

Adaptive Model-Based Visual Stabilization of Image Sequences Using Feedback

Evgeniya Bogatyrenko and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,

Karlsruhe Institute of Technology (KIT), Germany.
Email: evgeniya.bogatyrenko@kit.edu, uwe.hanebeck@ieee.org

Abstract—Visual stabilization proposed in this paper compensates changes of the scene caused by motion and deformation of an observed object. This is of high importance in computer-assisted beating heart surgery, where the views of the beating heart should be stabilized. The proposed model-based method defines visual stabilization as a transformation of the current image sequence to a stabilized image sequence. This transformation incorporates physical model of the observed object and model of the measurement process. In contrast to standard approaches, the quality of the visual stabilization is continuously evaluated and improved in two aspects. On the one hand, discretization errors are reduced. On the other hand, the parameters of the underlying models are adjusted. The performance of the proposed method is evaluated in an experiment with a pressure-regulated artificial heart. Compared with standard methods, the model-based method provides higher accuracy, which is additionally improved by a feedback mechanism.

Keywords: heart surface motion compensation, estimation, model adaptation, video processing.

I. INTRODUCTION

Visual stabilization is of importance for many industrial applications, such as automated handling of elastic objects like an insertion a flexible beam into a hole [1] or servoing in presence of non-rigid motion [2], [3]. In medical applications, visual stabilization is essential for computer-assisted surgical operations on soft tissues [4]. In these applications, a moving object is observed by a camera system that provides image sequences displaying a changing scene. The aim of visual stabilization is to compensate the object motion and deformation for representing this scene as stationary. For that purpose, the image sequence is transformed to a stabilized image sequence, as shown in Fig. 1. As a result, the stabilized images depict a moving deformable object as motionless, whereby its changes, such as different coloring, remain visible.

The proposed approach for visual stabilization is suggested for an application in a computer-assisted beating heart surgery system, first introduced in [4]. On the one hand, this system synchronizes surgical instruments with a beating heart [5]. On the other hand, it gives a surgeon an impression of operating on a stabilized heart by representing the heart surface as motionless. Application of this system for beating heart operations will lead to higher accuracy and repeatability of surgical interventions. While approaches for synchronization of surgical instruments are widely investigated [6]–[12], only few methods for the visual stabilization can be found in literature. In this application, the main challenge of the visual

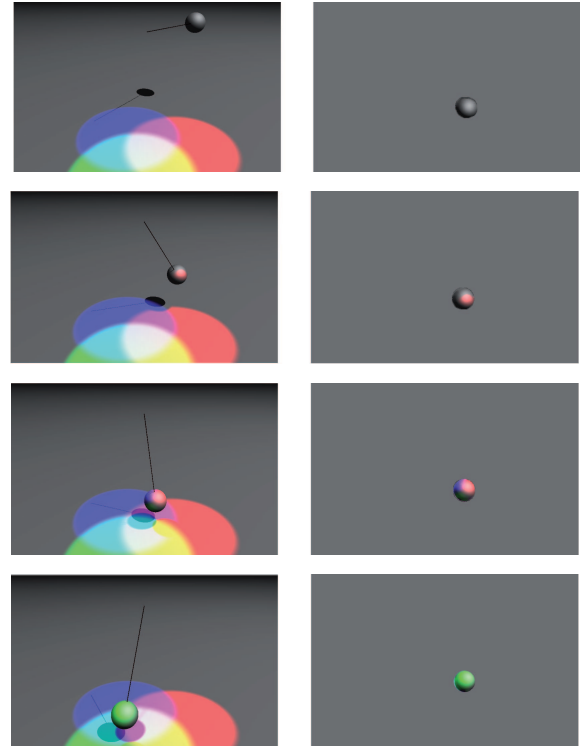


Figure 1. Visual stabilization of a moving pendulum. The image sequence provided by a camera is transformed to the stabilized image sequence.

stabilization is accurate reconstruction of continuous heart surface motion. This is difficult, because the space- and time-discrete measurement information provided by a camera system is uncertain and the heart dynamics are unknown.

Related work in the area of visual stabilization for industrial and medical application can be classified into four groups. First, the visual stabilization can be provided by sub-sampling, i.e., by selecting only those images from the image sequence that depict an object at the desired position. The method proposed in [13] is inspired by this idea. Here, electrocardiogram-triggered strobed light is used for making a heart appear still to a surgeon. The obtained experimental results testify that this method leads to increasing demands on surgeon's concentration and to fatigue. The main reason for that is neglected information between the sampled views. The second group includes approaches that transform the

current image sequence to the stabilized image sequence by a global transformation, e.g., by moving a camera according to the object motion [2], [4]. In [14], the camera motion is simulated by changing extrinsic camera parameters. Generally, these methods are unable to compensate the local motion of a deformable object. The third group combines methods using geometric image transformation. Here, the images acquired at every time step are mapped to the reference images, e.g., by linear interpolation [15] or by image transformation methods [16] such as warping [17] or morphing [18]. Some type of hybrid formulation is proposed in [19], where region-based deformable appearance models incorporate the combined parametrization provided by modal analysis and principal components analysis. Since this group of methods incorporates only geometric constraints in the transformation, motion artefacts can occur in the stabilized image sequence because the physical characteristics of the observed object are not considered. So, the image transformation by linear interpolation may become rough [17], since no smoothness constraints regarding the heart surface motion are considered. A method for visual stabilization related to the forth group was recently proposed in [20]. The key idea of this method is incorporating a physical model of an observed object and a model of the measurement process in the image transformation. According to the presented experimental results, the model-based visual stabilization provides higher accuracy than the geometric image transformation [21]. This is due to consideration of physical characteristics of the heart, so as model and measurement uncertainties.

The common disadvantage of all presented methods is that the quality of the visual stabilization after the transformation is not continuously monitored and evaluated. It should be noted that the measurement disturbances, such as stochastic uncertainties caused by camera noise, environmental disturbances like smoke due to tissue cutting, and inaccuracies of feature extraction will affect the quality of the visual stabilization. However, its deterioration in case of highly inaccurate measurements or transformation errors will not be detected and corrected.

The main contribution of the proposed adaptive model-based approach for visual stabilization is incorporating feedback from the stabilized image for improving the quality of the image transformation. Here, the model-based method proposed in [20] is extended in such way that the underlying models are adapted for reducing the image transformation errors.

After formulating the problem of image transformation in Section II, an overview of the proposed method with an illustration of its key idea is given in Section III. The main components of the method, such as image transformation function, involved models, estimation, and adaptation, are described in Section IV. In Section V, the model-based image transformation is evaluated in an experiment with a pressure-regulated artificial heart. It is compared to one of the standard methods based on geometric image transformation. Finally, Section VI outlines the main contributions and achieved results.

II. PROBLEM FORMULATION

This section deals with formulating the problem of visual stabilization. The setup of this problem is described with regard to the application of visual stabilization in a computer-assisted beating heart surgery. In this application, the motion of the intervention area of the heart surface should be visually stabilized. For that purpose, the beating heart is observed by a stereo endoscope or multicamera system.

For clarity of the problem formulation, we define the current camera image by a set of two-dimensional points, i.e., pixels

$$\mathcal{P}_k := \left\{ \underline{p}_k^i \right\}_i^P.$$

The size of this set P is determined by the resolution of the image. The position of each pixel

$$\underline{p}_k := [p_k^x, p_k^y]^T$$

in this image is identified by integer indices p_k^x and p_k^y in x and y directions of the image. Furthermore, the reference image is defined as one of the previous camera images. Therefore, when, for example, the reference image is initialized at time step t_{k-n} , it is described by a set of pixels $\mathcal{P}_{k-n}$.

The transformation of the current image to the reference image is provided by assigning the intensity $I(\underline{p}_k)$ of each pixel in the current camera image to each pixel $\underline{p}_{k-n}$ in the reference image. For that purpose, an image transformation function establishes the correspondences between the pixels of both images

$$T_k(\mathcal{C}_k(\mathcal{F}_{k-n}, \mathcal{F}_k), \mathcal{P}_k) : \mathcal{P}_k \rightarrow \mathcal{P}_{k-n} \quad (1)$$

based on the correspondences between the image features in these images

$$\mathcal{F}_{k-n} \rightarrow \mathcal{F}_k,$$

where

$$\mathcal{F}_k := \left\{ \underline{f}_k^i \right\}_{i=1}^F, \quad \underline{f}_k \subset \mathcal{P}_k.$$

The positions of the image features $\mathcal{F}_k$ in the current image are usually measured at the current time step t_k . Their positions $\mathcal{F}_{k-n}$ in the reference image acquired at time step t_{k-n} are exactly known. It should be noted that the transformation function depends on the unknown set of parameters $\mathcal{C}_k \in \mathbb{R}^L$. These parameters are used for approximating the positions of the pixels in the current image $\mathcal{P}_k$. In geometric image transformation approaches, i.e., by warping [16], [21], they incorporate the geometric constraints on the movement of the deformable object. In contrast, in the model-based image transformation method proposed in [20], the values of these parameters are influenced by geometric and physical constraints on the movement of the observed object. Due to physical constraints, unrealistic estimation of the object deformation is excluded.

Obviously, the quality of the approximation and therefore, of the visual stabilization strongly depends on the quality of the underlying models. However, in existing image transformation approaches, the models are assumed to be exactly known. Therefore, in this paper, an adaptive image transformation

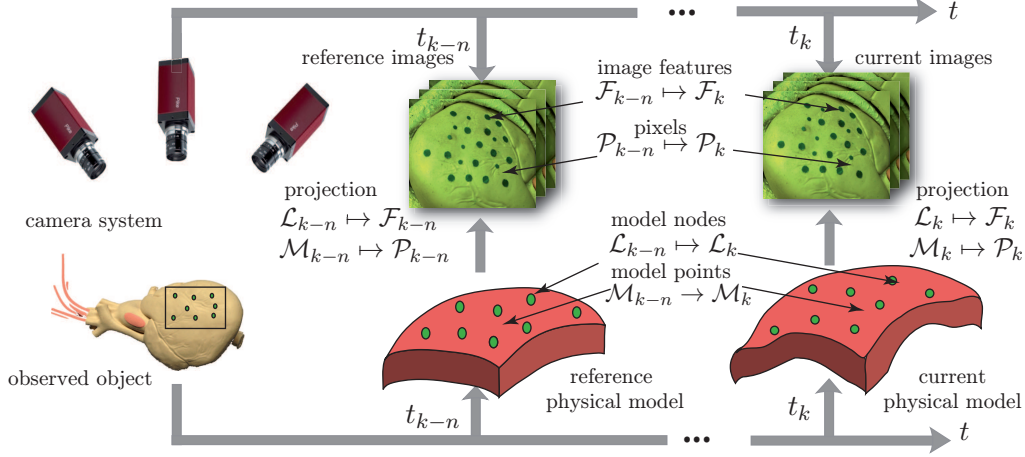


Figure 2. Key idea of the visual stabilization is a three-dimensional physics-based image transformation incorporating physical constraints provided by the three-dimensional physical model of the observed object. The transformation function maps the pixels $\mathcal{P}_{k-n}$ in the reference image acquired at time step t_{k-n} to the pixels $\mathcal{P}_k$ in the current image at time step t_k . This mapping is based on the correspondences between the image features $\mathcal{F}_k$ and $\mathcal{F}_{k-n}$ in both images (green marker) and is defined by a projection of the three-dimensional model with all its points $\mathcal{L}_k$ and $\mathcal{M}_k$ in camera images.

function should be derived. It should be able to improve the quality of the image transformation and accompanied models by using the feedback from the stabilized image sequence. In the course of this, the transformation errors should be reduced in the image regions where visual stabilization is inaccurate. For that purpose, the accuracy of the visual stabilization should be continuously evaluated. Furthermore, for avoiding deterioration of the image transformation accuracy, the uncertainties of the transformation function and measurements should be taken into account.

III. KEY IDEA

In this section, the key idea of the adaptive model-based visual stabilization, such as adaptation of the visual stabilization by the feedback from the stabilized image sequence, will be illustrated. For that purpose, a short introduction to a model-based image transformation incorporating physical constraints is given.

As described in [20], the model-based image transformation is based on a three-dimensional physical model, which is constructed on a set of three-dimensional points, i.e., model nodes

$$\mathcal{L}_k := \{\mathbf{l}_k^i\}_{i=1}^L$$

bounded by the model domain $\mathcal{L}_k \subseteq \Omega \subset \mathbb{R}^3$. These points may represent the landmarks on the surface of the object, as shown in Fig. 2. In this case, the three-dimensional position $\mathbf{l}_k^i$ of every landmark is reconstructed from the image features in the reference images $\mathcal{F}_{k-n} \mapsto \mathcal{L}_{k-n}$ according to [22].

When the object deforms, the model reproduces the object deformation. Then, the current three-dimensional positions $\mathbf{m}_k \in \Omega \subset \mathbb{R}^3$ of all model points from a set

$$\mathcal{M}_k := \{\mathbf{m}_k^i\}_{i=1}^M$$

are defined by approximating the deformations between the model nodes $\mathcal{L}_k \subset \mathcal{M}_k$. The parameters of this approximation represent the parameters $\mathcal{C}_k$ of the image transformation

function T_k . Finally, by projecting the reference position of the model points $\mathcal{M}_{k-n}$ and the current position of these points $\mathcal{M}_k$ in the respective images, the correspondences between the pixels $\mathcal{P}_{k-n} \mapsto \mathcal{P}_k$ of the reference image and the current image are defined.

It should be noted that the numerical accuracy of the physical model and therefore, of the image transformation function, depends on the number and the distribution of the model nodes $\mathcal{L}_k$. For example, for coping with strong deformations, a higher density of these nodes is desired. Their insufficient number can cause high discretization errors. However, the more nodes are in the model, the higher is the computational complexity of the transformation. Therefore, in this paper, the number of the model nodes is determined by using the *feedback from the stabilized image*. After evaluating the image transformation quality, the transformation errors are corrected by inserting additional model nodes in the areas where the image transformation can be improved. This allows to reduce the discretization errors and to adjust in these areas the physical behavior of the model to the behavior of the observed object.

IV. ADAPTIVE MODEL-BASED VISUAL STABILIZATION

The main components of the adaptive model-based stabilization such as image transformation, underlying models, estimation and adaptation will be described in this section. In Fig. 3, the functional interaction of these components is depicted.

A. Image Transformation

The image transformation function T_k in equation (1) sets correspondences between the pixels in the current and reference images. Due to the fact that these images represent camera projections of the observed heart surface, this function

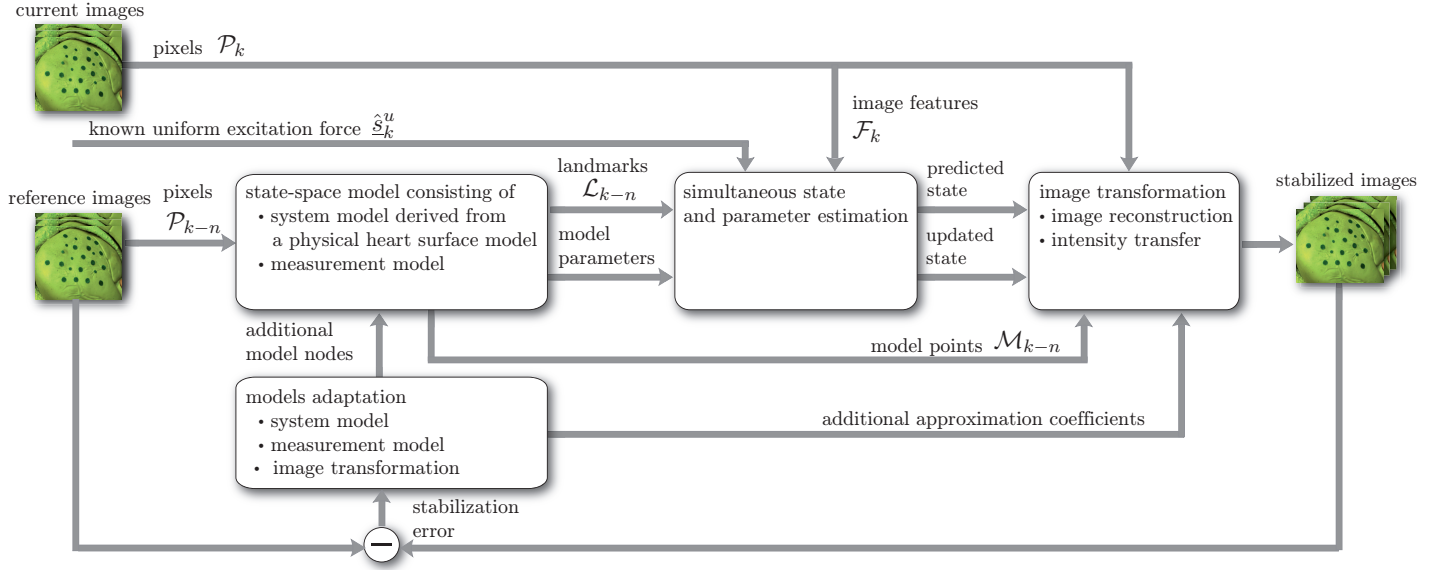


Figure 3. Adaptive visual stabilization of the image sequences with its main components: image transformation, underlying models, estimation and adaptation by the feedback mechanism.

is defined in [20] for every pixel in the reference image by

$$\begin{aligned} p_{k-n}^x &= p_k^x - \frac{r_{12}u_k^x + r_{12}u_k^y + r_{12}u_k^z + r_{14}}{r_{31}u_k^x + r_{32}u_k^y + r_{33}u_k^z + r_{34}}, \\ p_{k-n}^y &= p_k^y - \frac{r_{21}u_k^x + r_{22}u_k^y + r_{23}u_k^z + r_{24}}{r_{31}u_k^x + r_{32}u_k^y + r_{33}u_k^z + r_{34}}, \end{aligned} \quad (2)$$

where r_{ij} , $i = 1, \dots, 3$, $j = 1, \dots, 4$ represents the elements of the projection matrix $\mathbf{P}$ provided by a calibration of the cameras [23]. This transformation function depends on the three-dimensional displacement of the observed point on the heart surface $\mathbf{u}_k := [\mathbf{u}_k^x, \mathbf{u}_k^y, \mathbf{u}_k^z]^T$. Therefore, when the current displacements of the observed object are exactly known, accurate correspondences between the pixels of both images can be established. Then, the intensities of the pixels in the current image can be assigned to the corresponding pixels in the reference image. However, the displacements of the observed object are not exactly known, since the reconstruction of the three-dimensional information from camera data is corrupted by errors. Furthermore, approximation errors arise, because only the points $\mathcal{F}_k$ are measured and the other points are approximated. Therefore, the displacements of the entire object should be estimated.

B. Models of Observed Object and Measurement Process

For estimating the displacements of the heart surface, a state-space system including system and measurement equations is derived. The *system equation* defines a temporal propagation of the system state that characterizes the heart surface displacements. The specialty of the proposed system equation is the exploitation of physical background knowledge of the observed object in form of its physical model. The intervention area of the heart surface is modeled as a linear elastic physical body. The behavior of this body is described by a system of stochastic partial differential equations given

in [22]. Using the meshless collocation method [24] and implicit Euler integration, these equations are converted into a discrete state-space form [22], [25]. In this paper, the state-space model proposed in [20] is extended regarding the material inhomogeneity of the heart surface. The extended model is given by

$$\mathbf{x}_{k+1} = \mathbf{A}_k(\psi_k) \mathbf{x}_k + \mathbf{B}_k(\psi_k) (\hat{\mathbf{s}}_k^u + \mathbf{s}_k^n + \mathbf{w}_k), \quad (3)$$

where model uncertainties $\mathbf{w}_k$ are assumed white zero-mean Gaussian $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_k^w)$ with covariance $\mathbf{C}_k^w$. The unknown physical nodal parameters, such as material density ρ_k , Rayleigh damping coefficients $\eta_{1,k}$, $\eta_{2,k}$, Young's modulus vector $\mathbf{E}_k \in \mathbb{R}^L$, and Poisson's ratio ν_k are collected in the vector

$$\psi_k := [\mathbf{E}_k^T, \rho_k, \eta_{1,k}, \eta_{2,k}, \nu_k]^T. \quad (4)$$

They determine the physical characteristics of the model in the areas of the model nodes $\mathcal{L}_k$. In contrast to [20], in this paper, the Young's modulus $\mathbf{E}_k$ is defined for every model node, since the heart surface tissue is inhomogeneous. The heart excitation is assumed to be generated by a known uniformly distributed pressure inside the cardiac chambers $\hat{\mathbf{s}}_k^u$ and unknown non uniform system input $\mathbf{s}_k^n$. The system state

$$\mathbf{x}_k := [\mathbf{c}_k^T, \mathbf{c}_k^{dT}]^T \quad (5)$$

contains the three-dimensional values of L unknown coefficients $\mathbf{c}_k^i = [\mathbf{c}_k^{x,i}, \mathbf{c}_k^{y,i}, \mathbf{c}_k^{z,i}]^T \in \mathcal{C}_k$, also called nodal values, and their discrete derivatives $\mathbf{c}_k^d = (\mathbf{c}_k - \mathbf{c}_{k-1}) / \Delta t$, which are used for approximating the displacement $\mathbf{u}_k$ and velocity of every point on the heart surface. Here, the vector $\mathbf{c}_{k-1}$ denotes the coefficients at previous time step, and Δt stays for the time difference between the time steps. So, the three-dimensional displacement of a model point $\mathbf{m}_{k-n}$ from the

set $\mathcal{M}_{k-n}$ is approximated by

$$\underline{\mathbf{u}}_k \approx \Phi_k \underline{\mathbf{c}}_k, \quad (6)$$

where matrix

$$\Phi_k = \text{diag} \left\{ \varphi_k^T, \varphi_k^T, \varphi_k^T \right\} \quad (7)$$

consists of vector

$$\varphi_k = \left[\varphi_k(\underline{m}_{k-n}, l_{k-n}^1, \underline{\sigma}_k) \cdots \varphi_k(\underline{m}_{k-n}, l_{k-n}^L, \underline{\sigma}_k) \right]^T, \quad (8)$$

which elements are moving least-squares meshless functions $\varphi_k(\cdot)$, computed according to [26]. The support domain of these functions $\underline{\sigma}_k$ defines the relationship between the point $\underline{m}_{k-n}$ and model nodes $\mathcal{L}_k$. This domain is different for every point. In the proposed model, it is proportional to the distance between the point and its nearest neighboring nodes. For constructing the functions, for every point a minimum of 6 neighboring nodes is necessary.

The relationship between the system's state and measurements provided by a camera is described by the *measurement equation*, which is defined according to [20] for every image feature from the set $\mathcal{F}_{k-n}$ by

$$\begin{aligned} \hat{f}_k^x - f_{k-n}^x &= \frac{r_{12}\mathbf{u}_k^x + r_{12}\mathbf{u}_k^y + r_{12}\mathbf{u}_k^z + r_{14}}{r_{31}\mathbf{u}_k^x + r_{32}\mathbf{u}_k^y + r_{33}\mathbf{u}_k^z + r_{34}} + \mathbf{v}_k^x, \\ \hat{f}_k^y - f_{k-n}^y &= \frac{r_{21}\mathbf{u}_k^x + r_{22}\mathbf{u}_k^y + r_{23}\mathbf{u}_k^z + r_{24}}{r_{31}\mathbf{u}_k^x + r_{32}\mathbf{u}_k^y + r_{33}\mathbf{u}_k^z + r_{34}} + \mathbf{v}_k^y. \end{aligned} \quad (9)$$

The measured position of the image feature in the current camera image $\hat{\underline{f}}_k = [\hat{f}_k^x, \hat{f}_k^y]^T$ and the position of the image feature in the reference image $\underline{f}_{k-n} = [f_{k-n}^x, f_{k-n}^y]^T$ are extracted from images by a segmentation algorithm described in [22]. The measurement disturbances $\underline{\mathbf{v}}_k := [v_k^x, v_k^y]^T$ due to image noise and inaccurate projection are assumed white zero-mean Gaussian $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_k^v)$ with the covariance $\mathbf{C}_k^v$. It should be noted that the errors due to lens distortion are assumed negligible.

C. Estimation

In this section, the stochastic estimation of the heart surface displacement is described.

According to (3), the system state (5) involving approximation coefficients $\underline{\mathbf{c}}_k$ depends on the unknown parameters of the physical model, such as unknown nodal parameters $\underline{\psi}_k$ and non uniform heart excitation $\underline{\mathbf{s}}_k^n$. Therefore, these parameters should be simultaneously estimated with the system state. For that purpose, an arbitrary nonlinear estimator can be applied. In this paper, the estimation proposed in [20] is used.

As a result, when no measurements are available, e.g., in case of low image frame rate or total occlusions, the estimated three-dimensional displacements of every point from the set $\mathcal{M}_k$ at the current time step are characterized by mean and covariance

$$\hat{\underline{\mathbf{u}}}_k^p := \mathbf{H}_k \hat{\underline{\mathbf{x}}}_k^p, \quad \mathbf{C}_k^{u,p} := \mathbf{H}_k \mathbf{C}_k^{x,p} \mathbf{H}_k^T, \quad (10)$$

where the matrix $\mathbf{H}_k := \text{diag} \{ \Phi_k, \mathbf{0} \}$ contains the approximation functions (7) of this point. The vector $\hat{\underline{\mathbf{x}}}_k^p$ and matrix

$\mathbf{C}_k^{x,p}$ denote the first two moments of the predicted state $\underline{\mathbf{x}}_k$ represented by its mean and covariance. When measurements are available, the estimated three-dimensional displacements are also computed by (10), where the first two moments of the predicted state are replaced with the moments of the updated state, denoted by $\hat{\underline{\mathbf{u}}}_k^e$ and $\mathbf{C}_k^{u,e}$.

When the estimated three-dimensional displacements of the object are available, the correspondences between the pixels in the reference and current images can be determined using image transformation function given in equation (2). Finally, the current image can be transformed to the reference image by assigning the intensity of each pixel in the current camera image to each pixel in the reference image.

D. Adaptation of the Image Transformation

In case of an inaccurate physical model, the quality of the image transformation can deteriorate. It should be pointed out that the quality of the model depends on the number and the distribution of the points $\mathcal{L}_k$ used for constructing the model. As was described in Section IV-B, the coupling between these points depends on their support domains. On the one hand, the larger is the support domain of the point, the more points influence the displacement of this point. On the other hand, a small support domain of the point leads to the independence of its motion from distant points. Therefore, the interconnection between the points can be governed by changing size of the support domain. However, this is hardly possible if the model is initialized based on a small number of model nodes. The reason for that is that the approximation matrix Φ_k becomes singular if the demand on the minimum number of the neighboring nodes is not satisfied.

The insufficient number of model nodes $\mathcal{L}_k$ represents a discretization problem that can be solved by introducing additional model nodes. However, the more model nodes construct the model, the higher is the computational complexity of the image transformation. In itself, one model node introduces ten state variables in the proposed model, which are then propagated by a nonlinear estimator. In order to constrain the computational complexity and improve the accuracy of the image transformation, the transformation function and the models are adapted on the basis of the feedback from the stabilized image. For that purpose, at first, the stabilization errors are defined and detected. Then, the models are extended by additional model nodes, which are placed in regions, where the stabilization can be improved. Furthermore, the physical parameters are adapted in these regions. As a result, the image transformation function is enhanced.

1) *Feedback Computation*: For detecting the stabilization error, the difference image is constructed by subtracting the obtained compensated image from the reference image. These images are collected over a certain time interval, which is defined here by a period of the heart motion. It should be noted that the reference image is continuously updated at this time interval for considering the changing structure of the heart surface, e.g., by cutting arteries or due to bleeding. Then, the obtained image is converted to a gray scale intensity image and binarized, where only the gray values inside the defined gray scale level are converted to white pixels, as shown in Fig. 4.

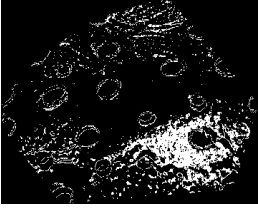


Figure 4. Stabilization error as a binarized difference between the compensated and reference image. The binarization allows to constrain the number of points, which are of interest for the feedback.

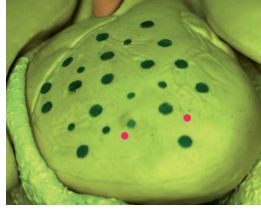


Figure 5. Additional model points (red) inserted for reducing the stabilization error. The small green marker are used for evaluating stabilization quality. The large green marker represent the selected image points.

The binarization allows to filter out the intensity differences due to changing light conditions and to constrain the number of points, which are of interest for the feedback. As a result, the image points with high stabilization error are identified.

2) *Model Adaptation*: For improving the quality of the stabilization, additional model nodes are introduced in the model in the areas, where the stabilization is inaccurate. In this section, the position of these points is defined and the changes caused in the models and the transformation function are explained.

The pixels, where a high stabilization error occurs, are uniquely assigned to the three-dimensional points by the model. From these, those points are selected

$$\mathcal{A}_k := \{\underline{a}_k^i\}_{i=1}^D, \mathcal{A}_k \subset \mathcal{M}_k$$

from the set $\mathcal{M}_k$ that have the highest divergence between their predicted and estimated position. The estimated position is used, as no measurement information about the motion of these points is available. The divergence is evaluated by the Mahalanobis distance [27]

$$e_k := (\hat{\underline{u}}_k^p - \hat{\underline{u}}_k^e)^T (\mathbf{S})^{-1} (\hat{\underline{u}}_k^p - \hat{\underline{u}}_k^e), \quad (11)$$

where the predicted $\hat{\underline{u}}_k^p$ and estimated $\hat{\underline{u}}_k^e$ displacements of the selected point are computed according to (10) using the first two moments of the a priori and a posteriori state. The matrix $\mathbf{H}_k$ contains the approximation functions of these points and the error covariance matrix is defined by

$$\mathbf{S} := \mathbf{C}_k^{u,p} + \mathbf{C}_k^{u,e}. \quad (12)$$

The points with the distance e_k larger than the parameter k are selected for inserting into the model. The parameter k ensures with the certain selected probability $P_k := P(e_k < k)$ sufficient prediction quality.

The adaptation of the physics-based model occurs by inserting at every defined position at least two model nodes, which are placed on the upper and lower surface of the model. In this way, the number of the model nodes $\mathcal{L}_k$ used for the construction of the heart surface model is enhanced by the inserted nodes, so that the set of the model nodes is now defined by

$$\mathcal{L}_k^a := \{\underline{l}_{k,i}^a\}_{i=1}^{L+A}, \underline{l}_k^a \in \mathcal{L}_k \cup \mathcal{A}_k. \quad (13)$$

This leads to the extension of the state vector (5) by new coefficients

$$\underline{x}_k^a := [\underline{c}_k^T, \underline{c}_k^{a,T}, \underline{c}_k^{d,T}, \underline{c}_k^{d,a,T}]^T, \quad (14)$$

where the coefficients $\underline{c}_k^a$ and $\underline{c}_k^{d,a}$ represent the nodal values of the inserted nodes. Furthermore, the vector of the unknown parameters (4) is enlarged, since the elasticity of the heart tissues in the area of the inserted nodes is defined by corresponding Young's modulus vector $\underline{E}_k^a$. Moreover, the matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ in equation (3) are computed using the adjusted parameter vector and the new approximation matrix (7), wherein the inserted nodes are considered. It should be noted that the support domain of the already existing model nodes $\mathcal{L}_k$ is not changed, when the new nodes are inserted. The support domain of the additional model nodes is defined by a minimum number of the neighboring nodes.

The adapted measurement model is computed according to (9), where the displacement $\underline{u}_k$ is calculated by (6) using the extended approximation matrix.

3) *Adapted Image Transformation Function*: Due to adaptation of the number of the model nodes, the image transformation function is also adjusted. It is computed by (2), where the three-dimensional displacement of the object is estimated by (10). In both equations, the adjusted state $\underline{x}_k^a$ and the extended approximation matrix are used.

V. EVALUATION

The adaptive model-based visual stabilization is evaluated in an experiment with a pressure-regulated artificial heart. The obtained results are compared with the standard geometric approach for visual stabilization.

A. Experimental Setup

The experimental setup shown in Fig. 6 is used in this paper. Here, an artificial beating heart is observed by a trinocular camera system. This system is installed at a distance of 50 cm



Figure 6. In computer-assisted beating heart surgery, an intervention area of a beating heart should be visually stabilized. With regard to this application, the artificial beating heart is observed by a camera system.

from the observed object. It consists of three PIKE F-210C

cameras [28] with a resolution of $1920 \text{ pixels} \times 1080 \text{ pixels}$. The camera baselines are about 57 cm, their focal length is about 35 mm. The image size of every camera is reduced by cutting out the defined regions of interest. The motion of the artificial heart is induced by a pressure signal with the amplitude 100 hPa and frequency 1.2 Hz.

The evaluation results are averaged over three runs, every of each consists of three image sequences including 400 images. Every of the image sequences is transformed to the reference image. The physical model of the heart surface is initialized based on the set of three-dimensional landmarks $\mathcal{L}$ reconstructed from the set of the image features $\mathcal{F}$. These image features are represented by large green marker in Fig. 5. The small green marker are inserted for the evaluation of the visual stabilization. They are not used for initialization of the model and their measurements are ignored by the estimation. The red points are inserted in the physical model based on the feedback from the stabilized image sequence. For the purpose of the feedback, the stabilization error is computed over 23 frames, whereby the intensity values between 8 and 12 gray levels are binarized.

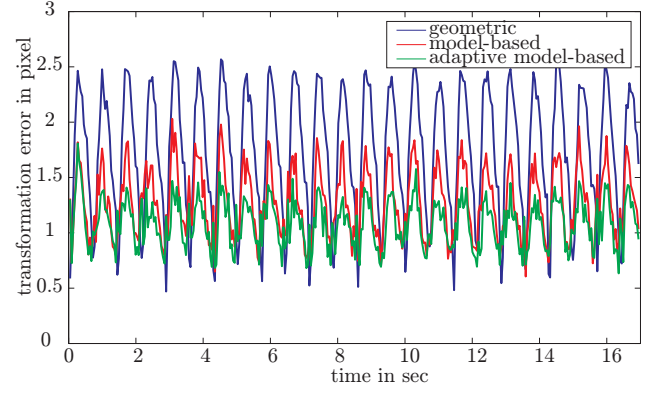
B. Experimental Results

The quality of the proposed adaptive method for visual stabilization is compared to the geometric image transformation that is introduced in [21]. The experimental results depicted in Fig. 7 show the transformation error averaged over three image sequences.

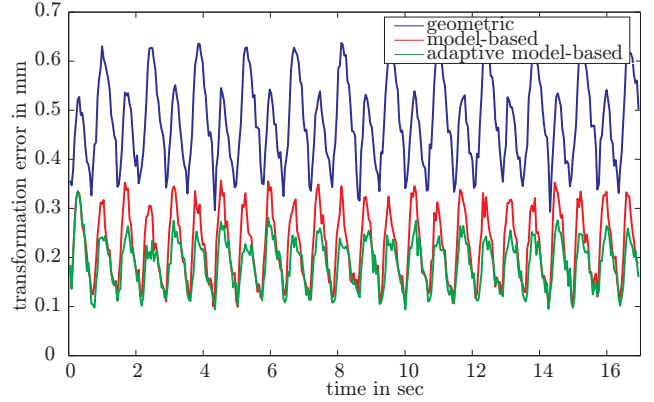
In two dimensions, this error is defined as the Euclidean distance between the positions of the image features in the reference image and stabilized images. The image features are extracted by a segmentation algorithm with sub-pixel accuracy. This algorithm is proposed in [22]. For evaluating the three-dimensional transformation error, the reference images and the stabilized images of all cameras are considered. In this way, the three-dimensional positions of evaluation points in reference and stabilized images can be reconstructed. Then, the Euclidean distance between the reconstructed positions of the reference points and stabilized points is computed.

As shown in Fig. 7, the accuracy of the model-based image transformation, which is identified by the maximum of the two-dimensional transformation error, is 20% higher than the accuracy of the geometric approach. Furthermore, the quality of transformation is increased by the feedback by 20%. In three dimensions, the accuracy of the adaptive model-based visual stabilization is 62% higher than the accuracy of the geometric approach and 29% higher than the accuracy of the model-based visual stabilization without feedback. It should be noted that the three-dimensional motion of the evaluation points achieve 10 mm. The displacement of the image features motion achieve 36 pixel.

While the error presented in Fig. 7 depends on the distribution of the evaluation points, the contour plots in Fig. 8 illustrate the performance of the approach over the entire stabilized area for one of the image sequence. The high quality of the model-based image transformation is evident. In the regions highlighted in Fig. 8(d), this transformation is significantly improved by the feedback mechanism.



(a) The two-dimensional transformation error.



(b) The three-dimensional transformation error.

Figure 7. The adaptive model-based visual stabilization in comparison to the model-based visual stabilization without feedback and the geometric approach. The transformation error is averaged over three evaluation runs containing 400 images. According to the maximum transformation error, the adaptive model-based image transformation provides 40% higher accuracy than the geometric image transformation in two dimensions. In three dimensions, the accuracy of the adaptive model-based visual stabilization is 62% higher than the accuracy of the geometric approach.

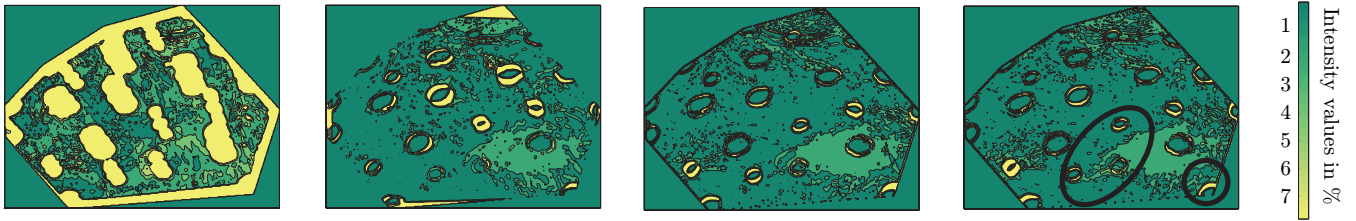
VI. CONCLUSIONS

The visual stabilization is highly important in many industrial and medical applications. It compensates changes of the scene caused by a motion and deformation of an observed object.

In this paper, visual stabilization is formulated as a model-based image transformation that incorporates a physical model of the observed object and a model of the measurement process. In this way, the physical characteristics of the observed object are considered.

The specialty of the proposed transformation is the continuous improvement of the image transformation quality by feedback from the stabilized image sequence. On the one hand, discretization errors are reduced by introducing additional model points in areas with high stabilization errors. On the other hand, the physical behavior of the model is adjusted to the behavior of the observed object.

The performance of the proposed method is evaluated with regard to computer-assisted beating heart surgery. It aims to be used for the virtual representation of the beating heart as



(a) Averaged difference between camera images and the reference image. (b) Geometric image transformation. (c) Model-based image transformation without feedback. (d) Adaptive model-based image transformation with feedback. (e) Error.

Figure 8. Contour plots representing the stabilization error, which is defined for one image sequence as a difference between the stabilized image and reference image. The error is averaged over 400 time steps. The model-based image transformation provides more accurate stabilization than the geometric image transformation. The quality of the model-based transformation is significantly improved by the feedback in the regions highlighted in Fig. 8(d).

motionless. This allows to extend the surgeon's capabilities during an operation on a beating heart and achieves a higher precision of surgical interventions. In an experiment with a pressure-regulated artificial heart, the proposed method provides higher accuracy than standard methods. The importance of the continuous monitoring of stabilization quality is emphasized by a significant improvement of stabilization accuracy due to feedback mechanism.

VII. ACKNOWLEDGMENTS

This work was partially supported by the German Research Foundation (DFG) within the Research Training Groups RTG 1126 "Intelligent Surgery - Development of new computer-based methods for the future working environment in visceral surgery"

REFERENCES

- [1] H. Nakagaki, H. Nakagaki, H. Nakagaki, and H. Nakagaki, "Study of Deformation and Insertion Tasks of a Flexible Wire," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1997.
- [2] D. Kumar and C. Jawahar, "Visual Servoing in Presence of Non-Rigid Motion," *International Conference on Pattern Recognition*, 2006.
- [3] S. Yue and D. Henrich, "Manipulating Deformable Linear Objects: Attachable Adjustment-Motions for Vibration Reduction," *Journal of Robotic Systems*, vol. 18(7), pp. 375–389, 2001.
- [4] Y. Nakamura, K. Kishi, and H. Kawakami, "Heartbeat Synchronization for Robotic Cardiac Surgery," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2001)*, Seoul, Korea, May 2001, pp. 2014–2019.
- [5] V. Falk, "Manual Control and Tracking - A Human Factor Analysis Relevant for Beating Heart Surgery," *The Annals of Thoracic Surgery*, vol. 74, pp. 624–628, 2002.
- [6] R. Richa, P. Poinet, and C. Liu, "Three-dimensional Motion Tracking for Beating Heart Surgery Using a Thin-plate Spline Deformable Model," *International Journal of Robotics Research*, vol. 20, no. 2-3, pp. 218–230, Feb. 2010.
- [7] T. Ortmaier, M. Groeger, D. H. Boehm, V. Falk, and G. Hirzinger, "Motion Estimation in Beating Heart Surgery," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 10, pp. 1729–1740, Oct. 2005.
- [8] Özkan Bebek and M. C. Çavuşoğlu, "Intelligent Control Algorithms for Robotic-Assisted Beating Heart Surgery," *IEEE Transactions on Robotics*, vol. 23, no. 3, pp. 468–480, Jun. 2007.
- [9] S. G. Yuen, P. M. Novotny, and R. D. Howe, "Quasiperiodic Predictive Filtering for Robot-Assisted Beating Heart Surgery," in *Proceeding of the IEEE International Conference on Robotics and Automation (ICRA 2008)*, Pasadena, CA, USA, May 2008, pp. 3875–3880.
- [10] 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 *Proceedings of the Medical Image Computing and Computer Assisted Interventions (MICCAI 2005)*, vol. 2, 2005, pp. 139–146.
- [11] T. Bader, A. Wiedemann, K. Roberts, and U. D. Hanebeck, "Model-based Motion Estimation of Elastic Surfaces for Minimally Invasive Cardiac Surgery," in *Proceedings of the 2007 IEEE International Conference on Robotics and Automation (ICRA 2007)*, Rome, Italy, Apr. 2007, pp. 2261–2266.
- [12] E. Bogatyrenko, U. D. Hanebeck, and G. Szabo, "Heart Surface Motion Estimation Framework for Robotic Surgery Employing Meshless Methods," in *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*, October 2009.
- [13] T. J. Gilhuly, S. E. Salcudean, and S. V. Lichtenstein, "Evaluating Optical Stabilization of the Beating Heart," *IEEE Engineering in Medicine and Biology Magazine*, pp. 133–140, 2003.
- [14] D. Stoyanov and G. Yang, "Stabilization of Image Motion for Robotic Assisted Beating Heart Surgery," *Medical Image Computing and Computer-Assisted Intervention*, vol. 10, pp. 417–424, 2007.
- [15] M. Gröger and G. Hirzinger, "Image Stabilisation of the Beating Heart by Local Linear Interpolation," in *Medical Imaging: Visualization, Image-Guided Procedures and Display*, 2006.
- [16] B. Zitová and J. Flusser, "Image Registration Methods: a Survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.
- [17] C. A. Glasbey and K. V. Mardia, "A Review of Image Warping Methods," *Journal of Applied Statistics*, vol. 25, pp. 155–171, 1998.
- [18] G. Wolberg, "Image Morphing: a Survey," *The Visual Computer*, vol. 14, pp. 360–372, 1998.
- [19] S. Sclaroff and J. Isidoro, "Active Blobs: Region-Based, Deformable Appearance Models," *Computer Vision and Image Understanding*, 2003.
- [20] E. Bogatyrenko and U. D. Hanebeck, "Visual Stabilization of a Beating Heart Motion by Model-Based Transformation of Image Sequences," in *Proceedings of the 2011 American Control Conference (ACC 2011) (to appear)*, San Francisco, California, USA, Jun. 2011.
- [21] N. Arad and D. Reisfeld, "Image Warping Using Few Anchor Points and Radial Funcions," 1994.
- [22] E. Bogatyrenko, P. Pompey, and U. D. Hanebeck, "Efficient Physics-Based Tracking of Heart Surface Motion for Beating Heart Surgery Robotic Systems," *International Journal of Computer Assisted Radiology and Surgery (IJCARS 2010)*, vol. 6, no. 3, pp. 387–399, Aug. 2010, doi:10.1007/s11548-010-0517-5.
- [23] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2000.
- [24] Atluri, *The Meshless Method (MLPG) for Domain and BIE Discretisation*. Tech Science Press, 2004.
- [25] E. Bogatyrenko and U. D. Hanebeck, "Simultaneous State and Parameter Estimation for Physics-Based Tracking of Heart Surface Motion," in *Proceedings of the 2010 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2010)*, Salt Lake City, Utah, Sep. 2010.
- [26] G.-R. Liu, *Mesh Free Methods: Moving Beyond the Finite Element Methods*. CRC Press, 2003.
- [27] P. Mahalanobis, "On the Generalised Distance in Statistics,," in *Proceedings of the National Institute of Science of India*, vol. 12, pp. 49–55, 1936.
- [28] *PIKE Technical Manual V4.2.0*, Allied Vision Technologies GmbH, Sep. 2009.