Published in final edited form as:

Comput Med Imaging Graph. 2019 March ; 72: 1–12. doi:10.1016/j.compmedimag.2018.12.001.

Uncertainty-aware Asynchronous Scattered Motion Interpolation Using Gaussian Process Regression

Bojan Koccev^{a,b,c,*}, Horst Karl Hahn^{b,c}, Lars Linsen^d, William M. Wells^e, and Ron Kikinis^a

^aDepartment of Mathematics and Computer Science, University of Bremen, Bremen, Germany

^bFraunhofer Institute for Medical Image Computing MEVIS, Bremen, Germany

^cDepartment of Computer Science and Electrical Engineering, Jacobs University Bremen, Bremen, Germany

^dInstitute of Computer Science, Westfälische Wilhelms-Universität Münster, Germany

^eDepartment of Radiology, Harvard Medical School and Brigham and Women's Hospital, Boston MA 02115 USA

Abstract

We address the problem of interpolating randomly non-uniformly spatiotemporally scattered uncertain motion measurements, which arises in the context of soft tissue motion estimation. Soft tissue motion estimation is of great interest in the field of image-guided soft-tissue intervention and surgery navigation, because it enables the registration of pre-interventional/pre-operative navigation information on deformable soft-tissue organs. To formally define the measurements as spatiotemporally scattered motion signal samples, we propose a novel motion field representation. To perform the interpolation of the motion measurements in an uncertainty-aware optimal unbiased fashion, we devise a novel Gaussian process (GP) regression model with a non-constant-mean prior and an anisotropic covariance function and show through an extensive evaluation that it outperforms the state-of-the-art GP models that have been deployed previously for similar tasks. The employment of GP regression enables the quantification of uncertainty in the interpolation result, which would allow the amount of uncertainty present in the registered navigation information governing the decisions of the surgeon or intervention specialist to be conveyed.

Keywords

Gaussian processes; interpolation; motion estimation; uncertainty

1. Introduction

Registration of pre-interventional/pre-operative navigation information on deformable soft-tissue organs requires estimating the motion that soft-tissue organs undergo during an intervention or surgery (Baumhauer et al. (2008)). The motion is to be estimated in an uncertainty-aware fashion, so as to allow conveying the amount of uncertainty present in the

*Corresponding author: bkocev@uni-bremen.de (Bojan Koccev).

registered navigation information influencing the surgeon's or intervention specialist's decisions (Risholm et al. (2010)).

For accurate motion estimation, one needs to fuse motion measurements (i.e., signal samples) with motion prior (e.g., motion dynamics, shape, etc.) information. According to the proposed general *data centric* taxonomy by Khaleghi et al. (2013), the motion signal samples may be imperfect, correlated, inconsistent, and/or in disparate forms/modalities. The imperfection aspect is manifested through uncertainty (Irani and Anandan (2000); Kanazawa and Kanatani (2001); Leedan and Meer (2000); Zhou et al. (2004, 2005)), imprecision, and/or granularity. The inconsistency problem is in terms of conflict, outlier (spurious data), and/or disorder (out-of-sequence data). Furthermore, the imprecision is expressed in the form of vagueness, ambiguity, and/or incompleteness.

In soft tissue navigation, the motion measurements information source is, in general, composed variously of real-time surface (e.g., electromagnetic (EM) (Kocev et al. (2014); Zhang et al. (2006)), optical (OP) (Meinzer et al. (2008)), point cloud (Cash et al. (2005); Kocev et al. (2013)), etc.) and volumetric (e.g., ultrasound-ultrasound (US-US) correlation (Kocev et al. (2014); Wang et al. (2013)), etc.) data. In this regard, we believe that using such various multimodal tracking systems could allow for better sampling of the motion signal function and therefore could improve the motion estimation accuracy. In return, this could improve the soft tissue navigation accuracy and the overall outcome of the surgery or intervention. We assume that the tracked soft-tissue organ is represented in a discrete (point-based) fashion (Lim et al. (2015)), while its motion state at any discrete time point is assumed to be directly defined by the positions of all points that constitute its discrete representation. In this context, the motion measurements impose three main challenges: (1) they capture the real-time motion of the tracked soft-tissue organ at, in general, randomly non-uniformly scattered points that do not necessarily correspond either in number or physically to the state points (Kocev et al. (2014)), in this case the information source is incomplete; (2) the temporal resolution of different measurement mechanisms may vary, i.e., the measurements may arrive at multiple rates, in which case the source contains out-of-sequence data; (3) the measurements are contaminated with noise, i.e., they are uncertain. For an abstract visualization of the nature of motion measurements and the motion estimation problem itself, please see Figure 1.

To address the challenges imposed by the motion measurements information source in the context of soft tissue navigation, we propose an algorithm for uncertainty-aware interpolation of motion signal samples that are randomly, non-uniformly scattered in the spatiotemporal domain. To formally define out- of-sequence, i.e., asynchronous, motion signal samples, we propose a novel motion field representation (see Section 3). Our proposed algorithm employs Gaussian process (GP) regression (Rasmussen (2006)) (see Section 4) to perform the interpolation in an optimal unbiased fashion. The GP embeds the motion prior, which is then conditioned on the motion signal samples when performing the regression. The conditioning leads to an estimate of the (full) posterior distribution over motion signal functions, which could be directly used for summarizing and conveying the uncertainty (Risholm et al. (2010)) in any registered pre-interventional/pre-operative soft-tissue navigation information. In this regard, we believe that providing information about the

uncertainty in the registered pre-interventional/pre-operative soft-tissue navigation information to the surgeon or intervention specialist could improve their decisions and the overall outcome of the surgery or intervention. The work of Lüthi et al. (2011) is similar to ours on this point. They use a zero-mean GP to model the motion prior and perform regression on motion signal samples that are randomly non-uniformly scattered in the spatial domain. Their motion signal samples are, in contrast to ours, in-sequence, i.e., synchronous. Hence, they do not need to use any temporal information. In contrast to their work, we propose using a non-zero mean function (see Subsection 5.1) and enable the interpolation in the spatiotemporal domain. We demonstrate that non-zero-mean GPs outperform zero-mean ones independent of whether the data are randomly non-uniformly scattered in the spatial or in the spatiotemporal domain. To enable the interpolation in the spatiotemporal domain, we propose using a squared exponential covariance function with anisotropic distance measure (ADM) (Rasmussen and Williams (2006)) (see Subsection 5.2). The spatiotemporal GP formalism facilitates the prediction of tissue displacements at any location, and for any time interval, from observed deformations that are sparse in space and time. We evaluated our algorithm by interpolating simulated randomly non-uniformly spatiotemporally scattered uncertain motion signal samples and comparing the interpolation results with the respective simulated ground truth (see Section 6). As in some applications the randomly non-uniformly spatiotemporally scattered uncertain motion signal samples may be drawn over some restricted region of space, e.g., on the surface of the soft-tissue organ, we also evaluated our proposed method on such simulated measurements.

2. Related Work

Our proposed GP regression model can be seen as a deformation model (Sotiras et al. (2013)) because it is intended to be used for interpolation of soft, i.e., deformable, tissue motion signal samples. Sotiras et al. (2013) classified all deformation models according to what drives the geometric transformations that are computed by these models. According to their top-level classification, the geometric transformation could be: 1) inspired by physical models (Wassermann et al. (2014)), 2) inspired by interpolation/approximation theory (Ledesma-Carbayo et al. (2005); Perperidis et al. (2005); Rohr et al. (2001); Vandemeulebroucke et al. (2011); Wörz and Rohr (2008)), and/or 3) knowledge-based, i.e., embedding prior information regarding the sought deformation, (Glocker et al. (2009); Kocev et al. (2014); Lüthi et al. (2011)). For a more detailed classification of all deformation models, we refer the reader to the work of Sotiras et al. (2013). In the following, we discuss in detail the above classified related work and contrast it against our own work.

Rohr et al. (2001) employed approximating thin-plate splines (TPS) to account for both isotropic and anisotropic landmark localization errors, where the corresponding TPS result from a minimizing functional with respect to the sought transformation. They weight the quadratic/squared Euclidean distance between the corresponding landmarks according to the landmarks localization uncertainty and, in this way, control the influence of the landmarks on the registration result. In follow-up work, Wörz and Rohr (2008) improved the accuracy of the approximating TPS (Rohr et al. (2001)) by incorporating Gaussian elastic body splines (GEBS) resulting in a new approximation approach with an improved underlying physical deformation model. They used an energy functional related to the Navier equation

under Gaussian forces, while the landmarks are still individually weighted in a similar fashion according to their localization uncertainties. The use of Gaussian forces (physically more plausible as they do not diverge, and decrease with distance) provided them with a free parameter controlling the locality of the transformation. One problem in the two above approaches is that they treat the landmark and regularization terms independently, which could be improved because the landmarks can constrain the regularization term itself as well (Lüthi et al. (2011)). We incorporate the landmark information (in general, the motion signal samples) during the training phase as part of the regularization as a priori knowledge on the admissible deformations. Moreover, in contrast to our approach, they do not handle asynchronous data and do not estimate the uncertainty in the resulting transformation.

Several groups have investigated the problem of spatiotemporal image registration (Bersvendsen et al. (2016); Ledesma-Carbayo et al. (2005); Perperidis et al. (2005); Shi et al. (2013); Vandemeulebroucke et al. (2011)). Perperidis et al. (2005) addressed the spatiotemporal iconic, i.e., intensity-based, registration of 3D image sequences by devising 4D affine and free-form deformation (FFD) (based on a 4D B-Spline model) models that are separated into spatial and temporal components. Ledesma-Carbayo et al. (2005) investigated the spatiotemporal iconic registration of 2D image sequences by using a semi-local spatiotemporal parametric linear model, based on 3D B-splines, that is also separable in time and space. Both Perperidis et al. (2005) and Ledesma-Carbayo et al. (2005) place the space- and time-axis basis functions on a uniform rectangular spatial grid and regularly spaced time intervals, respectively. Ledesma-Carbayo et al. (2005) discuss in depth the choice of time- and space-axis scale parameters, governing 130 the knot spacing, which, in our case, would be handled by the characteristic length-scales of our covariance function (see Subsection 5.2). Shi et al. (2013) revisited the classic FFD approach and devised a novel sparse representation for FFD using the principles of compressed sensing. They reconstruct the deformation from a pair of images (or image sequences) and apply a sparsity constraint to the parametric space. They extended the sparsity constraint to the temporal domain and proposed a temporal sparse free-form deformation (TSFFD) model which enabled capturing of fine local details, e.g., motion discontinuities in the spatiotemporal domain. They addressed the trade-off between robustness and accuracy for FFD-based registration by deploying sparsity constraints as an additional regularization term. Bersvendsen et al. (2016) addressed the temporal alignment by optimizing the alignment of the normalized cross correlation (NCC)-over-time curves of the sequences (Perperidis et al. (2005) also employed NCC, in the optimization of their temporal component), within their proposed fully automatic method for spatiotemporal (spatially rigid) registration between 145 two partially overlapping 3D image sequences. However, Bersvendsen et al. (2016), Ledesma-Carbayo et al. (2005), Perperidis et al. (2005), and Shi et al. (2013) do not estimate the transformation in an uncertainty-aware fashion. Moreover, in contrast to our approach, they do not address the problem of interpolating randomly non-uniformly scattered spatiotemporal motion signal samples, one that arises in the context of spatiotemporal geometric, i.e., landmark- based, image registration.

Our proposed deformation model falls into the group of knowledge-based statistically constrained geometric transformations. Similar to Glocker et al. (2009), we determine a probability density function modeling the prior distribution of the sought deformation field.

In contrast to our approach, Glocker et al. (2009) use Gaussian mixture models (GMMs) to represent their probability density functions and employ a Markov random field (MRF)-based formulation of the registration problem when computing the maximum a posteriori (MAP) estimate of the deformation field on a regular grid of control points (as in FFDs). For any other point in the domain, they employ a B-spline basis functions-based interpolation between the estimated control point displacements. Regarding the MRF graph cost function, they set the edge penalty costs based on the negative log likelihood, which in return, is defined by their prior probability density functions. As a result, they compute a single optimal deformation estimate, i.e., the MAP estimate, while we estimate the full a posteriori deformation field distribution. Therefore, in addition, we estimate the uncertainty in the deformation estimate at any point in the domain. We would, in general, be able to incorporate their clustering idea by training different instances of our proposed GP model for each identified cluster.

To the best of our knowledge, the work of Lüthi et al. (2011) is the most similar to ours. They model the deformations as a zero-mean vector-valued GP and regard the landmarks as additional information on the admissible deformations. In the general setting, they reduce the vector-valued case to the scalar case by constructing a matrix-valued covariance function as the product of a scalar-valued covariance function and a symmetric positive definite $N \times N$ matrix (N is the number of output dimensions), encoding the a priori knowledge about the correlation between the output dimensions. Our assumption with respect to the independence between the motion signal output dimensions is equivalent to using an $N \times N$ diagonal matrix in this case. They make the same assumption in all of their test examples. Furthermore, in contrast to our approach, they work with input landmarks that define synchronous motion signal samples that are randomly non-uniformly scattered in the spatial domain. In contrast to their approach, we propose to use a non-zero mean GP and allow the motion signal samples to be randomly non-uniformly scattered in the 185 spatiotemporal domain. In a follow-up work, Gerig et al. (2014) devised a new method for spatially-varying (allowing for different regularization properties in different regions) iconic registration using GP priors. As a result, they came up with a non-stationary GP, that allowed to model different amount of smoothness in different regions. In this way, they were able to differentiate between tissue types or to make the regularization stronger in regions with noisy data. In theory, their proposed GP can have any mean function (Lüthi et al. (2013)). However, in practice they use a zero-mean function in all of their test examples, except for their statistical shape model for which they estimate the mean function based on example shapes. To present their approach in the medical setting, 195 they demonstrated a solution for the challenging problem of atlas-based skull registration of cone-beam CT images. Similar to Gerig et al. (2014), Zhao et al. (2017) presented an alternative method for spatially-varying registration which is also anisotropic. Zhao et al. (2017) presented physically realistic deformations for the task of surface registration, by modeling the surface as an orthotropic elastic thin shell. They devised a statistical framework (Physical-Energy-Based MRF model) that can be deployed for estimating spatially varying anisotropic shell elasticity parameters with the input being only a set of known surface deformations. In parallel to estimating the elasticity parameters, they estimate the registration as well. They applied their approach in the context of 3D endoscopy reconstruction, which requires to

generate a 3D reconstruction surface from multiple endoscopic movie frames. They managed to register all single-frame 3D reconstruction surfaces into a single surface. We view the novelties presented by Gerig et al. (2014) and Zhao et al. (2017) as complementary to the novelties presented in this paper.

Wassermann et al. (2014) employed a stochastic differential equation (SDE) for modeling the deformations as the evolution of a time-varying velocity field. The SDE is defined based on a deterministic ordinary differential equation (ODE) and a covariance function that is calculated as the matrix Green's function of the linear differential operator that regularizes the deterministic velocity field. The linear differential operator restricts the deformations to the space of diffeomorphisms. Hence, their framework is suitable for modeling large deformation diffeomorphisms. In principle, they place a zero-mean GP prior with the above covariance function on the stochastic velocity field. They focus on operators that regularize in space but not in time. In contrast to their approach, we regularize in the spatiotemporal domain and employ a non-zero-mean GP prior.

In other medical imaging contexts, GPs have been used to quantify the uncertainty in image segmentation of deformable objects by defining a probability distribution of image segmentation boundaries as a GP and measuring the effect of using various plausible segmentation samples therefrom (Lê et al. (2015)).

GPs have generally been employed in several areas of visual computing for interpolation of uncertain data (Schlegel et al. (2012); Stytz and Parrott (1993); Wachinger et al. (2014)). In all of these cases, the data are given on a uniform spatial grid, which renders the zero-mean GP prior assumption by Schlegel et al. (2012) and Wachinger et al. (2014) as not very critical. In other words, if one assumes a constant-mean latent process function and uniform data contaminated with Gaussian white noise, one could model a zero-mean process by subtracting the empirical mean from the input data and adding it back after processing, if necessary (Schlegel et al. (2012)). However, if the data are given on a structured grid (e.g., if one applies the method of Wachinger et al. (2014) in the context of non-rigid image registration), or on an unstructured grid, or are randomly non-uniformly scattered, we believe that the empirical mean of the data is then less representative of the latent process mean even if the latent process mean is constant throughout the domain mainly because of the presence of noise. If, in reality, there is a large deviation from the assumption of a constant-mean latent process function, then one cannot, by default, model, as described above, a zero-mean GP prior. To address these aspects, we propose non-constant-mean GP models that can handle data that are randomly non-uniformly scattered in the spatiotemporal domain. This imposes the challenges of identifying and training a suitable non-constant mean function as well as an appropriate covariance function that can handle the aspects of drift (Stytz and Parrott (1993)) and spatiotemporal distribution. The randomness and non-uniformity in the distribution of the given data in the spatiotemporal domain are also addressed during the training phase.

For a general overview of traditional/classical and generalized interpolation techniques, we refer to the work of Thévenaz et al. (2000a, b).

3. Motion Field Representation

We define the motion signal function as

$$u: \mathbb{R}^{N+2} \rightarrow \mathbb{R}^N, \quad (1)$$

where the first N dimensions in the input domain are used to specify the location of a given N -dimensional spatial point, and the last two are used to define the time interval in which the motion of this spatial point took place. Each $(N+2)$ - dimensional spatiotemporal point is mapped to an N -dimensional displacement vector specifying the motion that the given spatial point underwent within the given time interval. The displacement vector can be seen as the position of the given spatial point at the end of the time interval t_{end} , where the position is defined in a local coordinate system with the origin at the position of the spatial point at the beginning of the time interval t_{begin} . In contrast to the classical Lagrangian specification of a flow field, our representation of the motion field allows for specifying the motion between any time interval (t_{begin}, t_{end}) . This is especially required for specifying asynchronous motion signal samples.

4. Gaussian Process Regression

A GP is defined as a collection of random variables, any finite number of which have (consistent) joint Gaussian distributions (Rasmussen (2006)). A scalar-valued GP is uniquely defined through its mean $m(x): \Omega \rightarrow \mathbb{R}$ and covariance $k(x, x'): \Omega \times \Omega \rightarrow \mathbb{R}$ functions, where Ω is an index set. GPs can be used to define distributions over functions. We use the following notation:

$$f \sim \mathcal{GP}(m, k) \quad (2)$$

to denote that the function f is distributed as a GP with mean function m and covariance function k . Moreover, GP models can be used to formulate a Bayesian framework for regression. In that context, a GP is used as a prior for Bayesian inference. In order to make predictions for unseen test cases x_* , one needs to compute the posterior by conditioning the prior on a given training data set of n observations

$$\mathcal{D} = \{(x_i, y_i) \mid i = 1, \dots, n\}. \quad (3)$$

where x_i are the training data locations and y_i are the observations (usually noisy, see Subsection 5.3) of the function values $f(x_i)$. It is generally assumed that the noise is additive independent and identically distributed zero-mean Gaussian, i.e.,

$$y(x) = f(x) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma_n^2), y \sim \mathcal{GP}(m, k + \sigma_n^2 \delta_{ii'}) \quad (4)$$

where $\delta_{ii'}$ is the Kronecker's delta function, i.e., $\delta_{ii'} = 1$ if and only if $i = i'$. After conditioning the prior process on the observations, one obtains the posterior process as follows:

$$f | \mathcal{D} \mathcal{G}p(m_{\mathcal{D}}, k_{\mathcal{D}}) m_{\mathcal{D}}(x) = m(x) + \sum (X, x)^T (\sum + \sigma_n^2 I)^{-1} (y - \mu) \kappa_{\mathcal{D}}(x, x') = \kappa(x, x') - (5)$$

$$\sum (X, x)^T (\sum + \sigma_n^2)^{-1} \sum (X, x')$$

where $\Sigma(X, x)$ is a vector of covariances between every training case and x (analogous for $\Sigma(X, x')$), μ is the vector of the function mean values at the training data locations, i.e., $\mu = m(x_i)$, $i = 1, \dots, n$, Σ is the covariance between the training data, and I is the identity matrix.

In the context of motion estimation, we would generally need to use a vector-valued Gaussian process (Lüthi et al. (2011)). The displacement $u(x)$, at any input location x , would need to be modeled as an N -dimensional random vector. The mean and covariance functions would then need to have the following form:

$$\begin{aligned} m(x): \Omega &\rightarrow \mathbb{R}^N \\ k(x, x'): \Omega \times \Omega &\rightarrow \mathbb{R}^{N \times N}. \end{aligned} \quad (6)$$

However, we assume that the motion signal output dimensions are independent (see Section 2 for a discussion on how our work is related to the work of Lüthi et al. (2011) on this point). Therefore, we are able to employ a separate scalar-valued GP for each output dimension (Chan (2013)). For a more general treatment of vector-valued GPs, see Hein and Bousquet (2004) and Micchelli and Pontil (2005).

5. Model Selection

The Gaussian process model we employ is a hierarchical non-parametric (i.e., it needs access to all training data in the process of making predictions) model with two levels. At the first level are the free (hyper-) parameters θ of the underlying modeling functions (mean function (see Subsection 5.1), covariance function (see Subsection 5.2), and likelihood function (see Subsection 5.3)). The (hyper-) parameters control the distribution of the target values. At the second/top level, we have a (discrete) set of possible model structures, $\mathcal{H}_i$ out of which we should choose. On a side note, one could consider a zero level with the noise-free latent function values f at the training inputs as the parameters of the Gaussian process model. We select the model structure $\mathcal{H}_i$ at the second/top level by specifying the function families to which we believe the mean (see Subsection 5.1), covariance (see Subsection 5.2), and likelihood (see Subsection 5.3) functions belong. As we normally have only vague information about the (hyper-) parameters, we deploy a mechanism for learning them from the training data (see Subsection 5.4).

5.1. Mean Function

Generally, the mean function of a scalar-valued GP $f(x)$ is defined as follows

$$m(x) = \mathbb{E}[f(x)], \quad (7)$$

where $\mathbb{E}$ is the expectation. In order to keep notations simple, the mean function of the prior GP is often set to zero (Rasmussen and Williams (2006)). The zero-mean assumption for the prior GP does not imply that the posterior GP would be zero-mean, i.e., from this point of view, one could see this assumption as not being critical. However, several problems may arise with respect to the interpretability of the model, the expressiveness of prior information, etc. (Rasmussen and Williams (2006)). Schlegel et al. (2012) model a zero-mean process on their data by subtracting the empirical mean from the input data and adding it back after processing, if necessary. However, Kuss (2006) pointed out that, in general, the mean of the data is not necessarily the mean of the process. Furthermore, we argue that the use of the empirical mean in this way is even more critical when dealing with non-uniformly scattered training data in contrast to when the data are uniformly scattered or even on a Cartesian grid. Therefore, we propose to model the mean function of the prior GP explicitly in order to be able to specify an appropriate non-zero mean function. One could use a fixed (deterministic) mean function or alternatively a few fixed basis functions with a set of coefficients/parameters which would need to be inferred from the training data. Then, one normally optimizes over the hyperparameters of the covariance function (see Subsection 5.2) jointly with the parameters of the mean function when fitting the model (see Subsection 5.4) on the training data. We build upon the work of O'Hagan and Kingman (1978) and propose to couple their linear mean function $m_1(x) = \beta^T x$ with a constant mean function $m_2(x) = c$, when modeling the mean function of the GP that embeds the motion prior. The two functions are simply added, i.e.,

$$m(x) = m_1(x) + m_2(x) = \beta^T x + c \quad (8)$$

which results in a composite mean function (Rasmussen and Nickisch (2010)). Hence, the prior mean is realized as the sum of a linear and constant function. The parameters β and c are inferred from the data (see Subsection 5.4).

5.2. Covariance Function

In general, the covariance function of a scalar-valued GP $f(x)$ is defined as follows

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))]. \quad (9)$$

As such, the choice of a covariance function induces the properties, e.g., stationarity, isotropy, smoothness, periodicity, etc., of the functions that are likely under the GP prior (Barber (2012); Rasmussen and Williams (2006)). If the covariance function is a function only of $|x - x'|$ then it is stationary, while if it is a function only of $|x - x'|$ then it is isotropic

(Rasmussen and Williams (2006)). Suitable properties for the covariance function generally need to be learned from the training data. Some of these properties are encoded through the hyperparameters (e.g., characteristic length-scale, variance, etc.) of the chosen covariance function. In the context of doing regression on spatiotemporal motion signal samples, we need to use covariance functions whose input domain Ω is a subset of $\mathbb{R}^D$. In this regard, we propose using the Squared Exponential (SE) covariance function with distance measure with a different characteristic length-scale (hyper-) parameter per input dimension

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}(x - x')^T P^{-1}(x - x')\right) \quad (10)$$

where P is a diagonal matrix with length-scale (hyper-) parameters $l_1^2, l_2^2, \dots, l_D^2$ as entries on the diagonal, and σ_f^2 is the signal variance (Rasmussen and Williams (2006)). As this covariance function uses different length-scales l_i on different dimensions, it is an anisotropic covariance function. In simple words, P encodes how far, along individual axes in input space, the input locations need to be so that the function values at those locations become uncorrelated. In our case, this is particularly suited for anisotropically adjusting the distance measure along the spatial and temporal axes in our input space. In other words, we can handle the anisotropy between space and time. These (hyper-) parameters will be learned from the data (see Subsection 5.4). In general, the functions in the SE covariance function family are infinitely differentiable, i.e., a GP process with a covariance function from this family has mean square derivatives of all orders, i.e., it is very smooth. Therefore, the use of this covariance function could also be interpreted as a mechanism for regularization that penalizes non-smooth solutions (Lüthi et al. (2011)). We therefore believe that the SE covariance function is particularly suitable for modeling soft-tissue motion prior within our target application because soft-tissue organs undergo smooth motion.

5.3. Likelihood Function

We employ a Gaussian likelihood function

$$p(y|f, \sigma_n) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(y - f)^2}{2\sigma_n^2}\right) \quad (11)$$

for regression, where y is the actual observation/measurement of the true latent value (of a component of the displacement field) f (see Section 4), and σ_n^2 is the noise variance (hyper-) parameter. In other words, this function defines how the noisy observations/measurements are assumed to diverge from the noise-free function f . The incorporation of a Gaussian likelihood function together with the Gaussian process prior allows for a posterior Gaussian process over functions and keeps things analytically tractable (Rasmussen and Williams (2006)). Our Gaussian noise model assumes homoscedasticity, i.e., the noise variance is assumed to be constant throughout the domain. For modeling heteroscedastic Gaussian

noise, we refer to the work of Goldberg et al. (1997) who uses the noise variance as a function of x .

5.4. (Hyper-) parameters Training

The (hyper-) parameters are optimized by maximizing the probability of the model, given the training data. The probability of the model, given the training data, is computed based on the marginal likelihood (ML), also known as model evidence,

$$\log p(y|X, \theta, \mathcal{H}_i) = -\frac{1}{2}(y - \mu)^T \Sigma_y^{-1}(y - \mu) - \frac{1}{2} \log |\Sigma_y| - \frac{n}{2} \log 2\pi, \quad (12)$$

where $\Sigma_y = \Sigma + \sigma_n^2 \mathcal{I}$. The ML is a probability distribution over the observations y , conditioned on the input locations x , the (hyper-) parameters θ (i.e., $\theta = (\beta, c, l_1, l_2, \dots, l_D, \sigma_f, \sigma_n)$, see Subsections 5.1, 5.2, and 5.3), and the chosen model structure $\mathcal{H}_i$. The use of log ML is appropriate because it automatically incorporates a trade-off between data-fit (first term of Eq. 12, i.e. $-\frac{1}{2}(y - \mu)^T \Sigma_y^{-1}(y - \mu)$) and model complexity (second term of Eq. 12, i.e., $\frac{1}{2} \log |\Sigma_y|$), i.e., it does not require an external parameter for controlling this trade-off (Rasmussen and Williams (2006)). The third term of Eq. 12, i.e., $\frac{n}{2} \log 2\pi$, is a normalization constant. Generally, the mechanism of using the data to estimate the prior parameters is known as empirical Bayes (Gelman et al. (2014)). This approximation of the complete hierarchical Bayesian analysis eliminates the need to put a probability model over all (hyper-) parameters. In principle, instead of putting prior distributions over the (hyper-) parameters and marginalizing them out, the (hyper-) parameters are set with the values that maximize the ML. The use of ML avoids over-fitting, which is associated with the classical maximum likelihood approach, by marginalizing out the model parameters and allows comparison of different models on all training data, i.e., it eliminates the need to use cross-validation (Bishop (2006)). However, due to possible sensitivity of the ML on the prior, it is still recommended to use an independent test dataset for the comparison of different models in practical applications (Bishop (2006)). To demonstrate the effect of the characteristic length-scale parameter on the probability of an example model given example training data, we try to fit only the length-scale parameter l_j while the remaining hyperparameters are set in accordance with the process from which the example training data are drawn. The training data are drawn from a zero-mean GP with an SE covariance function with ADM defined over a 2D index set, with (hyper-) parameters $(l_1, l_2, \sigma_f, \sigma_n) = (0.25, 0.125, 1, 0.1)$. In Figures 2 and 3, one can observe the fitting of the characteristic length-scales l_1 and l_2 respectively on the training data. The plots show the corresponding log ML and the decomposition into its constituents as a function of the respective characteristic length-scale hyperparameter. In both plots, the negative complexity term increases (i.e., the model loses complexity) as the respective length-scale hyperparameter l_j increases. Furthermore, in both plots, the data-fit term decreases monotonically as the length-scale hyperparameter l_j increases, because the model loses

flexibility. The marginal likelihood in Figure 2 reaches its peak value for $l_1 = 0.25$, while in Figure 3 the peak of the marginal likelihood is reached for $l_2 = 0.125$. This agrees with the respective characteristic length-scales of the GP from which the training data are drawn.

In the general setting, we use conjugate gradients (Polak and Ribiere (1969)) to optimize the (hyper-) parameters jointly, while maximizing the probability of the model given the training data. To circumvent bad local extrema, we perform random restarts and select the (hyper-) parameters configuration that gives the maximum probability of the model given the training data.

6. Evaluation

In this section, we aim to demonstrate that, when doing regression on randomly non-uniformly scattered motion signal samples, employing a GP prior with the proposed non-zero-mean function yields better results than using a GP prior with a zero-mean function. Furthermore, we will show that, in the case that the observations of the function are randomly non-uniformly scattered in the spatiotemporal domain, our proposed covariance function with ADM is more suitable for defining the GP prior of a given motion signal function than the one of the same family with an isotropic distance measure (IDM). Therefore, we compare the obtained results when doing regression using each of the GP priors configurations in Table 1.

Each of the GP model configurations in Table 1 is used to perform regression on simulated soft-tissue motion data. We simulated ground-truth soft-tissue motion data (see Subsection 6.1) out of which a relatively small portion (contaminated with noise) was used as training data and the rest (noise-free) as ground-truth test data. The (hyper-) parameters of GP models having each of the above configurations are first trained (see Subsection 6.2) by maximizing the likelihood, as defined in Eq. 12, and then the prior is conditioned on the training dataset according to Eq. 5, when making predictions for unseen test cases. The predictions are then evaluated in terms of accuracy against the respective ground truth (see Subsection 6.3). We train three different scalar-valued GP models with the same configuration for each motion signal output dimension (assumed to be independent, see Section 4). Hence, predictions are made for each output dimension separately.

6.1. Data Simulation

The ground-truth motion data are simulated over time at the points that constitute the discrete representation of the given deformable object. For this purpose, we employed a finite element (FE) model (Georgii and Westermann (2008, 2005)) for physics-based simulation of motion data. For this evaluation, we created an FE model of the CIRS triple modality breast biopsy training phantom with 1,962 vertices (see Figure 4.) and used it to simulate a physically plausible non-linear motion that a soft-tissue organ is likely to undergo during a biopsy intervention. The motion that a soft tissue organ is likely to undergo during a biopsy intervention is basically the motion that takes place upon pressing the target organ with a biopsy needle. We have defined a biopsy insertion point (the point in yellow in Figure 5) on the surface of the breast phantom and simulated an external point force acting on this point towards the breast phantom centroid. The largest simulated deformation is about 16.19

mm. It is important to note that it is not necessary for the material properties of the model to match those of the CIRS phantom exactly, i.e., no possible deviations influence the accuracy of our evaluation on the specific simulated motion data, as long as the defined material properties describe some realistic (soft-tissue) material. In principle, one could use virtually any deformable FE model for simulating such non-linear motion data provided that the FE model simulates the defined (realistic) dynamics accurately. Formally, the simulated ground-truth test data are in the form of samples from the output of the motion signal function as defined in Eq. 1 with $N=3$. We simulated a ground-truth test dataset

$$\mathcal{D}_{gt} = \left\{ \left((x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-1)\Delta t, k\Delta t), f_i \right) \mid i = 1, \dots, n_*; k = 1, \dots, s \right\} \quad (13)$$

with cardinality $s n_*$. Here, n_* corresponds to the number of vertices of the employed FE model, t corresponds to the time step size used by the time integration scheme, s is the total number of discrete time points, and $x_i^{(l)}$ is the l -th component of the ground-truth test location x_i . It is assumed that the state points and the FE vertices have 1-to-1 correspondence (Kocev et al. (2014)). For evaluating our algorithm on in-sequence (i.e., synchronous) uncertain motion signal samples, a training dataset

$$\mathcal{D}_{sync} = \left\{ \left((x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-1)\Delta t, k\Delta t), y_i \right) \mid i = 1, \dots, n \right\} \quad (14)$$

is created, for a discrete time point $k, 1 \leq k \leq s$, by selecting n ($n \ll n_*$) randomly non-uniformly scattered samples out of the $S n_*$ simulated ground-truth test samples and then contaminating them with heteroscedastic Gaussian noise. In this regard, $y_i = f_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \Sigma_i)$ being the Gaussian distributed noise value at x_i . For accuracy analysis of our algorithm on out-of -sequence (i.e., asynchronous) samples, another training data set

$$\begin{aligned} \mathcal{D}_{async} &= \left\{ \left((x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-1)\Delta t, k\Delta t), y_i \right) \mid i = 1, \dots, n_1 \right\} \\ &\cup \left\{ \left((x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-2)\Delta t, k\Delta t), y_i \right) \mid i = 1, \dots, n_2 \right\} \\ &\cup \left\{ \left((x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-3)\Delta t, k\Delta t), y_i \right) \mid i = 1, \dots, n_3 \right\} \end{aligned} \quad (15)$$

is created, for a discrete time point $k, 3 \leq k \leq s$, by taking the union of three different subsets (with cardinalities n_1, n_2 , and n_3 respectively) of the simulated ground-truth test data (i.e., in total $n = n_1 + n_2 + n_3 \ll n_*$ randomly non-uniformly scattered samples) and then contaminating the samples with heteroscedastic Gaussian noise. Note that each subset

contains samples with different time intervals, i.e., the resulting training dataset, D_{async} , contains asynchronous motion signal samples.

For the purpose of this evaluation, we set

$n_* = 1,962$, $s = 6$, $\Delta t = 33 \text{ ms}$, $n = 26$, $n_1 = 9$, $n_2 = 9$, and $n_3 = 8$. We chose this setting because it is a realistic one for the medical context that drives our developments.

6.2. Training Results

Let us denote each trained GP model with $\mathcal{M}_{\mathcal{D}}^{(c,d)}$, where D is the dataset on which the model is trained, $c \in \{1, 2, 3\}$ corresponds to the ID of the employed GP prior configuration (see Table 1), and $d \in \{1, 2, 3\}$ is the index of the motion signal output dimension for which the model is trained. We optimized the (hyper-) parameters of each model by maximizing its evidence (estimated using Eq. 12, given the respective training dataset) as explained in Subsection 5.4. The training has been performed on example synchronous (see Eq. 14) and asynchronous (see Eq. 13) training datasets. The values of the optimized (hyper-) parameters of the proposed GP models $\mathcal{M}_{\mathcal{D}_{sync}}^{(3,1)}$, $\mathcal{M}_{\mathcal{D}_{sync}}^{(3,2)}$, and $\mathcal{M}_{\mathcal{D}_{sync}}^{(3,3)}$ are presented in Tables 2 and 3. The resulting probability of each model $\mathcal{M}_{\mathcal{D}}^{(c,d)}$, given the respective training dataset, is presented in Table 4. Please note that we do not train GP models having the third configuration on synchronous datasets as in that case the temporal information is irrelevant. In the case of training on synchronous data, i.e., $\mathcal{D} = \mathcal{D}_{sync}$, the resulting model probabilities show that models having the second configuration, i.e., $\mathcal{M}_{\mathcal{D}_{sync}}^{(2,d)}$, are able to explain the example training data better than models having the first configuration, i.e., . In regard to training on asynchronous data, i.e., $\mathcal{D} = \mathcal{D}_{async}$, models having the third configuration, i.e., $\mathcal{M}_{\mathcal{D}_{async}}^{(3,d)}$, accommodate the training data better than models having any of the other two prior configurations. Therefore, our proposed modeling of the mean (see Subsection 5.1) and covariance (see Subsection 5.2) functions allows for GP models that can be better trained to explain training data given in the form of randomly non-uniformly scattered motion signal samples.

6.3. Accuracy Analysis

Using Eq. 5, each trained GP model $\mathcal{M}_{\mathcal{D}}^{(c,d)}$ is conditioned on the training dataset D in order to make predictions, at the original n_* test locations $(x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, (k-1)\Delta t, k\Delta t)$, for the d -th motion signal output dimension using the c -th GP prior configuration. The resulting predictions for the motion signal function mean values at the n_* test locations are represented as an n_* -dimensional vector $\bar{f}_*$, which is evaluated in terms of accuracy against the respective n_* -dimensional vector of ground-truth function mean values $\bar{f}_{*gt}$. The difference between the vectors $\bar{f}_*$ and $\bar{f}_{*gt}$ is measured using:

1. Euclidean distance;

2. cosine distance;
3. (Pearson) correlation distance;
4. root-mean-square error (RMSE);
5. mean absolute error (MAE), and
6. Wilcoxon two-sided rank sum test (Wilcoxon (1945)), with the null hypothesis that data in $\bar{f}_*$ and in $\bar{f}_{*_{gt}}$ are samples from continuous distributions with equal medians.

By combining the predictions (obtained using the c — th GP prior configuration and training dataset D) $\bar{f}_*$ for each output dimension into an $n_* \times 3$ matrix, we obtain the estimate of the full motion signal function mean values at the given n_* test locations. Each such $n_* \times 3$ matrix of (3D) displacement predictions together with the respective n_* test locations define an $n_* \times 3$ matrix $\bar{S}_*$ with its rows being the (3D) position mean values of all displacement vectors' endpoints, which is then evaluated in terms of accuracy against the respective ground truth $\bar{S}_{*_{gt}}$ using:

1. $L_{2,1}$ norm of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)^T$, and
2. $L_{2,2}$, i.e., Frobenius, norm of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$.

In the following, we present accuracy analysis results on both synchronous and asynchronous training datasets.

6.3.1. Accuracy Analysis Results

Synchronous Data.: In performing regression from space and time to motion signal function values using a synchronous training dataset, $\mathcal{D} = \mathcal{D}_{sync}$, all predictions $\bar{f}_*$ by models $\mathcal{M}_{\mathcal{D}_{sync}}^{(2,d)}$ are closer, in terms of Euclidean distance, cosine distance, and RMSE, to the ground truth $\bar{f}_{*_{gt}}$ than those by models $\mathcal{M}_{\mathcal{D}_{sync}}^{(1,d)}$ (see Tables 5, 6, and 8). In terms of (Pearson) correlation distance and MAE, the predictions for the second output dimension, $d=2$, are closer to the ground truth when using the first GP prior configuration, $c=1$. For the other two output dimensions, the second GP prior configuration, $c=2$, allows for better predictions (see Tables 7 and 9). $\mathcal{M}_{\mathcal{D}_{sync}}^{(1,2)}$ is better, in terms of MAE but not in terms of RMSE, than $\mathcal{M}_{\mathcal{D}_{sync}}^{(2,2)}$ mainly because in that case the first configuration gives higher-value outliers in the component-wise absolute differences to which RMSE gives higher weights (Chai and Draxler (2014)). However, RMSE may be a more appropriate metric for deciding which algorithm is better suited for safety-critical applications (Knight (2002)), e.g., navigated surgery (Mezger et al. (2013)) where large errors are to be avoided. An additional argument supporting that RMSE is more appropriate to represent model performance than

MAE in this case is that the distribution of the error $\bar{f}_* - \bar{f}_{*_{gt}}$ yielded by both configurations is Gaussian (Chai and Draxler (2014)). To confirm that the error is Gaussian-distributed, we used a t-test with the null hypothesis being that the data in $\bar{f}_* - \bar{f}_{*_{gt}}$ come from a normal distribution with unknown variance and mean value equal to the error sample set mean. Moreover, we visually analyzed the histogram of $\bar{f}_* - \bar{f}_{*_{gt}}$ yielded by both configurations (see Figures 6 and 7). As the means of the error sample sets yielded by both configurations are not exactly zero, the error sample sets are slightly biased (Chai and Draxler (2014)). Therefore, we provide, in addition, the respective standard error (SE) information (Chai and Draxler (2014)) (see Table 10). Note that when the error distribution and sample set are unbiased, the SE is equivalent to the RMSE (Chai and Draxler (2014)). While the results from the Wilcoxon two-sided rank sum tests show that there is no strong evidence supporting that there is a significant (at the 5% significance level) difference between the predictions yielded by $\mathcal{M}_{\mathcal{D}_{sync}}^{(2,d)}$ and the respective ground truth, however, there is strong evidence supporting that there is significant difference (except for $d=1$) between the predictions made using GP prior configuration 1 (i.e., $c=1$) and the respective ground truth. Hence, there is stronger evidence that, in general, the second GP prior configuration allows for better estimation of the true motion function median values (see Table 11). Furthermore, both the $L_{2,1}$ norm of $(\bar{S}_* - \bar{S}_{*_{gt}})^T$ and $L_{2,2}$ (i.e., Frobenius) norm of $(\bar{S}_* - \bar{S}_{*_{gt}})$ are smaller when using the second GP prior configuration (see Tables 12 and 13), i.e., the overall motion estimation error is larger when using the first GP prior configuration.

Asynchronous Data.: When mapping from spatiotemporal locations to motion signal function values using an asynchronous training dataset (i.e., $D = D_{async}$), all predictions $\bar{f}_*$ by models $\mathcal{M}_{\mathcal{D}_{async}}^{(3,1-2)}$ are closer, in terms of Euclidean distance, (Pearson) correlation distance, RMSE, and MAE, to the ground truth $\bar{f}_{*_{gt}}$ than those by models $\mathcal{M}_{\mathcal{D}_{async}}^{(1,1-2)}$ and $\mathcal{M}_{\mathcal{D}_{async}}^{(2,1-2)}$ (see Tables 5, 7, 8, and 9). Further-more, the predictions by model $\mathcal{M}_{\mathcal{D}_{async}}^{(3,3)}$ are better, in terms of cosine and (Pearson) correlation distances, than those by models $\mathcal{M}_{\mathcal{D}_{async}}^{(1,3)}$ and $\mathcal{M}_{\mathcal{D}_{async}}^{(2,3)}$ (see Tables 6 and 7). In all cases where the third GP prior configuration fails to produce better predictions than the other two configurations, the first GP prior configuration is the one that outperforms (except in terms of cosine distance for the second output dimension) the other configurations (see Tables 5, 6, 8, and 9). The results from the Wilcoxon two-sided rank sum tests show that there is strong evidence supporting that there is a significant (at the 5% significance level) difference between the predictions yielded by any $\mathcal{M}_{\mathcal{D}_{async}}^{(c,d)}$ and the respective ground truth (see Table 11). However, the Wilcoxon two-sided rank sum test does not provide strong evidence that any of the three different GP prior configuration allows for better estimation of the true motion function median values when compared to the others. On the other hand, the $L_{2,1}$ norm of $(\bar{S}_* - \bar{S}_{*_{gt}})^T$ and $L_{2,2}$ (i.e., Frobenius) norm of $(\bar{S}_* - \bar{S}_{*_{gt}})$ are clearly in favor of using the

third GP prior configuration. They are smallest when using the third GP prior configuration (see Tables 12 and 13), i.e., the overall motion estimation error is smallest when using our proposed GP prior configuration. In this regard, note that the overall motion estimation error is decreased by more than 50% in case of using our proposed mean and covariance functions (i.e., the third GP prior configuration) when doing regression on asynchronous training data. This major improvement is achieved mainly by using our proposed covariance function. For a visual depiction of the overall median errors yielded by different GP prior configurations trained on the same asynchronous dataset, we boxplot the L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by each GP prior configuration. In this regard, the boxplot notches in Figure 8 offer evidence of a statistically significant difference (at the 5% significance level) between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations. To formally confirm that there are statistically significant differences between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations, we employ a Wilcoxon two-sided rank sum test (Wilcoxon (1945)). The null hypothesis is that the values of the L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by one GP prior configuration trained on a given dataset and those yielded by another configuration trained on the same dataset are samples from continuous distributions with equal medians. Provided that there are statistically significant differences between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations (see Table 14) and the fact that the median L_2 norm of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the third GP prior configuration is the smallest, we conclude that there is strong evidence that our proposed modeling significantly decreases the overall motion estimation error when performing regression using an asynchronous training dataset. In some applications, the measurements may be restricted to some region, e.g., on the surface of the soft-tissue organ. To test whether our proposed method also improves the accuracy even under such constraints, we performed regression on an additional training dataset $D_{asyncSurf}$ composed of randomly non-uniformly spatiotemporally scattered uncertain surface motion measurements. In Figure 9, similar to Figure 8, one can observe the overall median errors yielded by the different GP prior configurations trained on $D_{asyncSurf}$. In this regard, the boxplot notches in Figure 9 also offer evidence of a statistically significant difference (at the 5% significance level) between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations. To formally confirm that there are statistically significant differences between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations trained on $D_{asyncSurf}$, we again employ a Wilcoxon two-sided rank sum test (Wilcoxon (1945)). The null hypothesis is again that the values of the L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by one GP prior configuration trained on $D_{asyncSurf}$ and those yielded by another configuration trained on the same dataset

are samples from continuous distributions with equal medians. The results from the Wilcoxon two-sided rank sum test are the same as those reported in Table 14 for the case $D = D_{asyn}$, i.e., it is formally confirmed that there are statistically significant differences between the median L_2 norms of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the three GP prior configurations. Provided this and the fact that the median L_2 norm of the rows of $\left(\bar{S}_* - \bar{S}_{*_{gt}}\right)$ yielded by the third GP prior configuration is the smallest, we conclude that there is strong evidence that our proposed modeling significantly decreases the overall motion estimation error also when performing regression on randomly non-uniformly spatiotemporally scattered uncertain surface motion measurements.

7. Conclusion and Future Work

We presented a novel algorithm for uncertainty-aware interpolation of randomly non-uniformly spatiotemporally scattered motion signal samples. By employing GP regression, we were able to perform the interpolation in an optimal unbiased fashion. The use of a composite (as the sum of a constant and linear function) prior mean function enabled the learning of global and local drifts, present in the latent process mean function, from randomly non-uniformly spatiotemporally scattered samples. By means of a squared exponential covariance function with ADM, we were able to model the nearness or similarity between pairs of random motion function values over a spatiotemporal domain. Through estimating the full a posteriori motion field distribution, we were able to quantify the uncertainty in the resulting MAP estimate of the soft tissue motion at any location in the spatiotemporal domain.

The evaluation of our devised interpolation algorithm on simulated randomly non-uniformly spatiotemporally scattered uncertain motion signal samples revealed that our proposed GP model is able to, at the same time, learn more and yield statistically significantly better predictions than the state-of-the-art GP models that employ a zero-mean function and do not make use of ADM. We identified strong evidence supporting our contention that our proposed modeling significantly decreases the overall motion estimation error when performing regression both in the case of using synchronous and asynchronous motion signal samples as training data.

In future work, we plan to identify appropriate formalisms and, if needed, approximation approaches to optimize the conditioning (see Eq. 5) of the proposed GP model on a given training dataset, especially required when the training dataset is large. We would like to apply our proposed model for registering real pre-interventional/pre-operative navigation data on deformable soft-tissue organs during a real intervention or surgery. Furthermore, we intend to deploy our proposed GP model for modeling organ(s) deformation in the context of radiotherapy.

Acknowledgments

This work was supported by the Fraunhofer Internal Programs under Grant No. MAVO 823 287. It was also supported by the DFG Creative Unit “Intra-Operative Information: What Surgeons Need, When They Need It”, and the NIH grants P41 EB015902, U24 CA180918, and P41 EB015898.

References

- Barber D, 2012 Bayesian reasoning and machine learning. Cambridge University Press.
- Baumhauer M, Feuerstein M, Meinzer HP, Rassweiler J, 2008 Navigation in endoscopic soft tissue surgery: perspectives and limitations. *Journal of endourology* 22, 751–766. [PubMed: 18366319]
- Bersvendsen J, Toews M, Danudibroto A, Wells WM, Urheim S, Estepar RSJ, Samset E, 2016 Robust spatio-temporal registration of 4d cardiac ultrasound sequences, in: *SPIE Medical Imaging*, International Society for Optics and Photonics. pp. 97900F–97900F.
- Bishop CM, 2006 Pattern recognition. Machine Learning 128.
- Cash DM, Miga MI, Sinha TK, Galloway RL, Chapman WC, 2005 Compensating for intraoperative soft-tissue deformations using incomplete surface data and finite elements. *Medical Imaging, IEEE Transactions on* 24, 1479–1491.
- Chai T, Draxler RR, 2014 Root mean square error (rmse) or mean absolute error (mae)?-arguments against avoiding rmse in the literature. *Geoscientific Model Development* 7, 1247–1250.
- Chan AB, 2013 Multivariate generalized gaussian process models. *arXiv preprint arXiv:1311.0360*.
- Gelman A, Carlin JB, Stern HS, Rubin DB, 2014 Bayesian data analysis. volume 2 Chapman & Hall/CRC Boca Raton, FL, USA.
- Georgii J, Westermann R, 2005 A multigrid framework for real-time simulation of deformable volumes, in: *Workshop On Virtual Reality Interaction and Physical Simulation*.
- Georgii J, Westermann R, 2008 Corotated finite elements made fast and stable, in: *Proceedings of the 5th Workshop On Virtual Reality Interaction and Physical Simulation*, pp. 11–19.
- Gerig T, Shahim K, Reyes M, Vetter T, Liithi M, 2014 Spatially varying registration using gaussian processes, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer pp. 413–420.
- Glocker B, Komodakis N, Navab N, Tziritas G, Paragios N, 2009 Dense registration with deformation priors, in: *International Conference on Information Processing in Medical Imaging*, Springer pp. 540–551.
- Goldberg PW, Williams CK, Bishop CM, 1997 Regression with input- dependent noise: A gaussian process treatment. *Advances in neural information processing systems* 10, 493–499.
- Hein M, Bousquet O, 2004 Kernels, associated structures and generalizations. Max-Planck-Institut fuer biologische Kybernetik, Technical Report.
- Irani M, Anandan P, 2000 Factorization with uncertainty, in: *Computer Vision-ECCV 2000*. Springer, pp. 539–553.
- Kanazawa Y, Kanatani K.i., 2001 Do we really have to consider covariance matrices for image features?, in: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, IEEE pp. 301–306.
- Khaleghi B, Khamis A, Karray FO, Razavi SN, 2013 Multisensor data fusion: A review of the state-of-the-art. *Information Fusion* 14, 28–44.
- Knight JC, 2002 Safety critical systems: challenges and directions, in: *Software Engineering, 2002. ICSE 2002. Proceedings of the 24rd International Conference on*, IEEE pp. 547–550.
- Kocev B, Georgii J, Linsen L, Hahn HK, 2014 Information Fusion for Real-time Motion Estimation in Image-guided Breast Biopsy Navigation, in: *Bender J, Duriez C, Jaillet F, Zachmann G (Eds.), Workshop on Virtual Reality Interaction and Physical Simulation, The Eurographics Association*. doi:10.2312/vriphys.20141227.
- Kocev B, Ritter F, Linsen L, 2013 Projector-based surgeon-computer interaction on deformable surfaces. *International Journal of Computer Assisted Radiology and Surgery*, 1–12URL: 10.1007/s11548-013-0928-1, doi:10.1007/s11548-013-0928-1.

- Kuss M, 2006 Gaussian process models for robust regression, classification, and reinforcement learning. Ph.D. thesis. TU Darmstadt.
- Lê M, Unkelbach J, Ayache N, Delingette H, 2015 Gpssi: Gaussian process for sampling segmentations of images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer pp. 38–46.
- Ledesma-Carbayo MJ, Kybic J, Desco M, Santos A, Suhling M, Hunziker P, Unser M, 2005 Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *IEEE transactions on medical imaging* 24, 1113–1126. [PubMed: 16156350]
- Leedan Y, Meer P, 2000 Heteroscedastic regression in computer vision: Problems with bilinear constraint. *International Journal of Computer Vision* 37, 127–150.
- Lim JH, Ong SH, Xiong W, 2015 Biomedical Image Understanding: Methods and Applications. John Wiley & Sons.
- Lüthi M, Jud C, Vetter T, 2011 Using landmarks as a deformation prior for hybrid image registration, in: *Pattern Recognition*. Springer, pp. 196–205.
- Lüthi M, Jud C, Vetter T, 2013 A unified approach to shape model fitting and non-rigid registration, in: *International workshop on machine learning in medical imaging*, Springer pp. 66–73.
- Meinzer HP, Maier-Hein L, Wegner I, Baumhauer M, Wolf I, 2008 Computer-assisted soft tissue interventions, in: *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, IEEE pp. 1391–1394.
- Mezger U, Jendrewski C, Bartels M, 2013 Navigation in surgery. *Langenbeck's Archives of Surgery* 398, 501–514.
- Micchelli CA, Pontil M, 2005 On learning vector-valued functions. *Neural computation* 17, 177–204. [PubMed: 15563752]
- O'Hagan A, Kingman JFC, 1978 Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)* 40, 1–42. URL: <http://www.jstor.org/stable/2984861>.
- Perperidis D, Mohiaddin RH, Rueckert D, 2005 Spatio-temporal free-form registration of cardiac mr image sequences. *Medical image analysis* 9, 441–456. [PubMed: 16029955]
- Polak E, Ribiere G, 1969 Note sur la convergence de methodes de directions conjuguées. *Revue française d'informatique et de recherche operationnelle, serie rouge* 3, 35–43.
- Rasmussen C, Williams C, 2006 *Gaussian Processes for Machine Learning Adaptive Computation and Machine Learning*, MIT Press, Cambridge, MA, USA.
- Rasmussen CE, 2006 Gaussian processes for machine learning .
- Rasmussen CE, Nickisch H, 2010 Gaussian processes for machine learning (gpml) toolbox. *The Journal of Machine Learning Research* 11, 3011–3015.
- Risholm P, Pieper S, Samset E, Wells III WM, 2010 Summarizing and visualizing uncertainty in non-rigid registration, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer pp. 554–561.
- Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn M, 2001 Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on medical imaging* 20, 526–534. [PubMed: 11437112]
- Schlegel S, Korn N, Scheuermann G, 2012 On the interpolation of data with normally distributed uncertainty for visualization. *Visualization and Computer Graphics, IEEE Transactions on* 18, 2305–2314.
- Shi W, Jantsch M, Aljabar P, Pizarro L, Bai W, Wang H, O'Regan D, Zhuang X, Rueckert D, 2013 Temporal sparse free-form deformations. *Medical image analysis* 17, 779–789. [PubMed: 23743085]
- Sotiras A, Davatzikos C, Paragios N, 2013 Deformable medical image registration: A survey. *IEEE transactions on medical imaging* 32, 1153–1190. [PubMed: 23739795]
- Styzt MR, Parrott RW, 1993 Using kriging for 3d medical imaging. *Computerized Medical Imaging and Graphics* 17, 421–442. [PubMed: 8287354]
- Thévenaz P, Blu T, Unser M, 2000a Image interpolation and resampling. *Handbook of medical imaging, processing and analysis* , 393–420.

- Thévenaz P, Blu T, Unser M, 2000b Interpolation revisited [medical images application]. IEEE Transactions on medical imaging 19, 739–758. [PubMed: 11055789]
- Vandemeulebroucke J, Rit S, Kybic J, Clarysse P, Sarrut D, 2011 Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Medical physics 38, 166–178. [PubMed: 21361185]
- Wachinger C, Golland P, Reuter M, Wells W, 2014 Gaussian Process Interpolation for Uncertainty Estimation in Image Registration. Springer International Publishing, Cham pp. 267–274. doi: 10.1007/978-3-319-10404-1_34.
- Wang Y, Georgescu B, Chen T, Wu W, Wang P, Lu X, Ionasec R, Zheng Y, Comaniciu D, 2013 Learning-based detection and tracking in medical imaging: a probabilistic approach, in: Deformation Models. Springer, pp. 209–235.
- Wassermann D, Toews M, Niethammer M, Wells III W, 2014 Probabilistic diffeomorphic registration: Representing uncertainty, in: International Workshop on Biomedical Image Registration, Springer pp. 72–82.
- Wilcoxon F, 1945 Individual comparisons by ranking methods. Biometrics bulletin 1, 80–83.
- Worz S, Rohr K, 2008 Physics-based elastic registration using non-radial basis functions and including landmark localization uncertainties. Computer Vision and Image Understanding 111, 263–274.
- Zhang H, Banovac F, Lin R, Glossop N, Wood BJ, Lindisch D, Levy E, Cleary K, 2006 Electromagnetic tracking for abdominal interventions in computer aided surgery. Computer Aided Surgery 11, 127–136. [PubMed: 16829506]
- Zhao Q, Pizer S, Alterovitz R, Niethammer M, Rosenman J, 2017 Orthotropic thin shell elasticity estimation for surface registration, in: International Conference on Information Processing in Medical Imaging, Springer pp. 493–504.
- Zhou XS, Comaniciu D, Xie B, Cruceanu R, Gupta A, 2004 A unified framework for uncertainty propagation in automatic shape tracking, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, IEEE pp. I–872.
- Zhou XS, Gupta A, Comaniciu D, 2005 An information fusion framework for robust shape tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, 115–129.

Figure 1:

Predicting motion from asynchronous observations. M_i , M_j , M_k , and M_l are example measurement points, which are updated asynchronously at different points in time t_0 , t_1 , and t_2 , while P_i is an example state point, which shall be predicted using uncertainty-aware asynchronous scattered motion interpolation.

Figure 2: Log marginal likelihood decomposition into its constituents: data-fit and complexity penalty, as a function of the characteristic length-scale l_1 .

Figure 3:
Log marginal likelihood decomposition into its constituents: data-fit and complexity penalty, as a function of the characteristic length-scale l_2

Figure 4:

Left: Breast phantom with 4 markers (one on the back side). Right: FE model composed of tetrahedral elements which are extracted from the MRI scan data of the breast phantom. The points in red are fixed, i.e., the FE nonlinear motion prediction model considers these vertices as not moving. (Image courtesy of Kocev et al. (2014).)

Figure 5:
Deformed breast phantom FE model with biopsy insertion point visualized in yellow. The points in red are fixed, while the point in yellow depicts the location where the external point force is applied.

Figure 6:
Histogram of $\bar{f}_* - \bar{f}_{*gt}$ yielded by $\mathcal{M}_{\mathcal{D}_{sync}}^{(1,2)}$.

Figure 7:
Histogram of $\bar{f}_* - \bar{f}_{*gt}$ yielded by $\mathcal{M}_{\mathcal{D}_{sync}}^{(2,2)}$.

Figure 8:

Boxplots of the L_2 norms of the rows of $(\tilde{S}_* - \tilde{S}_{*gt})$ yielded by different GP prior configurations trained on D_{async} : (i) using models $\mathcal{M}_{\mathcal{D}_{async}}^{(1,d)}$, (ii) using models $\mathcal{M}_{\mathcal{D}_{async}}^{(2,d)}$, (iii) using models $\mathcal{M}_{\mathcal{D}_{async}}^{(3,d)}$.

Figure 9:

Boxplots of the L_2 norms of the rows of $(\tilde{S}_* - \tilde{S}_{*gt})$ yielded by different GP prior

configurations trained on $D_{asyncSurf}$: (i) using models $\mathcal{M}_{\mathcal{D}_{asyncSurf}}^{(1,d)}$, (ii) using models

$\mathcal{M}_{\mathcal{D}_{asyncSurf}}^{(2,d)}$, (iii) using models $\mathcal{M}_{\mathcal{D}_{asyncSurf}}^{(3,d)}$.

Table 1:

GP priors configurations.

ID	Mean Func.	Cov. Func.	Lik. Func.
1	zero	SE w/ IDM	proposed
2	proposed	SE w/ IDM	proposed
3	proposed	proposed	proposed

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:
(Hyper-) parameters of the mean function of the proposed models $\mathcal{M}_{\mathcal{D}_{async}}^{(3,d)}$

	β	c
$d=1$	[0.000105, 0.001851, -0.001966, 0.003670, -0.008100]	0.999858
$d=2$	[-0.001354, 0.002312, 0.007315, -0.013983, 0.007166]	0.999935
$d=3$	[-0.001398, -0.007534, -0.000155, 0.000535, -0.004999]	0.999874

Table 3:

(Hyper-) parameters of the covariance and likelihood functions of the proposed models $\mathcal{M}_{\mathcal{D}}^{(3,d)}_{async}$.

	l_1	l_2	l_3	l_4	l_5	σ_f	σ_n
$d=1$	30.017140	30.021074	30.008921	600.005046	600.000000	0.005032	0.227006
$d=2$	23.156668	23.155694	23.119362	300.062788	300.000000	0.015149	0.238498
$d=3$	15.033298	15.028769	15.030861	600.007680	600.000000	0.010139	0.175856

Table 4:

Negative log marginal likelihoods of models $\mathcal{M}_{\mathcal{D}}^{(c,d)}$, with D being the training dataset (either synchronous or asynchronous), c the ID of the GP prior configuration (see Table 1), and d the index of the motion signal output dimension.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	-27.9233	-31.7146	-8.4969	-1.1611	-9.6756
$d = 2$	-9.3417	-13.1894	14.8096	13.9991	-8.6495
$d = 3$	-10.9099	-12.9253	-2.6988	-3.7344	-9.1918

Table 5:
Euclidean distances between $\bar{f}_*$ and $\bar{f}_{*gt}$.

	$D = D_{Sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	3.48379	3.28559	3.77276	6.75917	2.83698
$d = 2$	15.61828	15.52604	17.40310	14.72825	5.36545
$d = 3$	11.27818	11.04141	4.66890	8.25508	5.92519

Table 6:Cosine distances between $\bar{f}_*$ and $\bar{f}_{*gt}$.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	0.19607	0.17226	0.10324	0.15673	0.89327
$d = 2$	0.12244	0.12214	0.26239	0.13359	0.32024
$d = 3$	0.61894	0.57702	0.03873	0.06904	0.00839

Table 7:
(Pearson) correlation distances between $\bar{f}_*$ and $\bar{f}_{*_{gt}}$.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	0.63771	0.51480	1.05600	1.30947	0.53873
$d = 2$	0.58169	0.60213	0.97160	0.15555	0.04888
$d = 3$	0.90247	0.81330	0.07474	0.15025	0.01750

Table 8:

Root-mean-square errors between $\bar{f}_*$ and $\bar{f}_{*_{gt}}$.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	0.07865	0.07418	0.08517	0.15260	0.06405
$d = 2$	0.35260	0.35052	0.39290	0.33251	0.12113
$d = 3$	0.25462	0.24927	0.10541	0.18637	0.13377

Table 9:

Mean absolute errors between $\bar{f}_*$ and $\bar{f}_{*gt}$.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	0.03910	0.02879	0.08428	0.14592	0.05122
$d = 2$	0.07216	0.08021	0.38632	0.32558	0.10092
$d = 3$	0.05766	0.05222	0.08918	0.15518	0.10830

Table 10:Standard error between $\bar{f}_*$ and $\bar{f}_{*gt}$.

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	0.07863	0.07407	0.012308	0.044653	0.055790
$d = 2$	0.35084	0.34961	0.071571	0.067533	0.099143
$d = 3$	0.25241	0.24850	0.057916	0.131253	0.087739

Table 11:

Wilcoxon two-sided rank sum test null hypothesis (that data in $\bar{f}_*$ and in $\bar{f}_{*_{gt}}$ are samples from continuous distributions with equal medians) results ((p, h): “h=1” indicates a rejection of the null hypothesis, while “h=0” indicates a failure to reject the null hypothesis, at the 5% significance level, based on the estimated p-value p).

	$D = D_{sync}$		$D = D_{async}$		
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
$d = 1$	(0.110516, 0)	(0.360029, 0)	(0, 1)	(0, 1)	(0, 1)
$d = 2$	(0.000556, 1)	(0.077498, 0)	(0, 1)	(0, 1)	(0, 1)
$d = 3$	(0, 1)	(0.326926, 0)	(0, 1)	(0, 1)	(0, 1)

Table 12:

$L_{2,1}$ norm of $(\bar{S}_* - \bar{S}_{*_{gt}})^T$

$D = D_{sync}$		$D = D_{async}$		
$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
228.0773	219.4586	803.9132	783.8536	343.4022

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 13:

L_{2,2} (i.e., Frobenius) norm of $(\bar{S}_* - \bar{S}_{*_{gt}})$

$D = D_{sync}$		$D = D_{async}$		
$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 3$
19.5771	19.3330	18.4092	18.1866	8.4820

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 14:

Wilcoxon two-sided rank sum test null hypothesis (that the values of the L_2 norms of the rows of $(\bar{S}_* - \bar{S}_{*_{gt}})$ yielded by one GP prior configuration trained on a given dataset and those yielded by another configuration trained on the same dataset are samples from continuous distributions with equal medians) results ((p, h): “h=1” indicates a rejection of the null hypothesis, while “h=0” indicates a failure to reject the null hypothesis, at the 5% significance level, based on the estimated p-value p).

$D = D_{sync}$	$D = D_{async}$		
$(c = 1, c = 2)$	$(c = 1, c = 2)$	$(c = 1, c = 3)$	$(c = 2, c = 3)$
(0.006665, 1)	(0, 1)	(0, 1)	(0, 1)