Gaussian Process Interpolation for Uncertainty Estimation in Image Registration

Christian Wachinger[1,2], Polina Golland[1],
Martin Reuter[1,2], and William Wells[1,3]

[1] Computer Science and Artificial Intelligence Lab, MIT
[2] Massachusetts General Hospital, Harvard Medical School
[3] Brigham and Women's Hospital, Harvard Medical School

Abstract. Intensity-based image registration requires resampling images on a common grid to evaluate the similarity function. The uncertainty of interpolation varies across the image, depending on the location of resampled points relative to the base grid. We propose to perform Bayesian inference with Gaussian processes, where the covariance matrix of the Gaussian process posterior distribution estimates the uncertainty in interpolation. The Gaussian process replaces a single image with a distribution over images that we integrate into a generative model for registration. Marginalization over resampled images leads to a new similarity measure that includes the uncertainty of the interpolation. We demonstrate that our approach increases the registration accuracy and propose an efficient approximation scheme that enables seamless integration with existing registration methods.

1 Introduction

Registration is a fundamental tool in medical imaging for image alignment. Intensity-based registration commonly finds the transformation between images by an iterative procedure that resamples images on a common grid to evaluate their similarity. An inherent problem is the variation of the interpolation uncertainty across the image. Fig. 1 illustrates two images and an overlay of the corresponding grids. Intensity values on the moving grid (blue) are used to interpolate values on the fixed grid (red) to enable the comparison of both images. We point out two locations on the fixed grid that have very different distances to neighboring points on the moving grid. This difference causes variations in the interpolation uncertainty. Both locations contribute equally to the calculation of the similarity measure, although the interpolation from observations that are far away may not be very trustworthy.

To address this problem, we formulate the interpolation as Bayesian regression. The intensity values on the transformed grid serve as observations and the prediction yields samples on the fixed grid. We employ a Gaussian process (GP) prior over images and assume Gaussian noise on the observations. The inferred predictive distribution is Gaussian with mean and covariance functions serving as an interpolator and a confidence estimate. Depending on the design of the

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 267–274, 2014.

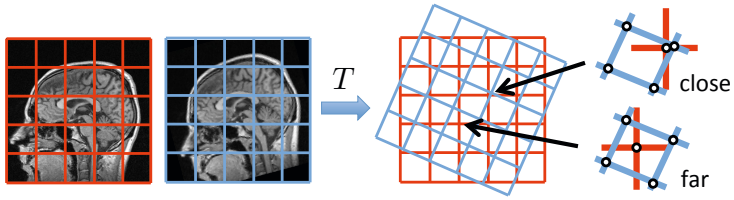


Fig. 1. Fixed (red) and moving (blue) images and the overlay of both grids after transformation (middle). The interpolation uncertainty varies across the resampled image due to different distances to neighboring points on the moving grid. Arrows point to two exemplary locations on the fixed grid where neighbors from the moving grid are close and far, respectively (right).

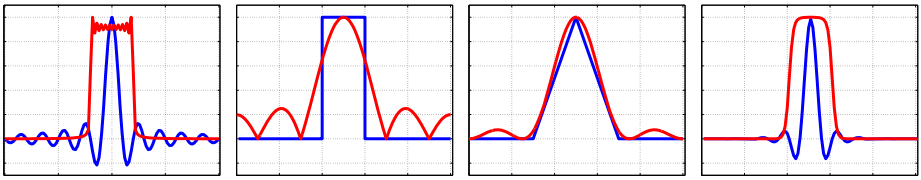
covariance matrix of the GP prior and the magnitude of the presumed noise in the images, we can account for smoothing and noise reduction in the prediction. This makes Gaussian processes a versatile framework for modeling image processing steps in registration.

The application of Gaussian processes introduces a new paradigm for the use of image interpolation in registration. Instead of only comparing the resampled intensity values, the similarity measure now takes into account the quality of the interpolation, which can vary dramatically across the image. To enable this change, we present a generative model for image registration with Gaussian processes. The inferred similarity measure emphasizes locations where samples are close to the original grid and deprecates locations that are equidistant from grid points. This is especially beneficial for anisotropically sampled data, frequently acquired in the clinical practice.

Related Work. The most common methods for interpolation are nearest neighbor, linear, cubic, and spline interpolation. The application of cubic B-splines for interpolation was proposed in [5]. Several excellent surveys of image interpolation exist [7,14]. Image interpolation in the context of registration is discussed in [4]. Further studies have been conducted to investigate the generation of interpolation artifacts and their influence on image registration, see for instance [1] and references therein. Gaussian processes have been applied in several fields of machine learning [11], *e.g.*, image denoising [8], interpolation [13] and segmentation [15]. Gaussian processes were also used to model flow fields [6] and deformation fields in hybrid registration [9]. Gaussian processes have not yet been used for image resampling in registration.

2 Method

Given two images I and J defined on discrete grids Ω_I and Ω_J, we calculate the transformation T that aligns the two images. We transform the grid Ω_J of the moving image J, yielding the transformed grid $T(\Omega_J) = \{T(\mathbf{x}), \mathbf{x} \in \Omega_J\}$.



(a) Truncated sinc (b) Nearest (c) Linear (d) GP

Fig. 2. Comparison of interpolation functions in spatial (blue) and frequency (red) domains. The optimal frequency response would correspond to a box function.

Except for axis-aligned transformations, we have to resample the transformed image from the grid $T(\Omega_J)$ to the grid of the fixed image Ω_I to compare the two images. For the resampling, a continuous version of the discrete input image is constructed with interpolation [10]. Fig 2 characterizes common image interpolation methods by showing their responses in spatial and frequency domains.

2.1 Image Interpolation with Gaussian Process Regression

In this section, we formulate image interpolation as Gaussian process regression to obtain the interpolator and uncertainty estimates. A Gaussian process is a stochastic process consisting of an infinite collection of random variables, where any finite subset has a multivariate Gaussian distribution [11]. A Gaussian process $\mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, is entirely characterized by the mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$ functions. The mean and covariance functions specify a distribution over functions, corresponding to a distribution over images in our case. We make the common assumption of a zero mean function [11].

Given moving image J on the transformed grid $X = T(\Omega_J)$, we predict the resampled image J^* on the fixed image grid $X^* = \Omega_I$. We employ a Gaussian process prior on the resampled image, $J^* \sim \mathcal{GP}(\mathbf{0}, k)$. Considering Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_J)$, the observations are distributed according to $p(J|J^*, X, X^*) = \mathcal{N}(\mathbf{0}|k(X, X) + \sigma_J^2\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. Under these assumptions, the posterior distribution for predicting the transformed image is

$$p(J^* \mid J; X^*, X) = \mathcal{N}(\boldsymbol{\mu}_J, \Sigma_J), \tag{1}$$

with mean and covariance

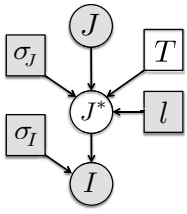
$$\boldsymbol{\mu}_J = k(X^*, X) \cdot [k(X, X) + \sigma_J^2\mathbf{I}]^{-1} \cdot J, \tag{2}$$

$$\Sigma_J = k(X^*, X^*) - k(X^*, X) \cdot [k(X, X) + \sigma_J^2\mathbf{I}]^{-1} \cdot k(X, X^*). \tag{3}$$

The covariance or kernel function k characterizes the properties of images. It captures the relation between the random variables, which correspond to the voxels in the image. We work with the squared exponential covariance function with length-scale l, $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2/(2 \cdot l^2)\right)$. The equivalent kernel characterizes the behavior of GP interpolation and is shown in Fig. 2 for the

squared exponential function. Theoretical connections to sinc interpolation exist for specific settings of the kernel [12]. The squared exponential kernel corresponds to a Bayesian linear regression model with an infinite number of Gaussian-shaped basis functions [11].

2.2 Generative Model for Gaussian Process Registration



We derive the registration method with uncertainty estimates by integrating the Gaussian process in a new generative model for registration (see graphical model on the right). We treat input images I and J as observed random variables affected by image noise $\varepsilon_I \sim \mathcal{N}(0, \sigma_I^2)$ and $\varepsilon_J \sim \mathcal{N}(0, \sigma_J^2)$, respectively. The resampled image J^* is a latent random variable. The amount of smoothing in the image J^* is controlled by the length-scale l of the kernel. Following the graphical model, the joint distribution of images I, J, J^* factorizes

$$p(I, J, J^*; T, \sigma_J, \sigma_I, l) = p(J^*|J; T, \sigma_J, l) \cdot p(I|J^*; \sigma_I). \tag{4}$$

The probability $p(J^*|J; T, \sigma_J, l)$ is the predictive distribution of the Gaussian process. From the previous section on Gaussian process interpolation we have $p(J^*|J; T, \sigma_J, l) \sim \mathcal{N}(\boldsymbol{\mu}_J, \Sigma_J)$. The likelihood $p(I|J^*; \sigma_I)$ accounts for noise in the fixed image I with respect to the prediction J^*. Under the assumption of i.i.d. Gaussian noise, this leads to the multivariate Gaussian distribution $p(I|J^*; \sigma_I) \sim \mathcal{N}(J^*, \sigma_I^2\mathbf{I})$. For calculating the optimal transformation $\hat{T}$, we perform maximum likelihood estimation on the joint distribution of images I and J

$$\hat{T} = \arg\max_T p(I, J; T, \sigma_J, \sigma_I, l). \tag{5}$$

For Bayesian inference, we marginalize over the latent random variable J^*

$$p(I, J; T, \sigma_I, \sigma_J, l) = \int p(I, J, J^*; T, \sigma_I, \sigma_J, l)\ \mathrm{d}J^* \tag{6}$$

$$= \int p(J^*|J; T, \sigma_J, l) \cdot p(I|J^*; \sigma_I)\ \mathrm{d}J^* \tag{7}$$

$$= \int \mathcal{N}(J^*; \boldsymbol{\mu}_J, \Sigma_J) \cdot \mathcal{N}(I; J^*, \sigma_I^2\mathbf{I})\ \mathrm{d}J^* \tag{8}$$

$$= \mathcal{N}(I; \boldsymbol{\mu}_J, \Sigma_J + \sigma_I^2\mathbf{I}), \tag{9}$$

where we applied the factorization from the graphical model and product properties of multivariate Gaussian distributions [11]. The log-likelihood function is

$$\log p(I, J; T, \sigma_I, \sigma_J, l) = \log\left((2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\right) - \frac{1}{2}(I - \boldsymbol{\mu}_J)^\top \Sigma^{-1}(I - \boldsymbol{\mu}_J), \tag{10}$$

with $\Sigma = \Sigma_J + \sigma_I^2\mathbf{I}$. This is the new similarity measure that we use for registration, where the covariance matrix Σ contains the uncertainty estimates. The presented approach models forward mapping in registration, where we obtain backward mapping by setting $X = \Omega_J$ and $X^* = T^{-1}(\Omega_I)$.

2.3 Practical Considerations

The computational cost of $\mathcal{O}(|\Omega_J|^3)$ for the matrix inversion $[k(X,X)+\sigma_J^2\mathbf{I}]^{-1}$ is challenging for large images. In order to reduce the computational cost, we split the volume into blocks. We perform the prediction for each block separately, where we identify the spatially closest observations. This comes at almost no additional cost, because the distances need to be calculated for constructing the kernel. Visual inspection has not shown boundary effects. With this approach, we do not construct the full covariance matrix Σ anymore, so that we cannot apply the similarity measure in Eq. (10). We consider only the diagonal entries of the covariance matrix Σ_{xx} and neglect the first term in Eq. (10), yielding

$$\log p(I, J; T, \sigma_I, \sigma_J, l) \approx -\sum_{x \in \Omega_I} \frac{(I(x) - \boldsymbol{\mu}_J(x))^2}{2 \cdot \Sigma_{xx}}. \tag{11}$$

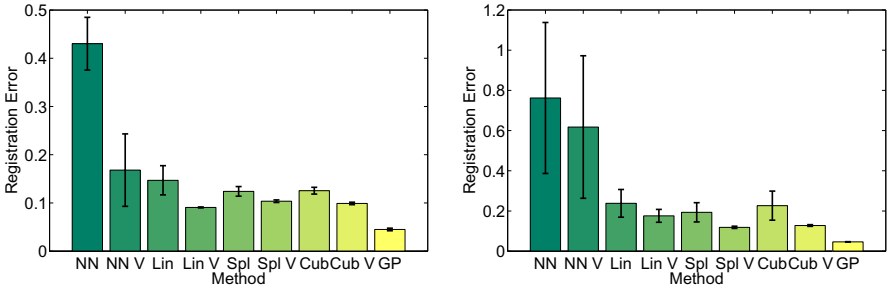
We use this similarity measure in combination with block-wise estimation. For constant variances Σ_{xx}, this corresponds to the common sum of squared differences (SSD).

To make the concept of uncertainty estimation in interpolation easy to integrate in existing applications, we propose an approximation for the variance values Σ_{xx} without performing GP regression. In this case, we use classic interpolation methods to construct the resampled image. Considering the covariance matrix in Eq. (3), we see that it only depends on the *locations* of the observations and predictions, but not on the observed values. We use the interpolation weights, as defined in linear interpolation, to approximate the elementwise variance values Σ_{xx}. We consider the prediction for a point $\mathbf{x}^*$ on the regular grid with spacing $\mathbf{s}$ and let $\mathbf{d} = \mathbf{x}^* - \mathbf{x}$ be the difference vector to the closest point on the base grid $\mathbf{x}$. We approximate the variance at location x^* with

$$v(\mathbf{x}^*) = \sum_{i=1}^{D} |d_i| \cdot (s_i - |d_i|), \tag{12}$$

where D is the dimensionality of the image. $v(\mathbf{x}^*)$ is the highest for locations that are equidistant from the base grid nodes, and zero when $\mathbf{x}^*$ lies on the base grid. We illustrate the variances for the approximation and the Gaussian process in 1D and 2D in the supplementary material, which shows that the approximation closely follows the true estimates from the Gaussian process.

There are two important parameters that affect the interpolation; the noise variance σ_J^2 and the length-scale l of the kernel. If we set $\sigma_J^2 = 0$, the interpolator passes exactly through the observations. For $\sigma_J^2 > 0$, the method accepts noise in the observations so that the images can deviate from the observations. The length-scale determines the region of influence of each observation. For shorter length-scale, the prediction is only dependent on a few observations, causing more sensitive results. For larger length-scale, we obtain smoother results. Noise reduction and smoothing are common pre-processing steps for image registration and they can be naturally modeled within the proposed Gaussian process framework. Finally, the interpolation on irregular grids does not pose problems because the method depends on pairwise distances between points only.



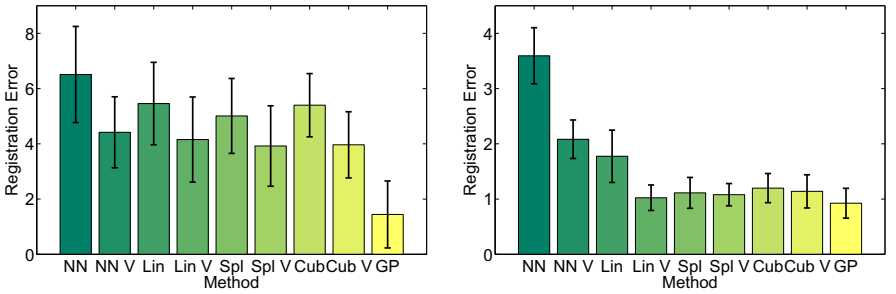
(a) 2D registration study BrainWeb (b) 2D registration study RIRE

Fig. 3. Bars indicate mean registration error; error bars show standard error. Nearest neighbor (NN), Linear (Lin), Spline (Spl), Cubic (Cub), and Gaussian Process (GP) interpolation is reported. The use of the variance approximation is indicated with 'V'.

3 Results

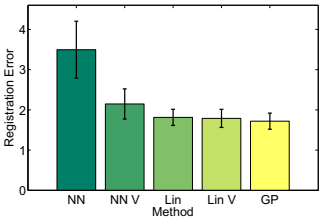
In our registration experiments, we focus on a rigid transformation model. This choice allows us to better isolate the effects of image interpolation in registration, which is the contribution of this work. Moreover, rigid registration enables exact computation of registration errors with respect to ground truth transformations on real data, which is challenging for transformation models with more degrees of freedom. We perform the first set of registration experiments on the publicly available BrainWeb [2] and RIRE [3] datasets. We set $\sigma_I^2 = \sigma_J^2 = 0.1$ in all experiments. First, we select axial slices and perform 2D registration. We downsample the images in one direction by a factor of 5 to simulate anisotropic data. Such anisotropy is commonly present in clinical practice. We transform the grid and create the fixed image by downsampling the original image. For this 2D registration experiment, we can calculate the GP interpolation ($l = 2.5$) without splitting the image into blocks. Consequently, we use the similarity measure in Eq. (10) with the full covariance matrix. For comparison, we perform nearest neighbor, linear, cubic, and spline interpolation with SSD as a similarity measure. Moreover, we compute the approximated variance in Eq. (12) and use it in the similarity measures in Eq. (11), indicated with 'V' in the plots. The mean image $\boldsymbol{\mu}_J$ from the Gaussian process regression is replaced by the nearest neighbor, linear, cubic, or spline interpolator in this case. Fig. 3 shows results over 50 runs from random initial transformations.

In a second experiment, we perform 3D experiments on the BrainWeb and RIRE datasets. Again we downsample the images in one direction by a factor of 5, to create anisotropic volumes. For the Gaussian process interpolation ($l = 2.5$), we split the image into $8 \times 8 \times 8$ cubes to limit the computational costs. Since we do not construct the entire covariance matrix Σ_J in this case, we work with a diagonal covariance matrix in the similarity measure in Eq. (10). The evaluation of the baseline methods with SSD and the variance approximation is analogous to the 2D experiment. Fig. 4 reports the mean RMS errors and standard errors.



(a) 3D registration study BrainWeb (b) 3D registration study RIRE

Fig. 4. Bars indicate mean registration error; error bars show standard error. Nearest neighbor (NN), Linear (Lin), Spline (Spl), Cubic (Cub), and Gaussian Process (GP) interpolation is reported. The use of the variance approximation is indicated with 'V'.



The final datatset consists of two MR images of the head that were acquired on two different grids in the MR scanner with a resolution of $3 \times 3 \times 3.6\text{mm}^3$. The primary slice direction is sagittal for the first image and axial for the second scan. We can access the transformation of each image with respect to the scanner coordinate system. Consequently, the ground truth transformation in our rigid registration experiments that relates both volumes is available. The registration is repeated 50 times for each configuration. The mean RMS errors and standard errors are plotted in the figure on the right. We compare to the nearest neighbor and linear interpolation. For the Gaussian process interpolation ($l = 2.5$), we divide the image into $8 \times 8 \times 8$ cubes to limit the computational costs. Again, we only use the variance model and not the full covariance matrix Σ_J.

Our results show a large decrease in registration error for more complex interpolation techniques than nearest neighbor interpolation. The decrease from linear interpolation to cubic or spline interpolation is less pronounced. Spline interpolation leads to the best registration results among the classical interpolation schemes. In all experiments, using uncertainty estimates in the similarity measure leads to more accurate registration results. The improvement is largest for nearest neighbor interpolation, where the interpolation quality decreases the most when moving further away from the grid points. This finding is interesting for the registration of categorical or label data, where more complex interpolation methods cannot be applied. For the other interpolation schemes, we also notice a substantial improvement for the uncertainty estimate, especially for linear interpolation. Registration with Gaussian processes achieved the best performance in all experiments. This supports the use of the mean function as high quality interpolator and the covariance matrix as uncertainty estimate for registration.

4 Conclusion

We proposed to integrate interpolation uncertainty into registration. To this end, we defined distributions over images based on Gaussian processes with the covariance of the posterior distribution serving as an uncertainty estimate. A novel generative model for registration with Gaussian processes yielded a similarity measure that incorporates interpolation uncertainty. Our results demonstrated improvement for image resampling and the necessity of integrating interpolation uncertainty in the similarity measure.

Acknowledgements. This work was supported in part by the Humboldt foundation, the National Alliance for Medical Image Computing (U54-EB005149), the NeuroImaging Analysis Center (P41-EB015902) and the National Center for Image-Guided Therapy (P41-EB015898).

References

1. Aljabar, P., Hajnal, J., Boyes, R., Rueckert, D.: Interpolation artefacts in non-rigid registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005, Part II. LNCS, vol. 3750, pp. 247–254. Springer, Heidelberg (2005)
2. Cocosco, C.A., Kollokian, V., Kwan, R.S., Evans, A.C.: Brainweb: Online interface to a 3d mri simulated brain database. NeuroImage 5(4) (May 1997)
3. Fitzpatrick, J.M., West, J.B., Maurer, C.R.: Predicting error in rigid-body point-based registration. IEEE Transactions on Medical Imaging 17(5), 694–702 (1998)
4. Hill, D., Batchelor, P., Holden, M., Hawkes, D.: Medical image registration. Physics in Medicine and Biology 46(3) (2001)
5. Hou, H., Andrews, H.: Cubic splines for image interpolation and digital filtering. IEEE Trans. on Acoustics, Speech and Signal Processing 26(6), 508–517 (1978)
6. Kim, K., Lee, D., Essa, I.: Gaussian process regression flow for analysis of motion trajectories. In: Int. Conference on Computer Vision, pp. 1164–1171 (2011)
7. Lehmann, T., Gonner, C., Spitzer, K.: Survey: Interpolation methods in medical image processing. IEEE Transactions on Medical Imaging 18(11), 1049–1075 (1999)
8. Liu, P.: Using Gaussian Process Regression to Denoise Images and Remove Artefacts from Microarray Data. Ph.D. thesis, University of Toronto (2007)
9. Lüthi, M., Jud, C., Vetter, T.: Using landmarks as a deformation prior for hybrid image registration. In: Mester, R., Felsberg, M. (eds.) DAGM 2011. LNCS, vol. 6835, pp. 196–205. Springer, Heidelberg (2011)
10. Parker, J., Kenyon, R., Troxel, D.: Comparison of interpolating methods for image resampling. IEEE Transactions on Medical Imaging 2(1), 31–39 (1983)
11. Rasmussen, C., Williams, C.: Gaussian processes for machine learning. MIT Press (2006)
12. Sollich, P., Williams, C.K.I.: Using the equivalent kernel to understand gaussian process regression. In: Neural Inform. Processing Systems, pp. 1313–1320 (2005)
13. Stytz, M.R., Parrott, R.W.: Using kriging for 3d medical imaging. Computerized Medical Imaging and Graphics 17(6), 421–442 (1993)
14. Thévenaz, P., Blu, T., Unser, M.: Image interpolation and resampling. In: Handbook of Medical Imaging, Processing and Analysis, pp. 393–420 (2000)
15. Wachinger, C., Sharp, G.C., Golland, P.: Contour-driven regression for label inference in atlas-based segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 211–218. Springer, Heidelberg (2013)