[1] *School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[2] *LIRMM, CNRS-UM, Montpellier, France*
[3] *Cardiac Surgery Center, Sichuan Provincial People's Hospital, Chengdu, China*
[*] *boyang@uestc.edu.cn*

Abstract: An efficient approach to dynamically reconstruct a region of interest (ROI) on a beating heart from stereo-endoscopic video is developed. A ROI is first pre-reconstructed with a decoupled high-rank thin plate spline model. Eigen-shapes are learned from the pre-reconstructed data by using principal component analysis (PCA) to build a low-rank statistical deformable model for reconstructing subsequent frames. The linear transferability of PCA is proved, which allows fast eigen-shape learning. A general dynamic reconstruction framework is developed that formulates ROI reconstruction as an optimization problem of model parameters, and an efficient second-order minimization algorithm is derived to iteratively solve it. The performance of the proposed method is finally validated on stereo-endoscopic videos recorded by da Vinci robots.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

In robotic-assisted off-pump heart surgery, heart beating considerably influences the accuracy of surgical operations, resulting in longer surgical duration and increased surgical risks. By measuring heart motion and actively synchronizing surgical instruments with this motion, a technique called active motion compensation, it is possible to provide a virtually stable operating environment to surgeons [1, 2]. Passive vision using stereo-endoscope, compared with other sensing techniques such as structured lighting [3, 4] and ultrasound [5], is more appropriate for measuring heart motion in a Minimally Invasive Surgery (MIS) because no additional instrument ports are required, as indicated in [6]. The 3D reconstruction with endoscope is also important for other advanced surgical techniques, e.g. augmented reality guidance [7], and multispectral [8] or multimodal [9] imaging. In addition, it is useful for offline applications as well, such as surgery simulation [10] and visual medical record [11], for learning, training or evaluation purposes.

However, in a highly dynamic MIS, it is very challenging to track and reconstruct Regions of Interest (ROIs) with complex soft-tissue deformations from real-time endoscopic videos. Model-based methods have been explored [12–18]. With a parameterized deformable model, an initial ROI is warped as a template to match pixels at subsequent stereo-pair frames. Model complexity (rank) can be simply measured by the number of model parameters. Low-rank models, such as rigid and affine models, are robust and computationally efficient but difficult to deal with complex soft-tissue deformations, while high-rank models generally suffer from problems of parameter convergence and heavy computational burden, difficult to meet real-time and robustness requirements, as indicated in [16].

On the other hand, most existing models, typically based on the mesh [2, 11, 13, 19] or spline [12, 14, 16–18], are designed for general deformations with certain continuity and smoothness constraints, which do not take into account the statistics and specificities of ROIs. Statistical shape models have been explored to reconstruct static anatomical structures, such as bone surfaces from X-ray images [20] and employed as *a priori* information in a wide range

#346135 https://doi.org/10.1364/BOE.9.006222
Journal © 2018 Received 17 Sep 2018; revised 26 Oct 2018; accepted 2 Nov 2018; published 13 Nov 2018

Biomedical Optics EXPRESS

of segmentation tasks from CT [21], MRI [22], 3D ultrasound [23], and optical coherence tomography [24, 25]. The training shapes in the literature are usually from CT databases or obtained manually by experts from pre-operative imaging. To the author's knowledge, statistical modeling in full 3D space for highly dynamic soft-tissues from stereo-endoscope is still open. It is almost impossible to establish a valid pre-operative shape database for beating heart. The shape similarity of anatomical structures from a given population, which has been used for statistical modeling in the literature, is unavailable for modeling beating heart due to high dynamics of cardiac soft-tissues.

In this work, we first derive a general 3D shape representation of ROI by decoupling a Thin Plate Spline (TPS) mapping into shape and position components. The high-rank version of the decoupled TPS is employed to pre-reconstruct a ROI for a number of frames to obtain ROI-specific shape data. Eigen-shapes are then learnt from these data, on which a low-rank 3D model is built and used to reconstruct subsequent frames. The linear transferability of Principal Component Analysis (PCA) is proved, which significantly reduces the computational burden of eigen-shape learning and thus makes real-time statistical modeling possible. A general model-based reconstruction framework with an iterative algorithm for estimating model parameter is presented. The feasibility and effectiveness of the proposed method is finally validated on stereo-endoscopic videos recorded by da Vinci (Intuitive Surgical, CA) robots.

2. Methods

2.1. Representing 3D shape with TPS

The TPS has been widely used in image alignment and shape matching, and recently used to handle soft-tissue deformations [14, 17]. We extend and decouple the classical TPS mapping into two independent components related to shape and position, respectively.

Let $\mathbf{m} \in \mathbb{R}^2$ denote pixel coordinates. The TPS of K Control Points (CPs) is written as

$$f(\mathbf{m}) = \boldsymbol{\varphi}(\mathbf{m})\,\boldsymbol{\alpha} + \begin{bmatrix}\mathbf{m}^{\mathrm{T}} & 1\end{bmatrix}\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\varphi}(\mathbf{m})$ is a row vector of elements $\{\varphi(\|\mathbf{m}-\mathbf{c}_k\|)\}_{k=1}^{K}$ with $\varphi(r) = r^2 \ln r$ and $\mathbf{c}_k$ being the k-th CP, and $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ are parameter vectors associated with non-affine and affine transformations, respectively. All vectors are column vectors, unless otherwise specified. To obtain square-integrable second derivatives, the following constraint is imposed.

$$\begin{bmatrix}\tilde{\mathbf{c}}_1 & \tilde{\mathbf{c}}_2 & \cdots & \tilde{\mathbf{c}}_K\end{bmatrix}\boldsymbol{\alpha} = \mathbf{0} \tag{2}$$

where $\tilde{\mathbf{x}}$ denotes the augmented vectors of $\mathbf{x}$ by adding 1 as the last element. By substituting it into Eq. (1), we can re-parameterize the non-affine part of the TPS as $\boldsymbol{\varphi}'(\mathbf{m})\boldsymbol{\alpha}'$, where $\boldsymbol{\alpha}'$ is a $K-3$ dimensional subvector of $\boldsymbol{\alpha}$ (three elements of $\boldsymbol{\alpha}$ are eliminated).

Let $\mathbf{p} \in \mathbb{R}^3$ denote 3D coordinates in the real space. By separately mapping in the x, y, z directions, the TPS can be extended to $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ that maps $\mathbf{m}$ to $\mathbf{p}$ with the parameter vectors $\{\boldsymbol{\alpha}'_{\mathrm{x}}, \boldsymbol{\beta}_{\mathrm{x}}, \boldsymbol{\alpha}'_{\mathrm{y}}, \boldsymbol{\beta}_{\mathrm{y}}, \boldsymbol{\alpha}'_{\mathrm{z}}, \boldsymbol{\beta}_{\mathrm{z}}\}$, which contain $3K$ model parameters (scalars) in total.

Let $\mathcal{R} = \{\mathbf{m}_n\}_{n=1}^{N}$ be a ROI. Assume that $\mathbf{m}_{\mathrm{o}}$ is the center of $\mathcal{R}$, and its 3D counterpart $\mathbf{p}_{\mathrm{o}}$ can denote the position of $\mathcal{R}$. The relative 3D coordinates of N pixels with respect to $\mathbf{p}_{\mathrm{o}}$ can represent the 3D shape of $\mathcal{R}$ that is position-independent. Let $\mathbf{p}' = \mathbf{p} - \mathbf{p}_{\mathrm{o}}$,

$$\mathbf{p}' = \begin{bmatrix}\mathbf{b}(\mathbf{m}) & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{b}(\mathbf{m}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{b}(\mathbf{m})\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{x}}\\ \boldsymbol{\theta}_{\mathrm{y}}\\ \boldsymbol{\theta}_{\mathrm{z}}\end{bmatrix} = \mathbf{B}(\mathbf{m})\,\boldsymbol{\theta}, \tag{3}$$

Biomedical Optics EXPRESS

where $\mathbf{b}(\mathbf{m}) = \left[\boldsymbol{\varphi}'(\mathbf{m}) - \boldsymbol{\varphi}'(\mathbf{m}_\mathrm{o}),\ (\mathbf{m} - \mathbf{m}_\mathrm{o})^\mathrm{T}\right]$, and $\boldsymbol{\theta}_{\mathrm{x/y/z}}$ is a parameter vector stacked by $\boldsymbol{\alpha}'_{\mathrm{x/y/z}}$ and the first two elements of $\boldsymbol{\beta}_{\mathrm{x/y/z}}$. Accordingly, the 3D shape of the ROI is represented as

$$\mathbf{s} = \mathbf{B}\boldsymbol{\theta}, \tag{4}$$

where $\mathbf{s}$ is a N'-dimensional shape vector stacked by $\{\mathbf{p}'_n\}_{n=1}^N$ with $N' = 3N$, and $\mathbf{B}$ is a $N' \times K'$ transformation matrix stacked by $\{\mathbf{B}(\mathbf{m}_n)\}_{n=1}^N$ with $K' = 3(K-1)$.

The shape and position parameters $\{\boldsymbol{\theta}, \mathbf{p}_\mathrm{o}\}$ together determine a 3D surface in the real space, and the decoupled TPS model can be rewritten as $\mathbf{p} = \mathbf{B}(\mathbf{m})\boldsymbol{\theta} + \mathbf{p}_\mathrm{o}$.

2.2. Low-rank shape modeling based on eigen-shapes

With enough CPs, the TPS model can accurately describe any 3D surface. Such general models, however, are inefficient for the ROI reconstruction of beating heart. Since heart motion is normally quasi-periodic with stable statistics [26], it is possible to learn a low-rank model for a specific ROI from its past shapes. For this, the PCA is applied to the shape data.

Assume that $\{\mathbf{s}_l\}_{l=1}^L$ are past shapes. A shape data matrix of row-wise zero mean is built:

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 - \bar{\mathbf{s}} & \mathbf{s}_2 - \bar{\mathbf{s}} & \cdots & \mathbf{s}_L - \bar{\mathbf{s}} \end{bmatrix}, \quad \bar{\mathbf{s}} = \sum_{l=1}^{L} \mathbf{s}_l / L \tag{5}$$

The PCA of $\mathbf{S}$ can be computed by using the Eigen Decomposition (ED):

$$\mathbf{S}\mathbf{S}^\mathrm{T} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\mathrm{T}, \tag{6}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of the eigenvalues $\{\lambda_j\}_{j=1}^{N'}$ sorted in descending order, and the columns of $\mathbf{U}$ are the corresponding eigenvectors. Each eigenvector identifies a specific shape in the 3D space, so we call it *eigen-shape*.

According to the Eckart-Young theorem, a data matrix can be optimally rebuilt on the first few eigenvectors. Assume that $\hat{\mathbf{S}}$ is the rebuilt matrix on the first J eigenvectors. The rebuilding error defined with the Frobenius norm is

$$\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_\mathrm{F}^2 = \sum_{j=J+1}^{N'} \lambda_j \tag{7}$$

Since the eigenvectors corresponding to the largest eigenvalues are reserved, information loss caused by $\hat{\mathbf{S}}$ is minimal of all possible J-rank rebuilding. In fact, the rebuilt data may be closer to ground truth, as the original data are usually contaminated by measurement noise, which is more likely to rest on the discarded eigenvectors.

For quasi-periodic heartbeat, the shape vectors of a ROI at different moments are correlated and sparse-distributed in the high-dimensional shape vector space. Accordingly, the past shapes of the ROI can be recovered accurately on the principal eigen-shapes. With the assumption of stable statistics, the shapes of the ROI in the future can also be reconstructed accurately with

$$\mathbf{s} = \mathbf{U}_J\mathbf{w} + \bar{\mathbf{s}}, \tag{8}$$

where $\mathbf{U}_J$ is the submatrix of $\mathbf{U}$ formed by the first J eigenvectors, and $\mathbf{w}$ is a new parameter vector of J dimension. Compared with the general shape model in Eq. (4), the new model in Eq. (8) is more efficient as statistics of the ROI are introduced.

However, the PCA for a large matrix is computationally expensive in both time and memory. The complexity for analyzing $\mathbf{S}$ is $\mathrm{O}\left(N'L^2\right)$ by assuming $L \le N'$. It is impractical for most real-time systems, as N' and L are very large for a typical reconstruction task. Here, we develop a fast eigen-shape algorithm, which reduces the complexity to $\mathrm{O}\left(LK'^2\right)$ where $K' \ll L$.

Biomedical Optics EXPRESS

Since rank $(\mathbf{B}) = K'$, for the shape matrix $\mathbf{S}$ there are at most K' valid eigen-shapes that correspond to non-zero eigenvalues. Assume that $\mathbf{\Theta}$ is the shape-parameter matrix of $\mathbf{S}$, i.e. $\mathbf{S} = \mathbf{B\Theta}$. Let $\mathbf{U}_\Theta \mathbf{\Lambda}_\Theta \mathbf{U}_\Theta^{\mathrm{T}}$ be the ED of $\mathbf{\Theta\Theta}^{\mathrm{T}}$, then it can be easily proved that:

Proposition 1: $\mathbf{\Lambda}_\Theta$ and $\mathbf{BU}_\Theta$ are the non-zero eigenvalues and eigenvectors of $\mathbf{SS}^{\mathrm{T}}$, respectively, if and only if $\mathbf{B}$ is column-orthonormal.

The proposition shows that with an orthonormal transformation matrix (called transferring condition), the PCA can be transferred from a large data matrix to a small parameter matrix. Unfortunately, the shape transformation matrix $\mathbf{B}$ of the TPS is not orthonormal, which distorts the mapping from the eigen-parameters ($\mathbf{U}_\Theta$) to eigen-shapes ($\mathbf{U}$).

To overcome this problem, we orthonormalize $\mathbf{B}$ with the QR decomposition $\mathbf{B} = \mathbf{B'R}$, and re-parameterize Eq. (4) as $\mathbf{s} = \mathbf{B'\theta'}$. Correspondingly, the parameter matrix $\mathbf{\Theta'}$ is built. By using Proposition 1, the eigen-shapes can be computed fast with $\mathbf{B'U}_{\Theta'}$, and the eigenvalues are obtained from $\mathbf{\Lambda}_{\Theta'}$ directly. The computational complexity is hence reduced to $\mathrm{O}(LK'^2)$.

Though proposed for shape modeling, Proposition 1 is useful to other PCA tasks associated with linear systems. We can extend it from the point view of function composition:

Proposition 2: Given two real vector spaces that are isomorphic, there is a linear function $\mathcal{F} : \mathbf{X} \mapsto \mathbf{B}_{\mathbb{V}}\mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathbf{X}$, mapping one vector set (matrix) $\mathbf{X}$ in $\mathbb{U}$ to another in $\mathbb{V}$, where $\mathbf{B}_{\mathbb{U}/\mathbb{V}}$ is a transformation matrix formed by an orthonormal basis of $\mathbb{U}/\mathbb{V}$. Let $\mathcal{P} : \mathbf{X} \mapsto \mathbf{U}_{\mathrm{X}}$ and $\mathcal{S} : \mathbf{X} \mapsto \mathbf{\Lambda}_{\mathrm{X}}$ be the functions of calculating eigenvectors and eigenvalues from a vector set, respectively. Then there are $\mathcal{P} \circ \mathcal{F} = \mathcal{F} \circ \mathcal{P}$ and $\mathcal{S} \circ \mathcal{F} = \mathcal{S}$ for any vector set in $\mathbb{U}$.

The proof is given in Appendix A. Proposition 2 shows that between any two isomorphic real vector spaces, we can always find an invertible linear transformation (*bijection*) that is commutative with the eigenvector operation and transparent to the eigenvalue operation, such that we can map the eigenvectors from one space to another and keep the eigenvalues unchanged.

2.3. Model-based tracking and reconstruction

The past shapes can be collected with high computational cost and even manual intervention by employing a high-rank TPS for pre-reconstruction. The shape transformation matrix of the TPS can be orthonormalized offline because it is constant for any ROI of given size.

The complete steps of the proposed reconstruction method are summarized as follows:

1) Given a ROI ($\mathcal{R}$) in the template image, the shape transformation matrix of a K-CP TPS is constructed and orthonormalized with $\mathbf{B'} = \mathbf{BR}^{-1}$.

2) The first L frames are pre-reconstructed with the orthonormalized TPS model, written as

$$\forall\, \mathbf{m} \in \mathcal{R}, \quad \mathbf{p} = \mathbf{B'}(\mathbf{m})\,\mathbf{\theta'} + \mathbf{p}_{\mathrm{o}}, \tag{9}$$

where $\mathbf{B'}(\mathbf{m})$ is the submatrix of $\mathbf{B'}$ associated with $\mathbf{m}$. As a result, $\{\mathbf{\theta}'_l\}_{l=1}^{L}$ are saved.

3) The parameter matrix $\mathbf{\Theta'}$ is built and decomposed by $\mathbf{\Theta'\Theta'}^{\mathrm{T}} = \mathbf{U}_{\Theta'}\mathbf{\Lambda}_{\Theta'}\mathbf{U}_{\Theta'}^{\mathrm{T}}$.

4) A low-rank deformable model is built on the first J columns of $\mathbf{U}_{\Theta'}$ (denoted as $[\mathbf{U}_{\Theta'}]_J$),

$$\forall\, \mathbf{m} \in \mathcal{R}, \quad \mathbf{p} = \mathbf{U}_J(\mathbf{m})\,\mathbf{w} + \bar{\mathbf{s}}(\mathbf{m}) + \mathbf{p}_{\mathrm{o}}, \tag{10}$$

where $\mathbf{U}_J(\mathbf{m}) = \mathbf{B'}(\mathbf{m})\,[\mathbf{U}_{\Theta'}]_J$ and $\bar{\mathbf{s}}(\mathbf{m}) = \mathbf{B'}(\mathbf{m})\,\bar{\mathbf{\theta}}'$.

5) The subsequent frames from $L+1$ are reconstructed with the new model in Eq. (10).

With a calibrated stereo-endoscope, the ROI reconstruction involved in the steps 2) and 5) can be converted into an optimization problem, i.e., finding the best model parameters that minimize the alignment error of ROI between the template image and the current stereo-pair images.

We discuss the problem, without loss of generality, based on a general $\mathbb{R}^2 \to \mathbb{R}^3$ model:

$$\mathbf{p} = \mathbf{T}(\mathbf{m})\,\mathbf{\xi} + \mathbf{d}(\mathbf{m}), \tag{11}$$

where $\mathbf{\xi}$ is the parameter vector, $\mathbf{T}(\mathbf{m})$ and $\mathbf{d}(\mathbf{m})$ are the transformation matrix and offset vector associated with $\mathbf{m}$, respectively. Many existing models can be written in this form. For linear

models, both $\mathbf{T}(\mathbf{m})$ and $\mathbf{d}(\mathbf{m})$ are linear with respect to $\mathbf{m}$, such as the affine model [2], while for nonlinear models at least one of them is nonlinear, such as the TPS, B-Spline [12], Piecewise Bilinear Mapping (PBM) [13], as well as the new statistical model in Eq. (10). The TPS in Eq. (9), for instance, is obtained by letting $\mathbf{T}(\mathbf{m}) = [\mathbf{B}'(\mathbf{m}),\ \mathbf{I}_{(3)}]$, $\mathbf{d}(\mathbf{m}) = 0$, and $\boldsymbol{\xi}$ be a combined vector of $\boldsymbol{\theta}'$ and $\mathbf{p}_{\mathrm{o}}$, where $\mathbf{I}_{(3)}$ denotes a 3-dimensional identity matrix.

As illustrated in Fig. 1, given a pixel $\mathbf{m}$ in the template image, its 3D coordinates $\mathbf{p}$ in the real space are determined by the model parameter $\boldsymbol{\xi}$ to be estimated by using Eq. (11). According to the pinhole camera model, the perspective projection of $\mathbf{p}$ onto the left/right camera plane of the stereo-endoscope can be represented by the homogeneous coordinates

$$\mathbf{p}^{\mathrm{L/R}} = \mathbf{C}_{\mathrm{L/R}}\tilde{\mathbf{p}}, \tag{12}$$

where $\mathbf{C}_{\mathrm{L/R}}$ is the 3×4 left/right camera matrix, which can be calibrated with Zhang's method [27] before surgery. The corresponding pixel coordinates can be calculated by

$$\mathbf{m}^{\mathrm{L/R}} = \mathcal{H}\left(\mathbf{p}^{\mathrm{L/R}}\right), \tag{13}$$

where $\mathcal{H}$ maps homogeneous coordinates to pixel coordinates,

$$\mathcal{H}\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \begin{bmatrix} a/c \\ b/c \end{bmatrix} \tag{14}$$

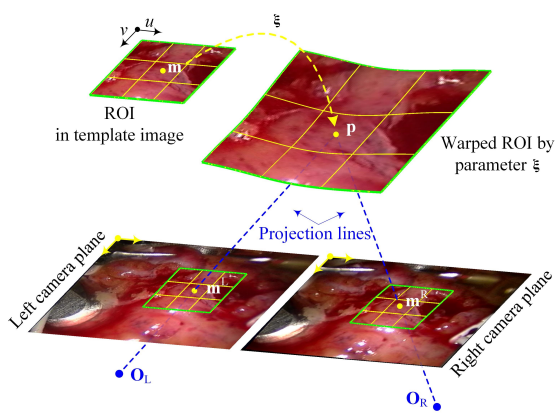


Fig. 1. Illustration of model-based ROI reconstruction

Let $\mathcal{I}_{\mathrm{T}}$ denote the intensity function of the template image and $\mathcal{I}_{\mathrm{L/R}}$ be that of the current left/right image. For the given ROI, there should be

$$\mathcal{I}_{\mathrm{T}}(\mathbf{m}) \equiv \mathcal{I}_{\mathrm{L/R}}\left(\mathbf{m}^{\mathrm{L/R}}|_{\boldsymbol{\xi}^*}\right),\ \forall\ \mathbf{m} \in \mathcal{R} \tag{15}$$

with an ideal parameter $\boldsymbol{\xi}^*$. However, in most cases, the ideal parameter does not exist, involving the solution of an overdetermined nonlinear system of equations. With the sum of squared pixel differences. The reconstruction problem can be formulated as

$$\arg\min_{\boldsymbol{\xi}}\left(\|\mathbf{I}_{\mathrm{L}} - \mathbf{I}_{\mathrm{T}}\|^2 + \|\mathbf{I}_{\mathrm{R}} - \mathbf{I}_{\mathrm{T}}\|^2\right), \tag{16}$$

Biomedical Optics EXPRESS

where $\mathbf{I}_{\mathrm{L/R}}$ and $\mathbf{I}_{\mathrm{T}}$ are N-dimensional intensity vectors stacked by $\{\mathcal{I}_{\mathrm{L/R}}(\mathbf{m}_n^{\mathrm{L/R}})\}_{n=1}^N$ and $\{\mathcal{I}_{\mathrm{T}}(\mathbf{m}_n)\}_{n=1}^N$, respectively. Regularization items can be added in Eq. (16) for robust solution, as in [28].

The Efficient Second-order Minimization (ESM) technique [29] is used to solve Eq. (16), which has a wider convergence basin and a faster convergence rate than traditional minimization algorithms. In the ESM, the model parameter is updated iteratively at each frame, $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} + \Delta\boldsymbol{\xi}$, until $\|\Delta\boldsymbol{\xi}\|$ is less than a preset threshold. The increment is calculated by

$$\Delta\boldsymbol{\xi} = -2\begin{bmatrix} \nabla_{\boldsymbol{\xi}}\mathbf{I}_{\mathrm{L}} + \nabla_{\boldsymbol{\xi}^*}\mathbf{I}_{\mathrm{L}} \\ \nabla_{\boldsymbol{\xi}}\mathbf{I}_{\mathrm{R}} + \nabla_{\boldsymbol{\xi}^*}\mathbf{I}_{\mathrm{R}} \end{bmatrix}^{+} \begin{bmatrix} \mathbf{I}_{\mathrm{L}} - \mathbf{I}_{\mathrm{T}} \\ \mathbf{I}_{\mathrm{R}} - \mathbf{I}_{\mathrm{T}} \end{bmatrix}, \tag{17}$$

where $\nabla_{\mathbf{x}}\mathbf{Y}$ denotes the Jacobian matrix of $\mathbf{Y}$ at $\mathbf{x}$, and "+" denotes the matrix pseudo-inverse. The computations of the Jacobians in Eq. (17) are given in Appendix B.

It should be noted that the proposed modeling method as well as the reconstruction algorithm, though derived from the TPS, apply equally to other deformable models. The shape representation in Eq. (4) and thus the statistical shape model in Eq. (8) can be derived from any $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ deformable model that can be written in the form of Eq. (11), including the B-spline and the PBM mentioned before. In fact, the $\mathbb{R}^2 \rightarrow \mathbb{R}$ B-spline based statistical shape model used for tissue segmentation in [23] is sub-optimal, because it is learnt simply from the parameter spaces without orthonormalizing the transformation matrix, which distorts the mapping from eigen-parameters to eigen-shapes, as proved in Proposition 1.

3. Results

Taking shape statistics into account, the low-rank model can be efficient in computation and robust against divergence of model parameter, when incorporated into the model-based reconstruction framework. Four stereo-endoscopic videos captured by the da Vinci robots are used to validate the proposed method. Each video consists of 800 stereo-pair frames with a frame rate of 25 Hz. As shown in Fig. 2, the videos I and II record the motion of a phantom heart model [30, 31], and the videos III and IV are in vivo beating hearts in real TECAB procedures [32]. Color images are converted into 8-bit grayscale for efficient processing.

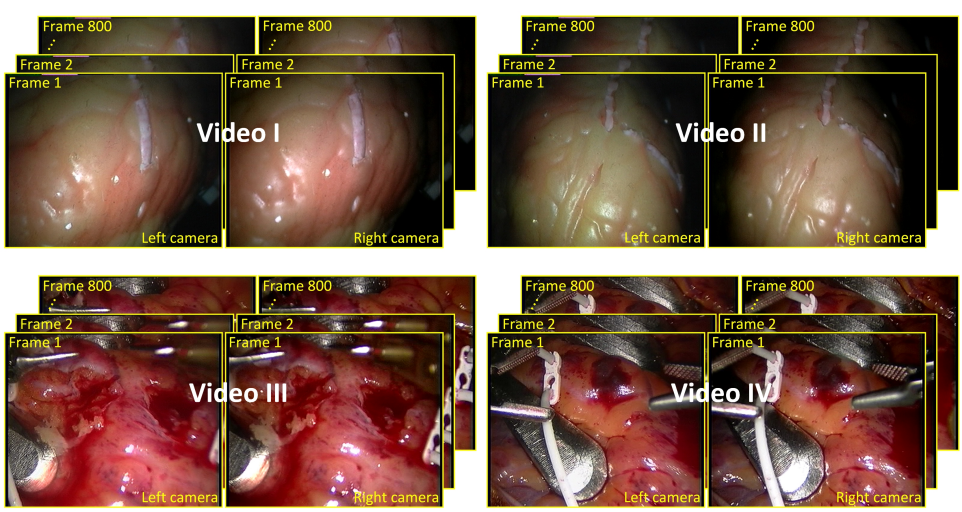


Fig. 2. Stereo-endoscopic videos (I-IV) for validation

Biomedical Optics EXPRESS

3.1. Sparsity and statistical stability of shape data

The rank of eigen-shape based model (J) is adjustable according to the efficiency and robustness requirements of reconstruction tasks in practice, as well as the sparsity of shape data itself, which can be indicated by the eigenvalues. A guideline for choosing J is given in our experiments.

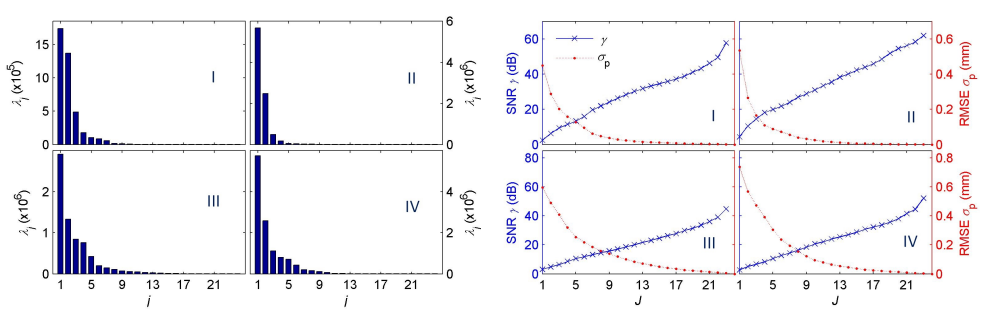
To obtain shape data, an orthonormalized 9-CP TPS is employed to reconstruct a 120×120 ROI on each video for 600 frames. At each frame, the ESM is used to estimate the shape and position parameters iteratively. Specular reflections caused by glossy tissue surfaces are removed by using a simple thresholding method [14]. The reconstruction flow is restarted manually when it is interrupted. We finally obtain four 24×600 shape-parameter matrices, which identify four hyper-high dimensional shape matrices (43200×600).

The eigenvalues and eigen-shapes are computed from shape-parameter matrices by using Proposition 1. As shown in Fig. 3(a), the eigenvalues on each video are sparse, indicating that there is obvious correlation between the shape vectors. The 24-dimensional shape subspace spanned by the 9-CP TPS is redundant for the ROI. Therefore, it is feasible to rebuild the shapes in a lower-dimensional space spanned by few eigen-shapes.

According to Eq. (7), the rebuilding error can be indicated more intuitively with the Signal-to-Noise Ratio (SNR) γ and the Root-Mean-Square Error (RMSE) σ_{p}, where

$$\gamma = 10\lg\left(\frac{\sum_{j=1}^{K'}\lambda_j}{\sum_{j=J+1}^{K'}\lambda_j}\right), \quad \sigma_{\mathrm{p}} = \sqrt{\frac{\sum_{j=J+1}^{K'}\lambda_j}{NL}} \tag{18}$$

Figure 3(b) gives the SNR and RMSE curves with respect to J. In our study, we choose the least J that ensures $\gamma > 20$ dB. Following this guideline, the first 8/6/12/10 eigen-shapes are reserved for statistical modeling on the video I/II/III/IV, corresponding to the RMSE of 0.048/0.072/0.070/0.094 mm/pixel, much lower than the measurement error (about 0.1 mm) of the stereo-endoscope image system itself and thus ignorable.



(a) Eigenvalues (b) SNR & RMSE

Fig. 3. PCA of shape data on videos I-IV

To show the statistical stability of shape data, we compared the PCA results computed from the first 600 frame and from all 800 frames. Figure 4 shows the corresponding mean shapes and first eigen-shapes on each video. The differences between the corresponding shapes are trivial (of the order of 0.01 mm/pixel), showing that the statistics of heartbeats are stable and thus reconstructing future frames with the statistical model trained from past frames is feasible.

3.2. Reconstruction experiments

In the reconstruction experiments, the first $L = 600$ frames in each video, serving as training frames, are pre-reconstructed with the 9-CP TPS, while the subsequent 200 frames, serving as

Biomedical Optics EXPRESS

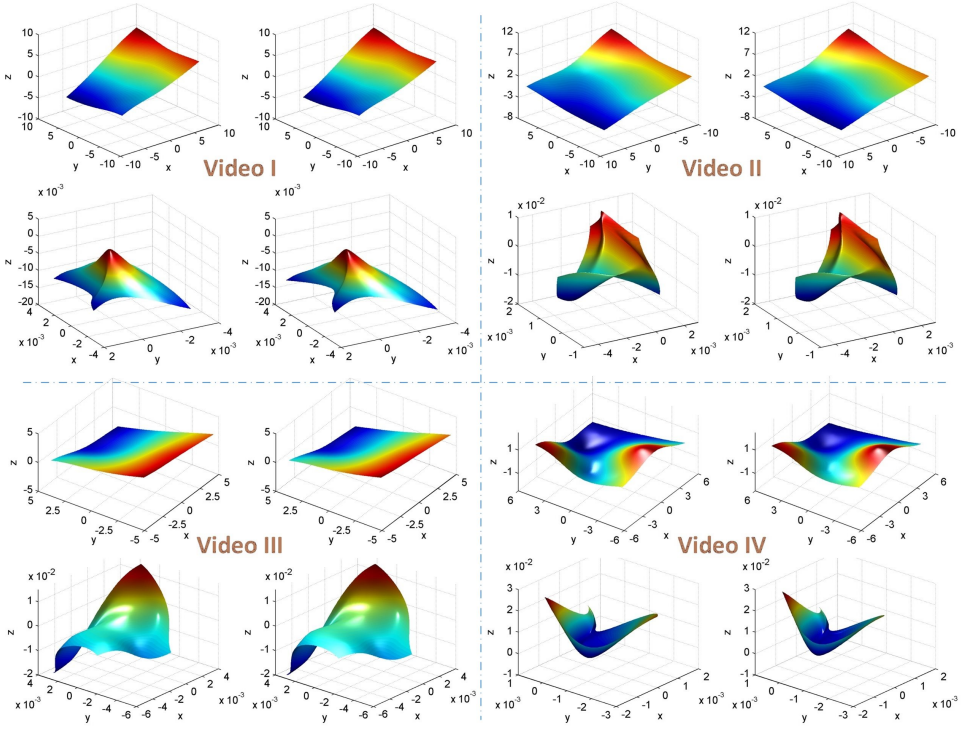


Fig. 4. Mean shapes and 1st eigen-shapes calculated from all 800 frames vs. from the first 600 frames. For each video, the first row shows the mean shapes, the second row the 1st eigen-shapes; the left shapes are from all 800 frames, the right shapes from the first 600 frame.

test frames, are reconstructed with the eigen-shape based Statistical Deformable Models (SDMs).

3.2.1. Qualitative analyses

Since the ground truth is unknown, reconstructed ROI is usually evaluated qualitatively by visually inspecting its concordance with heart beating and by analyzing its motion trajectories.

Figure 5 shows some reconstructed frames in each video. The SDMs can successfully handle soft-tissue deformations on beating hearts. The 3D shapes reconstructed at various frames are consistent with those perceived by human vision, and the center of the ROI is located accurately.

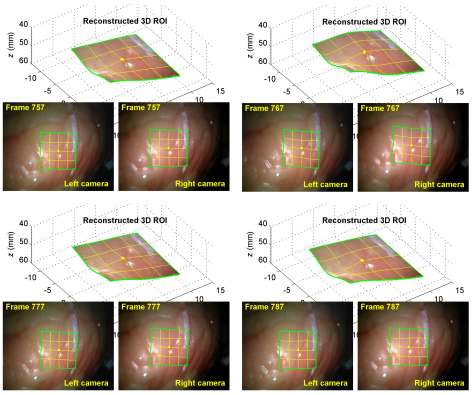
For comparison, the corresponding frames reconstructed by a 4-CP TPS on the video I are given in Fig. 6. Even though the 4-CP TPS has 12 parameters more than those (8 eigen-shape weights + 3 position parameters) of the SDM learnt for the video I, the 3D surfaces represented by the 4-CP TPS appear to be over-smooth with many 3D details missed out, comparing with those in Fig. 5(a), and therefore the advantages of the SDM are demonstrated.

The motion trajectories of the ROIs on the phantom videos (I and II) and their Power Spectral Densities (PSD) are shown in Fig. 7. The trajectories of $\mathbf{p}_o$ along x, y, z axes exhibit highly regular periodicity, and their power concentrates at 1.5 Hz and its second (3.0 Hz) and third (4.5 Hz) harmonics, which is consistent with the given rate of phantom heart beating.

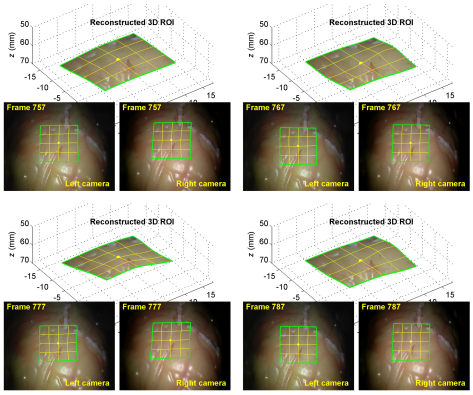
3.2.2. Quantitative evaluation

The SDMs are evaluated quantitatively and compared with well-established deformable models based on the ESM solution in terms of robustness, accuracy and speed. The reference models

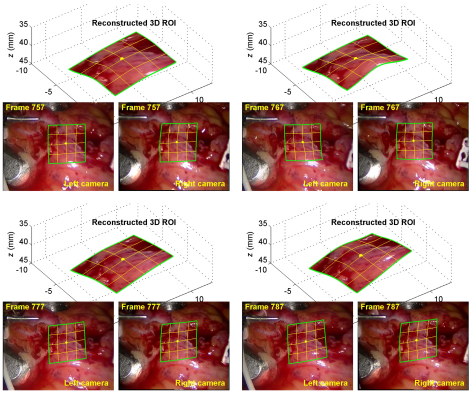
Biomedical Optics EXPRESS



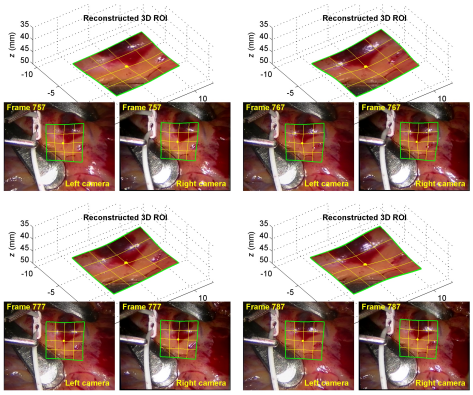
(a) Video I (see Visualization 1)



(b) Video II (see Visualization 2)



(c) Video III (see Visualization 3)



(d) Video IV (see Visualization 4)

Fig. 5. Reconstructed ROI with the statistical deformable models on various videos.

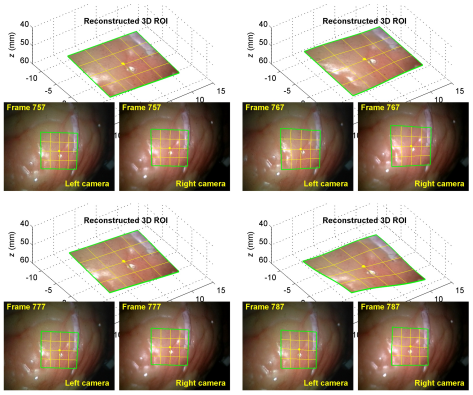
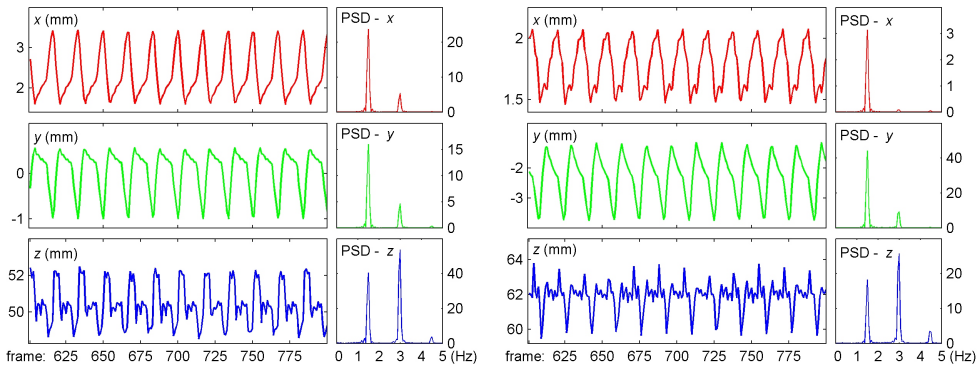


Fig. 6. Reconstructed ROI with the 4-CP TPS on video I.

Biomedical Optics EXPRESS



(a) Video I (b) Video II

Fig. 7. Motion trajectory analysis of ROI ($\mathbf{p}_o$) tracked by the statistical deformable model.

under comparison include the 9-CP TPS, 9-CP PBM and 9-CP B-Spline (all of them contain 27 parameters), and the Quasi-Spherical Triangle (QST) [16] of 10 parameters. In addition, the 4-CP TPS is tested on the phantom videos and the 5-CP TPS on the in vivo videos, since their ranks are close to those of the SDMs built for the corresponding videos.

Robustness: The ROI reconstruction with stereo-endoscope is very challenging. The processing chain for continuous image sequence is often interrupted by dynamic effects in the MIS. As recursive structure is usually adopted for efficiency, in which the final estimate at the current frame serves as an initial estimate for the next frame, reconstruction algorithms normally cannot recover by themselves when interrupted and must be restarted. The robustness of reconstruction algorithms can be measured by the number of frames processed continuously.

All 200 test frames on the phantom videos can be continuously processed by all investigated models, due to ideal imaging conditions and less dynamic disturbance. On the in vivo videos, the SDMs and the QST can take all 200 test frames, performing much better than the 9-CP TPS, 9-CP PBM and 9-CP B-Spline models which take averagely 58, 47 and 42 frames, respectively. It verifies that the high-rank models with too many parameters are susceptible to dynamic effects in the MIS, making the ESM prone to divergence, while the low-rank models with fewer parameters are easier to find optimal solutions in the noisy environments. The SDMs also perform better than the 5-CP TPS (taking averagely 152 frames) on the in vivo videos, though the ranks of these models are identical or similar. The result shows that for a specific ROI, the solution-space designed online by statistical modelling is more efficient than the fixed solution-space spanned by a general deformable model of identical or similar rank.

Accuracy: For evaluating accuracy, natural landmarks in the ROIs are manually tracked in both left and right image sequences. The 3D coordinates of the landmarks obtained by the model-based reconstruction are projected back onto the left and right images, and compared with the manually tracked results. To ensure the comparability of the results, interrupted algorithms are restarted manually. The joint pixel error is calculated at each frame

$$\epsilon = \sqrt{\left\|\mathbf{m}^{\mathrm{L}} - \acute{\mathbf{m}}^{\mathrm{L}}\right\|^2 + \left\|\mathbf{m}^{\mathrm{R}} - \acute{\mathbf{m}}^{\mathrm{R}}\right\|^2}, \tag{19}$$

where $\mathbf{m}^{\mathrm{L/R}}$ denotes the pixel coordinates projected onto the left/right image and $\acute{\mathbf{m}}^{\mathrm{L/R}}$ denotes those manually tracked. The Mean Error (ME) and Standard Deviation (SD) of ϵ over 200 test frames on each video are shown in Table 1.

On the phantom videos, good results are obtained by all models, as in the robustness test. The high-rank (9-CP) models have a slight advantage over the low-rank models (4-CP TPS, SDMs

Biomedical Optics EXPRESS

and QST). However, the SDM of 8/6 eigen-shapes on the video I/II performs better than the 4-CP TPS and QST, though have similar ranks. On the in vivo videos, dynamic effects in the MIS complicate the reconstruction task. The best results are obtained by the SDMs, showing that the proposed method is efficient for handling complicated deformations on in vivo hearts. Although the SDMs are trained from the shape data obtained by the 9-CP TPS, they are still able to achieve more accurate results than the latter in in vivo settings, thanks to the intrinsic advantage that the eigen-shapes extracted by the SDMs are insensitive to small noises in the training data.

Speed: The computational cost of reconstruction algorithms depends on the rank of the employed model. The most time-consuming operation in the ESM iteration is the pseudo-inverse calculation of the Jacobian matrix in Eq. (17). For a K-rank model on a N-size ROI, the computational complexity for one frame is $O(\mu NK^2)$, where μ is the number of the ESM iterations. Typically, on the in vivo videos, the average μ of the 12 eigen-shape SDM is 6.8, close to that of the QST and about half those of the 9-CP models, confirming that models with fewer parameters are easier to converge. For the same ROI, the computational cost of a typical 12 eigen-shape SDM is nearly 15% that of a 9-CP model.

To evaluate practical speed, the average CPU time is computed based on our experimental platform (Matlab R2012a, Intel Xeon E3-1231 CPU and 8GB RAM). The speed of the 12 eigen-shape SDM is 2.9 s/frame, much faster than 9.2 s/frame of the 9-CP TPS. By means of statistical modeling, real-time reconstruction is feasible if Graphics Processor Units (GPU) and more efficient programming languages (e.g. C/C++) are adopted.

Finally, the SDMs built with different SNR guidelines ($\gamma > 30$ dB and $\gamma > 10$ dB) are tested. For $\gamma > 30$ dB, the first 12/10/19/16 eigen-shapes are reserved on the video I/II/III/IV. There is not significant improvement in the reconstruction accuracy though more eigen-shapes are employed, while the robustness and speed are decreased obviously. Their performances are between those of the SDM ($\gamma > 20$ dB) and the 9-CP TPS. For $\gamma > 10$ dB, only the first 4/2/5/5 eigen-shapes are reserved on the video I/II/III/IV, which results in faster and more robust reconstructions. However, the reconstructed surfaces are inaccurate due to too few shape degrees of freedom. The tracking errors are even larger than those of the QST in Table 1.

Overall, high-rank models can render tissue surfaces more accurately in theory; however, in practice they often *confuse* parameter estimation algorithms due to too many degrees of freedom. For general deformable models based on spline or mesh, increasing CPs will result in a quadratic growth of computational burden and cause model parameters difficult to convergence, while reducing CPs may lead to over-smoothing and loss of 3D details. It is a dilemma for these general models to choose an appropriate rank, since they were designed for handling general surfaces without taking into account the statistics and specificities of the ROI. The proposed method

Table 1. Tracking errors (ME±SD) compared with manually tracked results.

Methods	Video I	Video II	Video III	Video IV
SDM	1.26±0.46	1.27±0.62	2.52±0.98	2.40±0.90
QST	1.42±0.51	1.51±0.70	3.09±1.10	2.80±1.09
4/5-CP TPS	1.40±0.52	1.44±0.68	3.03±1.11	2.75±1.01
9-CP TPS	1.21±0.51	1.23±0.64	2.83±1.04	2.59±0.87
9-CP PBM	1.23±0.53	1.29±0.66	2.92±1.09	2.64±0.92
9-CP B-Spline	1.21±0.53	1.27±0.63	2.90±1.10	2.60±0.90

provides a good solution to this dilemma, which can balance the robustness, accuracy and speed well. As shown in the experiments, the SDMs are efficient in computation and robust against dynamic disturbances without sacrificing reconstruction accuracy, and they are fully superior to the general models of the same or similar rank.

4. Conclusions

We propose a statistical modeling method to reconstruct ROIs on beating heart from stereo-endoscopic video. Statistical shape models are trained online by learning eigen-shapes from past shapes of ROIs. To obtain past shapes, the high-rank versions of classical deformable models (such as but not limited to the TPS) after decoupling and orthonormalizing, can be employed to pre-reconstruct a ROI for a number of frames. The linear transferability of PCA is proved, which can significantly reduce the computational burden of eigen-shape learning. Since the statistical model is trained specially for fitting the 3D structure of the ROI, it is very efficient when incorporated into a model-based reconstruction framework, which formulates the dynamic reconstruction as an optimization problem of model parameters for each stereo image frame. Based on a general model equation, the efficient second-order minimization algorithm is derived to iteratively solve model parameters. We validate the modeling method on the stereo-endoscopic videos collected by the da Vinci robots. The analyses of shape data from various videos show the sparsity of eigenvalues and statistical stability of eigen-shapes, and thus the feasibility of low-rank statistical modeling. The ROI reconstruction experiments demonstrate the efficiency of the statistical models and their advantages over the traditional models in balancing robustness, accuracy and speed. The proposed method provides an efficient and robust way to track and reconstruct dynamic soft-tissues with stereo-endoscope, which is important for further improving the functionality and applicability of surgical robots. Besides, the proposed method is also useful for offline tasks, e.g. surgical simulation or visual medical record, which require efficient representation and rendering of 3D tissues.

Appendix A. Proof of proposition 2

Since $\mathbb{U} \cong \mathbb{V}$, $\mathbb{U}$ and $\mathbb{V}$ have the same dimension, denoted to be M without loss of generality. Since $\mathbf{B}_{\mathbb{U}/\mathbb{V}}$ is a column-orthonormal, for any vector set $\mathbf{D} \subset \mathbb{R}^M$, using Proposition 1 gives

$$\mathcal{P}\left(\mathbf{B}_{\mathbb{U}/\mathbb{V}}\mathbf{D}\right) = \mathbf{B}_{\mathbb{U}/\mathbb{V}}\mathcal{P}\left(\mathbf{D}\right) \tag{20}$$

Suppose $\mathbf{X} \subset \mathbb{U}$. Since $\mathbb{U} \cong \mathbb{R}^M$, there is a unique coordinates set $\mathbf{D} \subset \mathbb{R}^M$ that identifies $\mathbf{X}$ with the orthonormal basis $\mathbf{B}$, i.e. $\mathbf{X} = \mathbf{B}_{\mathbb{U}}\mathbf{D}$ and $\mathbf{D} = \mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathbf{X}$. From Eq. (20), one can get

$$\mathcal{P}\left(\mathbf{B}_{\mathbb{V}}\mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathbf{X}\right) = \mathcal{P}\left(\mathbf{B}_{\mathbb{V}}\mathbf{D}\right) = \mathbf{B}_{\mathbb{V}}\mathcal{P}\left(\mathbf{D}\right) \tag{21}$$

$$\mathcal{P}\left(\mathbf{D}\right) = \mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathbf{B}_{\mathbb{U}}\mathcal{P}\left(\mathbf{D}\right) = \mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathcal{P}\left(\mathbf{B}_{\mathbb{U}}\mathbf{D}\right) = \mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathcal{P}\left(X\right), \tag{22}$$

and thus $\mathcal{P} \circ \mathcal{F}(\mathbf{X}) = \mathcal{P}(\mathbf{B}_{\mathbb{V}}\mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathbf{X}) = \mathbf{B}_{\mathbb{V}}\mathbf{B}_{\mathbb{U}}^{\mathrm{T}}\mathcal{P}(\mathbf{X}) = \mathcal{F} \circ \mathcal{P}(\mathbf{X})$.

Similarly, using Proposition 1 gives $\mathcal{S}(\mathbf{B}_{\mathbb{U}/\mathbb{V}}\mathbf{D}) = \mathcal{S}(\mathbf{D})$, and thus $\mathcal{S} \circ \mathcal{F}(\mathbf{X}) = \mathcal{S}(\mathbf{B}_{\mathbb{V}}\mathbf{D}) = \mathcal{S}(\mathbf{B}_{\mathbb{U}}\mathbf{D}) = \mathcal{S}(\mathbf{X})$.

Appendix B. Computations of Jacobians

B.1. Computation of $\nabla_{\xi}\mathbf{I}_{\mathrm{L/R}}$

The n-th row of $\nabla_{\xi}\mathbf{I}_{\mathrm{Y}}$ (with Y denoting L or R) can be decomposed into four sub-Jacobians

$$\nabla_{\xi}\mathcal{I}_{\mathrm{Y}}\left(\mathbf{m}_n^{\mathrm{Y}}\right) = \nabla_{\mathbf{m}_n^{\mathrm{Y}}}\mathcal{I}_{\mathrm{Y}}\ \nabla_{\mathbf{p}_n^{\mathrm{Y}}}\mathcal{H}\ \nabla_{\mathbf{p}}\mathbf{p}^{\mathrm{Y}}\ \nabla_{\xi}\mathbf{p}|_{\mathbf{m}_n}, \tag{23}$$

Biomedical Optics EXPRESS

where:

a) $\nabla_{\mathbf{m}_n^{\mathrm{Y}}} \mathcal{I}_{\mathrm{Y}}$ is the gradient (1×2) of the image $\mathcal{I}_{\mathrm{Y}}$ at $\mathbf{m}_n^{\mathrm{Y}}$.

b) $\nabla_{\mathbf{p}_n^{\mathrm{Y}}} \mathcal{H}$ is the derivative of function $\mathcal{H}$ at $\mathbf{p}_n^{\mathrm{Y}}$. From Eq. (14), one derives

$$\nabla_{\mathbf{x}} \mathcal{H} = \begin{bmatrix} 1/c & 0 & -a/c^2 \\ 0 & 1/c & -b/c^2 \end{bmatrix} \tag{24}$$

c) $\nabla_{\mathbf{p}} \mathbf{p}^{\mathrm{Y}} \equiv [\mathbf{C}_{\mathrm{Y}}]_3$ and $\nabla_{\boldsymbol{\xi}} \mathbf{p}|_{\mathbf{m}_n} \equiv \mathbf{T}(\mathbf{m}_n)$ are constant for a given ROI.

B.2. Computation of $\nabla_{\boldsymbol{\xi}^*} \mathbf{I}_{\mathrm{L/R}}$

The Jacobian cannot be computed directly because the ideal parameter $\boldsymbol{\xi}^*$ is unknown. An effective approximation is derived by using the template image and the current parameter $\boldsymbol{\xi}$.

Given a parameter $\boldsymbol{\xi}$, the mapping $\mathcal{M}_{\boldsymbol{\xi}} : \mathbf{m} \mapsto \mathbf{m}^{\mathrm{Y}}$ is normally invertible, i.e., there is $\mathbf{m} = \mathcal{M}_{\boldsymbol{\xi}}^{-1}\left(\mathbf{m}^{\mathrm{Y}}\right)$. By using Eq. (15), the n-th row of $\nabla_{\boldsymbol{\xi}^*} \mathbf{I}_{\mathrm{Y}}$ can be written as

$$\nabla_{\boldsymbol{\xi}^*} \mathcal{I}_{\mathrm{Y}}\left(\mathbf{m}_n^{\mathrm{Y}}\right) = \nabla_{\boldsymbol{\xi}^*} \mathcal{I}_{\mathrm{T}}\left(\mathbf{m}_n\right), \tag{25}$$

and similarly decomposed into

$$\nabla_{\mathbf{m}_n} \mathcal{I}_{\mathrm{T}} \; \nabla_{\mathbf{m}_n^{\mathrm{Y}}} \mathcal{M}_{\boldsymbol{\xi}^*}^{-1} \; \nabla_{\mathbf{p}_n^{\mathrm{Y}}|_{\boldsymbol{\xi}^*}} \mathcal{H} \; \nabla_{\mathbf{p}} \mathbf{p}^{\mathrm{Y}} \; \nabla_{\boldsymbol{\xi}} \mathbf{p}|_{\mathbf{m}_n} \tag{26}$$

where:

a) $\nabla_{\mathbf{m}_n} \mathcal{I}_{\mathrm{T}}$ is the gradient of $\mathcal{I}_{\mathrm{T}}$ at $\mathbf{m}_n$.

b) $\nabla_{\mathbf{p}} \mathbf{p}^{\mathrm{Y}}$ and $\nabla_{\boldsymbol{\xi}} \mathbf{p}|_{\mathbf{m}_n}$ are constant, independent of $\boldsymbol{\xi}^*$, the same as in Eq. (23).

c) $\nabla_{\mathbf{m}_n^{\mathrm{Y}}} \mathcal{M}_{\boldsymbol{\xi}^*}^{-1}$ is the inverse of the Jacobian

$$\nabla_{\mathbf{m}_n} \mathcal{M}_{\boldsymbol{\xi}^*} = \nabla_{\mathbf{p}_n^{\mathrm{Y}}|_{\boldsymbol{\xi}^*}} \mathcal{H} \; \nabla_{\mathbf{p}} \mathbf{p}^{\mathrm{Y}} \; \nabla_{\mathbf{m}_n} \mathbf{p}|_{\boldsymbol{\xi}^*}, \tag{27}$$

where $\mathbf{p}_n^{\mathrm{Y}}|_{\boldsymbol{\xi}^*}$, the ideal homogeneous coordinates determined by $\boldsymbol{\xi}^*$, can be approximated by $\mathbf{p}_n^{\mathrm{Y}}$, and $\nabla_{\mathbf{m}_n} \mathbf{p}|_{\boldsymbol{\xi}^*}$, the derivative (3×2) of the deformable model at $\mathbf{m}_n$ with $\boldsymbol{\xi}^*$, for $\mathbf{m} = [u \; v]^{\mathrm{T}}$,

$$\nabla_{\mathbf{m}_n} \mathbf{p}|_{\boldsymbol{\xi}^*} = \left[\left.\frac{\partial \mathbf{T}(\mathbf{m})}{\partial u}\right|_{u_n} \boldsymbol{\xi}^* \quad \left.\frac{\partial \mathbf{T}(\mathbf{m})}{\partial v}\right|_{u_n} \boldsymbol{\xi}^* \right] + \nabla_{\mathbf{m}_n} \mathbf{d}(\mathbf{m}), \tag{28}$$

can be approximated by letting $\boldsymbol{\xi}^* = \boldsymbol{\xi}$.

In Eq. (28), $\partial \mathbf{T}(\mathbf{m})/\partial \eta$ (with η denoting u or v for convenience) and $\nabla_{\mathbf{m}} \mathbf{d}(\mathbf{m})$ are constant and can be pre-computed for a given ROI. Take the statistical model in Eq. (10) for example:

$$\frac{\partial \mathbf{T}(\mathbf{m})}{\partial \eta} = \left[\frac{\partial \mathbf{B}(\mathbf{m})}{\partial \eta} \mathbf{R}^{-1} \left[\mathbf{U}_{\Theta'}\right]_J \quad \mathbf{0} \right], \tag{29}$$

$$\nabla_{\mathbf{m}} \mathbf{d}(\mathbf{m}) = \left[\frac{\partial \mathbf{B}(\mathbf{m})}{\partial u} \mathbf{R}^{-1} \bar{\boldsymbol{\theta}}' \quad \frac{\partial \mathbf{B}(\mathbf{m})}{\partial v} \mathbf{R}^{-1} \bar{\boldsymbol{\theta}}' \right] \tag{30}$$

where, according to Eq. (3), $\partial \mathbf{B}(\mathbf{m})/\partial \eta$ is formed diagonally with

$$\frac{\partial \mathbf{b}(\mathbf{m})}{\partial \eta} = \left[\frac{\boldsymbol{\varphi}'(\mathbf{m})}{\partial \eta} \quad \underbrace{1/0 \quad 0/1}_{\text{for } \eta = u/v} \right] \tag{31}$$

Biomedical Optics EXPRESS

Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two complementary submatrices of the left matrix in Eq. (2), where $\mathbf{C}_1$ is 3×3 invertible, corresponding to the three eliminated elements of $\boldsymbol{\alpha}$, and $\mathbf{C}_2$ corresponds to $\boldsymbol{\alpha}'$. Then

$$\boldsymbol{\varphi}'(\mathbf{m}) = \boldsymbol{\varphi}(\mathbf{m}) \begin{bmatrix} -\mathbf{C}_1^{-1}\mathbf{C}_2 \\ \mathbf{I}_{(K-3)} \end{bmatrix} \tag{32}$$

and the derivatives of the k-th radial basis kernel in $\boldsymbol{\varphi}(\mathbf{m})$ can be pre-computed by using

$$\begin{bmatrix} \dfrac{\partial \varphi(\|\mathbf{m}-\mathbf{c}_k\|)}{\partial u} & \dfrac{\partial \varphi(\|\mathbf{m}-\mathbf{c}_k\|)}{\partial v} \end{bmatrix} = (2\ln\|\mathbf{m}-\mathbf{c}_k\| + 1)(\mathbf{m}-\mathbf{c}_k)^{\mathrm{T}} \tag{33}$$

Funding

National Natural Science Foundation of China (61305022); Sichuan Science and Technology Program (2018SZDZX0013, 2017HH0054); State Key Laboratory Virtual Reality Technology and Systems (BUAA-VR-16KF-11).

Acknowledgments

We would like to acknowledge the da Vinci data provided by the Hamlyn Centre, Imperial College London.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. Y. Nakamura, K. Kishi, and H. Kawakami, "Heartbeat synchronization for robotic cardiac surgery," in Proceedings of IEEE Conference on Robotics and Automation (IEEE, 2001), pp. 2014–2019.
2. T. Ortmaier, M. Gröger, D. H. Boehm, V. Falk, and G. Hirzinger, "Motion estimation in beating heart surgery," IEEE T. Biomed. Eng. **52**(10), 1729–1740 (2005).
3. M. Hayashibe, N. Suzuki, and Y. Nakamura, "Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management," Med. Image Anal. **10**(4), 509–519 (2006).
4. N. T. Clancy, D. Stoyanov, L. Maier-Hein, A. Groch, G. Z. Yang, and D. S. Elson, "Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging," Biomed. Opt. Express **2**(11), 3119–3128 (2011).
5. O. Bebek and M. C. Cavusoglu, "Intelligent control algorithms for robotic-assisted beating heart surgery," IEEE T. Robot. **23**(3), 468–480 (2007).
6. P. Mountney, D. Stoyanov, and G. Z. Yang, "Three-dimensional tissue deformation recovery and tracking," IEEE Signal Proc. Mag. **27**(4), 14–24 (2010).
7. S. Bernhardt, S. A. Nicolau, V. Agnus, L. Soler, C. Doignon, and J. Marescaux, "Automatic localization of endoscope in intraoperative CT image: A simple approach to augmented reality guidance in laparoscopic surgery," Med. Image Anal. **30**, 130–143 (2016).
8. N. T. Clancy, D. Stoyanov, D. R. C. James, A. Di Marco, V. Sauvage, J. Clark, G. Z. Yang, and D. S. Elson, "Multispectral image alignment using a three channel endoscope in vivo during minimally invasive surgery," Biomed. Opt. Express **3**(10), 2567–2578 (2012).
9. K. L. Lurie, R. Angst, E. J. Seibel, J. C. Liao, and A. K. E. Bowden, "Registration of free-hand OCT daughter endoscopy to 3D organ reconstruction," Biomed. Opt. Express **7**(12), 4995–5009 (2016).
10. J. Zhang, J. Chang, X. Yang, and J. J. Zhang, "Virtual reality surgery simulation: A survey on patient specific solution," in Next Generation Computer Animation Techniques, J. Chang, J. J. Zhang, N. M. Thalmann, S. M. Hu, R. Tong, W. Wang, ed. (Springer, 2017), pp. 220–233.
11. K. L. Lurie, R. Angst, D. V. Zlatev, J. C. Liao, and A. K. E. Bowden, "3D reconstruction of cystoscopy videos for comprehensive bladder records," Biomed. Opt. Express **8**(4), 2106–2123 (2017).
12. W. Lau, N. A. Ramey, J. J. Corso, N. V. Thakor, and G. D. Hager, "Stereo-based endoscopic tracking of cardiac surface deformation," in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004, C. Barillot, D. R. Haynor, P. Hellier, ed. (Springer, 2004), pp. 494–501.
13. D. Stoyanov, A. Darzi, and G. Z. Yang, "A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery," Comput. Aided. Surg. **10**(4), 199–208 (2005).
14. 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. **29**(2-3), 218–230 (2010).

Biomedical Optics EXPRESS

15. 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. Ass. Rad. **6**(3), 387–399 (2011).
16. W. K. Wong, B. Yang, C. Liu, and P. Poignet, "A quasi-spherical triangle-based approach for efficient 3-D soft-tissue motion tracking," IEEE-ASME T. Mech. **18**(5), 1472–1484 (2013).
17. 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 Recogn. **47**(9), 2962–2973 (2014).
18. B. Yang, C. Liu, K. L. Huang, and W. F. Zheng, "A triangular radial cubic spline deformation model for efficient 3D beating heart tracking," Signal Image Video P. **11**(7), 1329–1336 (2017).
19. N. T. Clancy, S. Arya, J. Qi, D. Stoyanov, G. B. Hanna, and D. S. Elson, "Polarised stereo endoscope and narrowband detection for minimal access surgery," Biomed. Opt. Express **5**(12), 4108–4117 (2014).
20. N. Baka, B. L. Kaptein, M. de. Bruijne, T. van. Walsum, J. E. Giphart, W. J. Niessen, and B. P. F. Lelieveldt, "2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models," Med. Image Anal. **15**(6), 840–850 (2011).
21. E. M. A. Anas, A. Rasoulian, A. Seitel, K. Darras, D. Wilson, P. S. John, D. Pichora, P. Mousavi, R. Rohling, and P. Abolmaesumi, "Automatic segmentation of wrist bones in CT using a statistical wrist shape + pose model," IEEE T. Med. Imag. **35**(8), 1789–1801 (2016).
22. X. Albà, M. Pereañez, C. Hoogendoorn, A. J. Swift, J. M. Wild, A. F. Frangi, and K. Lekadir, "An algorithm for the segmentation of highly abnormal hearts using a generic statistical shape model," IEEE T. Med. Imag. **35**(8), 845–859 (2016).
23. J. Pedrosa, S. Queirós, O. Bernard, J. Engvall, T. Edvardsen , E. Nagel, and J. D'hooge, "Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based B-spline explicit active surfaces," IEEE T. Med. Imag. **36**(11), 2287–2296 (2017).
24. V. Kajić, M. Esmaeelpour, B. Považay, D. Marshall, P. L. Rosin, and W. Drexler, "Automated choroidal segmentation of 1060 nm OCT in healthy and pathologic eyes using a statistical model," Biomed. Opt. Express **3**(1), 86–103 (2012).
25. M. Pilch, Y. Wenner, E. Strohmayr, M. Preising, C. Friedburg, E. M. zu Bexten, B. Lorenz, and K. Stieger, "Automated segmentation of retinal blood vessels in spectral domain optical coherence tomography scans," Biomed. Opt. Express **3**(7), 1478–1491 (2012).
26. B. Yang, C. Liu, W. F. Zheng, and S. Liu, "Motion prediction via online instantaneous frequency estimation for vision-based beating heart tracking," Information Fusion **35**, 58–67 (2017).
27. Z. Zhang, "A flexible new technique for camera calibration," IEEE T. Pattern Anal. **22**(11), 1330–1334 (2000).
28. A. I. Aviles-Rivero, S. M. Alsaleh, and A. Casals A, "Sliding to predict: vision-based beating heart motion estimation by modeling temporal interactions," Int. J. Comput. Ass. Rad. **13**(3), 353–361 (2018).
29. S. Benhimane and E. Malis, "Homography-based 2D visual tracking and servoing," Int. J. Robot. Res. **26**(7), 661–676 (2007).
30. D. Stoyanov, M. V. Scarzanella, P. Pratt, and G. Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, T. Jiang, N. Navab, J. P. W. Pluim, M. A. Viergever, ed. (Springs, 2010), pp. 275–282.
31. P. Pratt, D. Stoyanov, M. V. Scarzanella, and G. Z. Yang, "Dynamic guidance for robotic surgery using image-constrained biomechanical models," in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, T. Jiang, N. Navab, J. P. W. Pluim, M. A. Viergever, ed. (Springs, 2010), pp. 77–85.
32. 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," in Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, J. S. Duncan, G. Gerig, ed. (Springs, 2005), pp. 139–146.