

# Registration of Cervical MRI Using Multifeature Mutual Information

Marius Staring\*, Uulke A. van der Heide, Stefan Klein, Max A. Viergever, and Josien P. W. Pluim

**Abstract**—Radiation therapy for cervical cancer can benefit from image registration in several ways, for example by studying the motion of organs, or by (partially) automating the delineation of the target volume and other structures of interest. In this paper, the registration of cervical data is addressed using mutual information (MI) of not only image intensity, but also features that describe local image structure. Three aspects of the registration are addressed to make this approach feasible. First, instead of relying on a histogram-based estimation of mutual information, which poses problems for a larger number of features, a graph-based implementation of  $\alpha$ -mutual information ( $\alpha$ -MI) is employed. Second, the analytical derivative of  $\alpha$ -MI is derived. This makes it possible to use a stochastic gradient descent method to solve the registration problem, which is substantially faster than nonderivative-based methods. Third, the feature space is reduced by means of a principal component analysis, which also decreases the registration time. The proposed technique is compared to a standard approach, based on the mutual information of image intensity only. Experiments are performed on 93 T2-weighted MR clinical data sets acquired from 19 patients with cervical cancer. Several characteristics of the proposed algorithm are studied on a subset of 19 image pairs (one pair per patient). On the remaining data (36 image pairs, one or two pairs per patient) the median overlap is shown to improve significantly compared to standard MI from 0.85 to 0.86 for the clinical target volume (CTV,  $p = 2 \cdot 10^{-2}$ ), from 0.75 to 0.81 for the bladder ( $p = 8 \cdot 10^{-6}$ ), and from 0.76 to 0.77 for the rectum ( $p = 2 \cdot 10^{-4}$ ). The registration error is improved at important tissue interfaces, such as that of the bladder with the CTV, and the interface of the rectum with the uterus and cervix.

**Index Terms**— $\alpha$ -mutual information, kNN graphs, cervical cancer, local image structure, nonrigid registration, radiation therapy, Shannon mutual information.

## I. INTRODUCTION

CERVICAL cancer is a common type of cancer in women, with 493 000 incidences and 274 000 deaths worldwide in 2002 [1]. If surgery is not possible, external-beam radiation therapy is the treatment of choice [2], in combination with

Manuscript received December 02, 2008; revised February 20, 2009. First published March 10, 2009; current version published August 26, 2009. This work was supported by the Netherlands Organisation for Scientific Research (NWO). Asterisk indicates corresponding author.

\*M. Staring was with the University Medical Center Utrecht, Image Sciences Institute, 3508 GA Utrecht, The Netherlands. He is now with Leiden University Medical Center, 2300 RA Leiden, The Netherlands (e-mail: m.staring@lumc.nl).

S. Klein, M. A. Viergever, and J. P. W. Pluim are with the University Medical Center Utrecht, Image Sciences Institute, 3508 GA Utrecht, The Netherlands (e-mail: stefan@isi.uu.nl; max@isi.uu.nl; josien@isi.uu.nl).

U. A. van der Heide is with the Department of Radiotherapy, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands (e-mail: u.a.vanderheide@umcutrecht.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2009.2016560

chemotherapy [3]. During therapy, a patient is irradiated several times in succession. In our hospital, a patient is irradiated five times a week for a duration of five weeks. Several factors are important for the success of radiotherapy. The first factor concerns the dose targeting. The region that should receive a high dose is the clinical target volume (CTV), which includes the gross tumor volume (GTV) and a region around the GTV with suspicion of micro-metastases. Neighboring tissue such as the bladder and rectum should receive a dose as low as possible, to reduce complications during and after therapy. Therefore, accurate localization of the cervix and surrounding structures is needed to effectively aim the dose delivery. A second factor is the ability to detect whether treatment plans need to be adapted to changes in the anatomy, for example due to response of the tumor to treatment. Adaptation of the treatment plans can improve sparing of bowel and rectum [4]. Pötter *et al.* [3] expect that it can have a minimizing effect on treatment related morbidity. A third important factor in radiation therapy concerns the treatment margins used to accommodate uncertainties in the position of the target volumes [5]. The uncertainty can be high due to the large movement of the bladder, rectum and intestines. Quantification of these geometrical changes (i.e., identification of the true movements) may help to reduce the treatment margin of these structures [6]–[9].

Image registration can aid in all these cases. Segmentation of the relevant structures can be achieved by an atlas-based registration approach, as has been done for the prostate [10]. Another approach is to automatically update the manually created treatment plan of the first day, to that of the current day, instead of creating new manual segmentations. Automatic updates can be realized by performing a nonrigid intrasubject registration, and propagating the segmentation of the first day to the current day. Even if the quality of these automatic segmentations is not sufficient for direct clinical use, it can be used as an initialization for annotators, which can save time and reduce observer variability [11]. To detect the need for adapting the radiotherapy treatment plans, one can register the current image with the baseline image. The volume of the GTV in combination with its motion are indicators for the necessity to update the treatment plans [12]. Geometrical changes can be quantified by the transformation resulting from image registration.

While traditionally CT images are used for radiotherapy, MRI is increasingly added, since it is superior to CT for staging, delineation of the relevant organs, and for measuring cervical carcinoma size and uterine extension, see [13]–[15]. Factors that hinder the success of image registration are MR imaging artifacts, the highly anisotropic voxels, and the large changes between successive treatment days (interfraction variability) of the

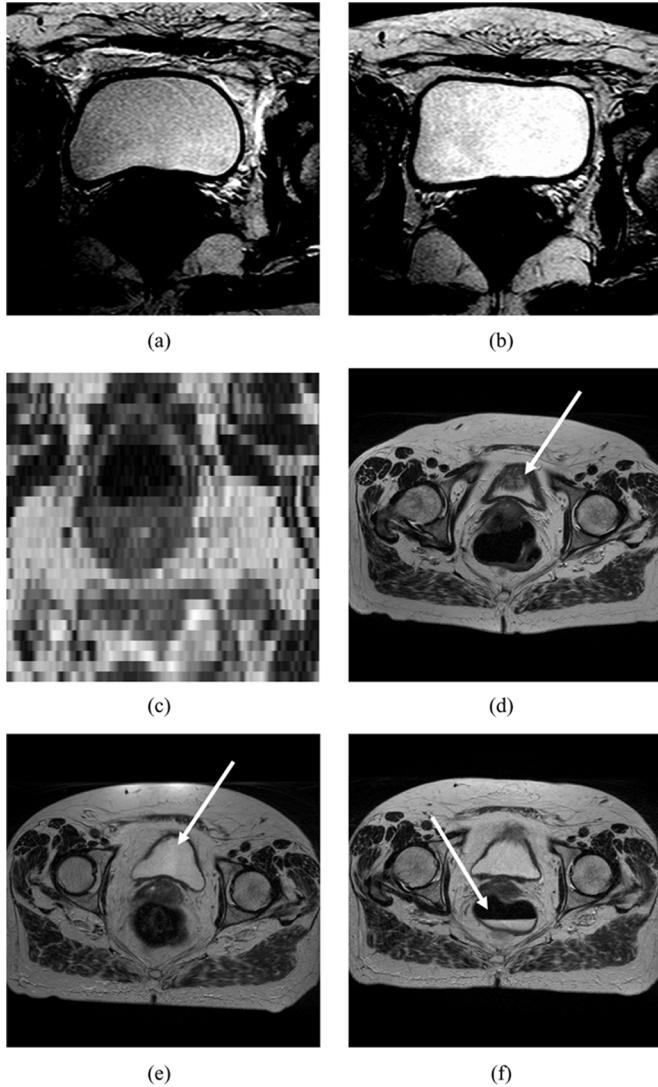


Fig. 1. Examples of limiting factors for image registration. Figures (a) and (b) show two scans of the same patient, intensity windowed for optimal display of the bladder. Figure (b) demonstrates an intensity inhomogeneity over the bladder. (c) Slice in the coronal plane to illustrate data anisotropy. In (d)–(f) three scans of another patient are shown, exhibiting large variations in bladder filling [compare (d) with (e)] and rectum filling [compare (d), (e) with (f)].

position, shape and size of the bladder, rectum, and intestines in particular. In Fig. 1, examples of these limiting factors are given.

In this paper, we present an intensity-based nonrigid registration method to automatically align intrapatient MR images of the cervix, for the purposes mentioned above. A standard approach, based on the mutual information of the intensity of the MR images only, may not be sufficient to overcome all of the limiting factors. Therefore, we propose to use an algorithm that additionally incorporates features, in our case features that describe local image structure. The feature information is utilized by a multidimensional  $\alpha$ -mutual information measure ( $\alpha$ -MI), and is shown to improve the registration quality.

In order to make  $\alpha$ -MI feasible for a large number of features, we address several aspects of the method.

- Multidimensional  $\alpha$ -mutual information is implemented by computing the length of entropic graphs [16]–[19]. Commonly, mutual information is implemented by the

estimation of a joint histogram. The joint histogram is of dimension  $2d$ , with  $d$  the number of features. Therefore, it suffers from the so-called curse of dimensionality: more and more samples are needed to fill the joint histogram up to a sufficient level, so that MI can be reliably estimated. The need to estimate the joint histogram can be eliminated, by directly relating entropy to the data, without estimating the probability density function. The choice for the entropic graphs implementation makes it possible to use multidimensional mutual information for a larger number of features.

- The analytical derivative of  $\alpha$ -MI with respect to the transformation parameters is derived. With the analytical derivative the registration problem can be solved using a stochastic gradient descent method. This is much faster, compared to a finite difference optimization routine or a nonderivative-based optimizer.
- The feature space is reduced by employing a principal component analysis (PCA) on the total set of features. This also decreases the registration time substantially.

$\alpha$ -Mutual information for image registration was introduced by Hero *et al.* [16], [17]. It has been used on 2-D data [18], [20], sometimes using a nonderivative-based optimizer [21]. In [22], a descent direction for a Euclidean minimal spanning tree is derived, and some results for rigid registration are given. Oubel *et al.* [23] apply  $\alpha$ -joint entropy to 3-D tagged MR sequences of the heart, for which they also introduce an analytical derivative. The method was evaluated on five sequences, using only two features: the intensity from two different viewing directions.

In Section II details about the registration framework are given,  $\alpha$ -MI is explained, and the analytical derivative is deduced. The features used for the registration of the cervical data are discussed in Section III. For tuning and evaluation, multiple T2-weighted MR data sets are available for 19 patients. Registration parameters are tuned on a subset of 19 image pairs. Evaluation is performed on the remaining 36 image pairs by comparing automatic segmentations with a manually established ground truth, see Sections IV and V.

## II. METHOD

### A. Registration Framework

Registration of a  $q$ -dimensional moving image  $I_M : \Omega_M \subset \mathbb{R}^q \rightarrow \mathbb{R}$  to a fixed image  $I_F : \Omega_F \subset \mathbb{R}^q \rightarrow \mathbb{R}$  can be formulated as an optimization problem

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \mathcal{S}(T\boldsymbol{\mu}; I_F, I_M, \Omega_F) \quad (1)$$

with  $T\boldsymbol{\mu} : \Omega_F \rightarrow \Omega_M$  the transformation modelled by the vector of parameters  $\boldsymbol{\mu} \in \mathbb{R}^M$ , and  $\mathcal{S}$  a suitable cost function (dissimilarity measure) that is minimal when  $I_F(x)$  and  $I_M(T\boldsymbol{\mu}(x))$  are aligned. It is possible to additionally include a regularization term in the cost function, but this is omitted in this work.

### B. Mutual Information

The mutual information [24], [25] of image intensities is defined as:  $MI = H(I_F) + H(I_M) - H(I_F, I_M)$ , with  $H(I) = -\sum p_i \log p_i$  the entropy of the intensities of image  $I$ ,

$H(I_F, I_M)$  the joint entropy, and  $p_i$  the intensity probabilities. It is usually implemented by constructing a joint histogram, which estimates the joint intensity probabilities. Details about the specific implementation used in our article can be found in [26], in which the analytical derivative of MI can also be found.

### C. Entropy, Entropic Graphs, and $\alpha$ -mutual Information

Mutual information is commonly calculated on the intensities only, i.e., on 1-D signals, by estimation of a joint histogram. For higher-dimensional signals, however, this computation method poses problems, as outlined in Section I. Therefore, we opt for an implementation of  $\alpha$ -mutual information introduced by Hero *et al.* [16], [17] based on entropic graphs.

Define  $\mathbf{z}(\mathbf{x}_i) = [z_1(\mathbf{x}_i), \dots, z_d(\mathbf{x}_i)]$  to be a vector of dimension  $d$  containing all feature values at point  $\mathbf{x}_i$ . For example, the first index  $z_1$  is the intensity of the image, and  $z_2, z_3$ , and  $z_4$  contain the values of the spatial derivatives of that image. It was shown by Beardwood *et al.* [27] (see also [28]) that the length of the graph that connects the feature points  $\{\mathbf{z}_i\}$ , is related to the Rényi  $\alpha$ -entropy [29]. It can be shown, using l'Hopital's rule, that the limit  $\alpha \rightarrow 1$  of  $\alpha$ -entropy equals the Shannon entropy. Not all graphs are suited to compute  $\alpha$ -entropy, but for example the minimal spanning tree (MST) and the  $k$ -nearest neighbor ( $k$ NN) graph are [17]. We choose to use the  $k$ NN graph, since it is computationally attractive relative to the MST [17], [30].

Let  $\mathbf{z}^f(\mathbf{x}_i)$  be the feature vector of the fixed image at a point  $\mathbf{x}_i$ , and  $\mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_i))$  that of the moving image at the transformed point  $\mathbf{T}_\mu(\mathbf{x}_i)$ . Let  $\mathcal{Z}_f = \{\mathbf{z}^f(x_1), \dots, \mathbf{z}^f(x_N)\}$  be the collection of  $N$  feature vectors drawn from the fixed image, and  $\mathcal{Z}_m$  that of the moving feature vectors. Let  $\mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_\mu(\mathbf{x}_i))$  be the concatenation of the two feature vectors:  $[\mathbf{z}^f(\mathbf{x}_i), \mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_i))]$ , with corresponding collection  $\mathcal{Z}_{fm}$ . Three  $k$ NN graphs can be constructed on the three collections, where the total distance of a feature vector  $\mathbf{z}$  to its  $k$  nearest neighbors is given by

$$\Gamma_i^f = \sum_{p=1}^k \|\mathbf{z}^f(\mathbf{x}_i) - \mathbf{z}^f(\mathbf{x}_{ip})\| \quad (2)$$

$$\Gamma_i^m(\boldsymbol{\mu}) = \sum_{p=1}^k \|\mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_i)) - \mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_{ip}))\| \quad (3)$$

$$\Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^k \|\mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_\mu(\mathbf{x}_i)) - \mathbf{z}^{fm}(\mathbf{x}_{ip}, \mathbf{T}_\mu(\mathbf{x}_{ip}))\| \quad (4)$$

with  $\mathbf{z}^f(\mathbf{x}_{ip})$ ,  $\mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_{ip}))$ , and  $\mathbf{z}^{fm}(\mathbf{x}_{ip}, \mathbf{T}_\mu(\mathbf{x}_{ip}))$  the  $p$ th nearest neighbor of  $\mathbf{z}^f(\mathbf{x}_i)$ ,  $\mathbf{z}^m(\mathbf{T}_\mu(\mathbf{x}_i))$ , and  $\mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_\mu(\mathbf{x}_i))$ , respectively. Note that we used the Euclidean distance in (2)–(4). Also note that the neighbors of the three graphs in general do not correspond. A graph-based estimator for  $\alpha$ -MI is defined as [17]:

$$\alpha\widehat{\text{MI}}(\boldsymbol{\mu}; \mathcal{Z}_f, \mathcal{Z}_m, \mathcal{Z}_{fm}) = \frac{1}{\alpha-1} \log \frac{1}{N^\alpha} \sum_{i=1}^N \left( \frac{\Gamma_i^{fm}(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}} \right)^{2\gamma} \quad (5)$$

with  $\gamma = d(1-\alpha)$ , and  $0 < \alpha < 1$  a user-defined constant. In this paper, we used  $\alpha = 0.99$  for all experiments.

### D. Optimization

To solve the minimization problem (1), an iterative stochastic gradient descent optimization strategy is employed [31]

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + a_t \tilde{\mathbf{g}}_t \quad (6)$$

with  $\tilde{\mathbf{g}}_t$  an approximation of the exact derivative of the similarity measure evaluated at iteration  $t$ :  $\partial S / \partial \boldsymbol{\mu}(\boldsymbol{\mu}_t)$ . The derivative is approximated by computing it over a subset of all points  $x \in \Omega_F$ . The subset contains  $N$  points, which were chosen randomly in every iteration. This was shown to substantially accelerate the registration, while retaining convergence properties [31]. Samples that are not in the overlap of the fixed and deformed moving image domain are excluded. For the step size  $a_t$  a decaying function of the iteration number is used:  $a_t = a/(A+t)^\tau$ , with  $a > 0$ ,  $A \geq 1$  and  $0 < \tau \leq 1$  user-defined constants [31].

### E. Derivative of $\alpha$ -mutual Information

For the multidimensional case, registration is written as:  $\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \alpha \widehat{\text{MI}}(\boldsymbol{\mu}; \mathcal{Z}_f, \mathcal{Z}_m, \mathcal{Z}_{fm})$ . In order to apply the stochastic gradient descent method (6), the derivative  $\partial / \partial \mu_j \alpha \widehat{\text{MI}}(\boldsymbol{\mu}; \mathcal{Z}_f, \mathcal{Z}_m, \mathcal{Z}_{fm})$  is required. Since computing the derivative using finite differences is extremely slow for a large number of transformation parameters  $M$ , we derive an analytical expression for it. The expression we derive is valid as long as graph topology does not change for small changes of  $\boldsymbol{\mu}$ .

For compact notation, define

$$G_i(\boldsymbol{\mu}) = \frac{\Gamma_i^{fm}(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}} \quad (7)$$

and

$$\mathbf{d}_{ip}^{fm}(\boldsymbol{\mu}) = \mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_\mu(\mathbf{x}_i)) - \mathbf{z}^{fm}(\mathbf{x}_{ip}, \mathbf{T}_\mu(\mathbf{x}_{ip})) \quad (8)$$

and similarly for  $\mathbf{d}_{ip}^m$ . Then the derivative of  $\alpha$ -MI equals

$$\begin{aligned} \frac{\partial}{\partial \mu_j} \alpha \widehat{\text{MI}}(\boldsymbol{\mu}; \mathcal{Z}_f, \mathcal{Z}_m, \Omega_F) \\ = \frac{-2d}{\sum_{i=1}^n G_i(\boldsymbol{\mu})^{2\gamma}} \sum_{i=1}^n G_i(\boldsymbol{\mu})^{2\gamma-1} \frac{\partial}{\partial \mu_j} G_i(\boldsymbol{\mu}). \end{aligned} \quad (9)$$

The derivative of  $G_i(\boldsymbol{\mu})$  is written as

$$\frac{\partial}{\partial \mu_j} G_i(\boldsymbol{\mu}) = \frac{\frac{\partial}{\partial \mu_j} \Gamma_i^{fm}(\boldsymbol{\mu}) - \frac{1}{2} \Gamma_i^{fm}(\boldsymbol{\mu}) \Gamma_i^m(\boldsymbol{\mu})^{-1} \frac{\partial}{\partial \mu_j} \Gamma_i^m(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}}. \quad (10)$$

From (10) we expand the derivative of  $\Gamma_i^{fm}(\boldsymbol{\mu})$

$$\frac{\partial}{\partial \mu_j} \Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^k \frac{\partial}{\partial \mu_j} \sqrt{\mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})^T \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})} \quad (11)$$

$$= \sum_{p=1}^k \frac{1}{2} \left\| \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu}) \right\|^{-1} \frac{\partial}{\partial \mu_j} (\mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})^T \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})) \quad (12)$$

$$= \sum_{p=1}^k \frac{\mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})^T}{\left\| \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu}) \right\|} \cdot \frac{\partial}{\partial \mu_j} \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu}) \quad (13)$$

TABLE I  
IRREDUCIBLE SET OF SECOND ORDER CARTESIAN  
IMAGE STRUCTURE INVARIANTS FOR 3-D

Einstein notation	matrix notation
$L$	$L$
$L_i L_i$	$\mathbf{g}^T \mathbf{g}$
$L_i L_{ij} L_j$	$\mathbf{g}^T H \mathbf{g}$
$L_i L_{ij} L_{jk} L_k$	$\mathbf{g}^T H H \mathbf{g}$
$L_{ii}$	$tr(H)$
$L_{ij} L_{ji}$	$tr(HH)$
$L_{ij} L_{jk} L_{ki}$	$tr(HHH)$

where (11) follows from (4) and the definition of the Euclidean distance metric. Only the derivative to the moving image features matter, since  $\partial/\partial\mu_j \mathbf{d}_{ip}^{fm}(\boldsymbol{\mu}) = [\mathbf{0}, \partial/\partial\mu_j \mathbf{d}_{ip}^m(\boldsymbol{\mu})]^T$ . So

$$\begin{aligned} \frac{\partial}{\partial\mu_j} \Gamma_i^{fm}(\boldsymbol{\mu}) &= \sum_{p=1}^k \frac{\mathbf{d}_{ip}^m(\boldsymbol{\mu})^T}{\|\mathbf{d}_{ip}^{fm}(\boldsymbol{\mu})\|} \cdot \left[ \frac{\partial \mathbf{z}^m}{\partial \mathbf{x}_i}(\mathbf{T}\boldsymbol{\mu}(\mathbf{x}_i)) \right. \\ &\quad \left. \cdot \frac{\partial \mathbf{T}\boldsymbol{\mu}}{\partial\mu_j}(\mathbf{x}_i) - \frac{\partial \mathbf{z}^m}{\partial \mathbf{x}_i}(\mathbf{T}\boldsymbol{\mu}(\mathbf{x}_{ip})) \frac{\partial \mathbf{T}\boldsymbol{\mu}}{\partial\mu_j}(\mathbf{x}_{ip}) \right]. \end{aligned} \quad (14)$$

The derivative  $\partial/\partial\mu_j \Gamma_i^m(\boldsymbol{\mu})$  can be derived similarly. Summarizing, the derivative of  $\alpha$ - $\widehat{\text{MI}}$  can be expressed in terms of the spatial derivative of the moving (feature) images  $\partial \mathbf{z}^m / \partial \mathbf{x}_i$ , the Jacobian of the transformation  $\partial \mathbf{T}\boldsymbol{\mu} / \partial \mu_j$ , the differences  $\mathbf{d}_{ip}^{fm}$  and  $\mathbf{d}_{ip}^m$ , and the graph distances  $\Gamma_i^f$ ,  $\Gamma_i^m(\boldsymbol{\mu})$  and  $\Gamma_i^{fm}(\boldsymbol{\mu})$ . Derivation of the analytical derivative is similar to that of Oubel *et al.* [23], who did the analysis for  $\alpha$ -joint entropy. Note that, in contrast with Oubel, we also take the derivative to the nearest neighbors into account in (14).

### III. FEATURES

Now that the registration framework for multidimensional mutual information has been defined, features need to be chosen. From the vast amount of choices, features that describe the local structure of images supply supplementary knowledge, which may improve the registration. We choose the set of Cartesian image structure features up to the second order derivatives [32], listed in Table I. Here,  $L$  denotes luminance or intensity,  $\mathbf{g} = (\partial L / \partial \mathbf{x})$  the spatial derivative,  $H$  is the Hessian of  $L$ , and  $tr(\cdot)$  denotes the matrix trace. Note that  $L_i L_i$  is the gradient magnitude, and  $L_{ii}$  the Laplacian. All features are invariant to rotation and translation. Additionally, this set is irreducible: other structure invariants of order up to two can be expressed in terms of features in this set, and the set is minimal. The invariants are computed using Gaussian derivatives at scale  $\sigma$ . Features at scale  $\sigma = 1$  and 2 were included. The additional use of  $\sigma = 4$  did not improve results, see Section V-A. In total we have 15 features: 14 features that describe local structure (the seven features from Table I at two scales), and the original intensity data. In Fig. 2(a)–(f), some examples are given. Note that all features are derived from the original image data, and can be computed before registration.

Since the computational complexity of (5) and (9) increases with the number of features  $d$ , and since not all features may have an equal contribution to the registration quality, it could be beneficial to select a subset of the total feature set for inclusion in the registration. We opt to perform a PCA on the complete

feature set, and select only the first  $P \leq d$  principal components for inclusion in the registration.

The features are normalized to have zero mean and unit variance. The PCA is performed on the concatenation of the fixed and moving images, resulting in the  $(N_F + N_M) \times d$  matrix  $[\mathbf{z}^f(\mathbf{x}_1), \dots, \mathbf{z}^f(\mathbf{x}_{N_F}), \mathbf{z}^m(\mathbf{y}_1), \dots, \mathbf{z}^m(\mathbf{y}_{N_M})]^T$ , with  $N_F = |\Omega_F|$  and  $N_M = |\Omega_M|$ . By performing the PCA on the concatenation, rather than on the separate images  $I_F$  and  $I_M$ , an identical linear combination is used for the feature images of the fixed and the moving set. The PCA algorithm returns the principal components (a linear combination of the input), together with the explained variance  $\lambda_i^2, i = 1, \dots, d$ , of each component. The percentage of total explained variance is given by  $\sum_{i=1}^P \lambda_i^2 / \sum_{i=1}^d \lambda_i^2$ . In Fig. 2(g)–(l), the first six principal components, derived from the total feature set of the image in Fig. 2(a), are shown.

### IV. EXPERIMENTS

The graph-based  $\alpha$ -mutual information measure was implemented in the registration package `elastix`, developed by the authors. It is publicly available from <http://elastix.isi.uu.nl>. This package is largely based on the Insight Toolkit (ITK) [33]. The implementation of the  $k$ NN graph is based on the approximate nearest neighbor (ANN) library, freely available from <http://www.cs.umd.edu/~mount/ANN/>.

#### A. Data

The MR data were acquired with a Philips 1.5 T scanner (Gyroscan NT Intera; Philips Medical Systems, Best, The Netherlands), using a T2-weighted sequence taken in the transversal direction. Nineteen patients were scanned. Each patient was scanned five times, one scan each week, except for two patients who were scanned four times.

The image dimensions were  $512 \times 512 \times 30$  voxels of size  $0.625 \times 0.625 \times 4.5$  mm. The data were cropped before registration to a size of  $210 \times 250 \times 30$  voxels such that the data included the relevant structures. Manual segmentations of the GTV, CTV, bladder, and rectum were available for each image. They were created by a radiation oncologist and approved by a radiologist. The segmentations were made with the clinical purpose in mind, and not for evaluation of registration quality specifically.

The data were divided into two sets. The first set contains the images at week 1 and 2 (19 image pairs), and is used for selection of the parameters. The second set consists of the remaining 55 images: 19 pairs of weeks 3 and 4, and 17 pairs of weeks 4 and 5; a total of 36 image pairs. The second set is used for comparing MI and  $\alpha$ -MI with parameters tuned on the first set. With this division, evaluation of the method could be performed on all available patients.

#### B. Registration Settings

The registration parameters were chosen by trial-and-error on the first data set. Images at one week were registered to that of the next week. A rigid registration based on the mutual information of intensities only, was performed prior to nonrigid registration, to get a rough alignment. For the nonrigid registration, the transformation was parameterized with B-splines [34]. We

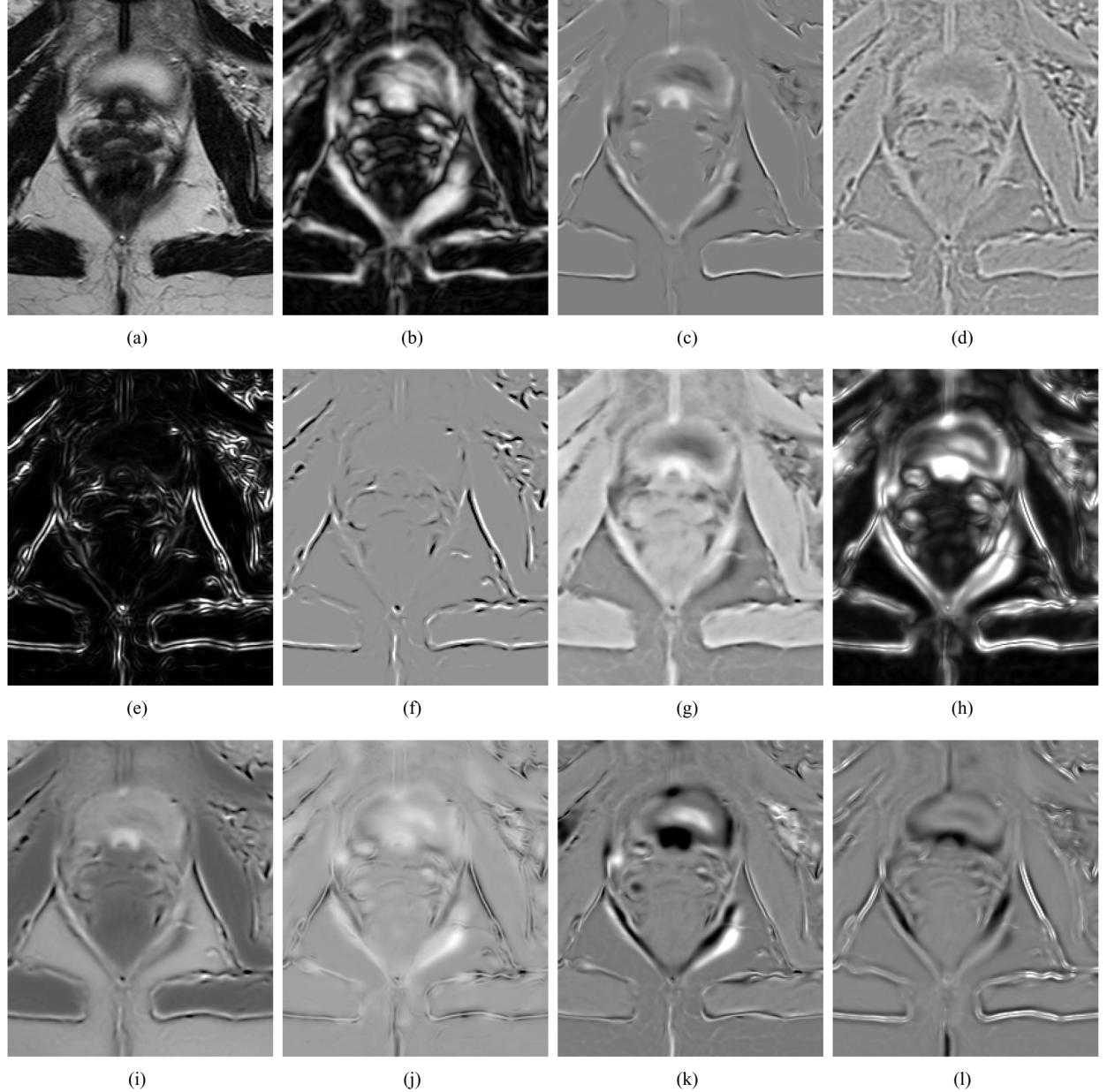


Fig. 2. Examples of the image structure features. (a) original image, (b)–(f) are the features that describe local structure ( $\sigma = 1$ ), with (b)  $L_iL_i$ , (c)  $L_iL_{ij}L_j$ , (d)  $L_{ii}$ , (e)  $L_{ij}L_{ji}$ , (f)  $L_{ij}L_{jk}L_{ki}$ . The first six principal components of the PCA are shown in (g)–(l). For this particular example the total explained variance for increasing  $P$  from 1 to 6 was 57%, 79%, 88%, 93%, 96%, and 98%. The images were intensity windowed for optimal display.

employ a multiresolution scheme with three resolution levels on all feature images. Gaussian smoothing was applied, but no downsampling to more accurately interpolate the moving image  $I_M(\mathbf{T}_\mu(\mathbf{x}))$ . Scales  $\sigma = 4.0, 2.0$ , and  $1.0$  voxels were used in the  $x$  and  $y$  directions. For the  $z$  direction,  $\sigma = 2.0, 1.0$ , and  $0.5$  voxel was used, because of the voxel anisotropy. A multigrid approach was used with a spacing of  $80, 40$ , and  $20$  mm between the B-spline control points for the first, second, and final resolution, respectively. This yielded  $M = 3300$  parameters in the final resolution. For the optimisation procedure,  $A = 50$ ,  $\tau = 0.6$ , and  $a = 2000$  were set. During the parameter selection stage, 300 iterations were used. MI and  $\alpha$ -MI were compared on the second data set with 600 iterations. The number of samples, randomly selected in every iteration was set to  $N = 5000$ ;

this yields good results for standard MI [31]. To be comparable to MI,  $\alpha$  was set to 0.99. For the kNN graph implementation the ANN package was used. We selected kD trees, a standard splitting rule, and a bucket size of 50. The parameter  $\epsilon$  that defines the amount of error that is acceptable when computing the nearest neighbors, see [30], was set to 10. The  $k = 20$  nearest neighbors were used to compute (the derivative of)  $\widehat{\alpha\text{-MI}}$  for data set 1, and  $k = 5$  for data set 2.

### C. Evaluation Measure

To evaluate the registration quality, manual segmentations of the CTV, bladder and rectum were used. Automatic segmentations were generated by transforming the manual segmentation

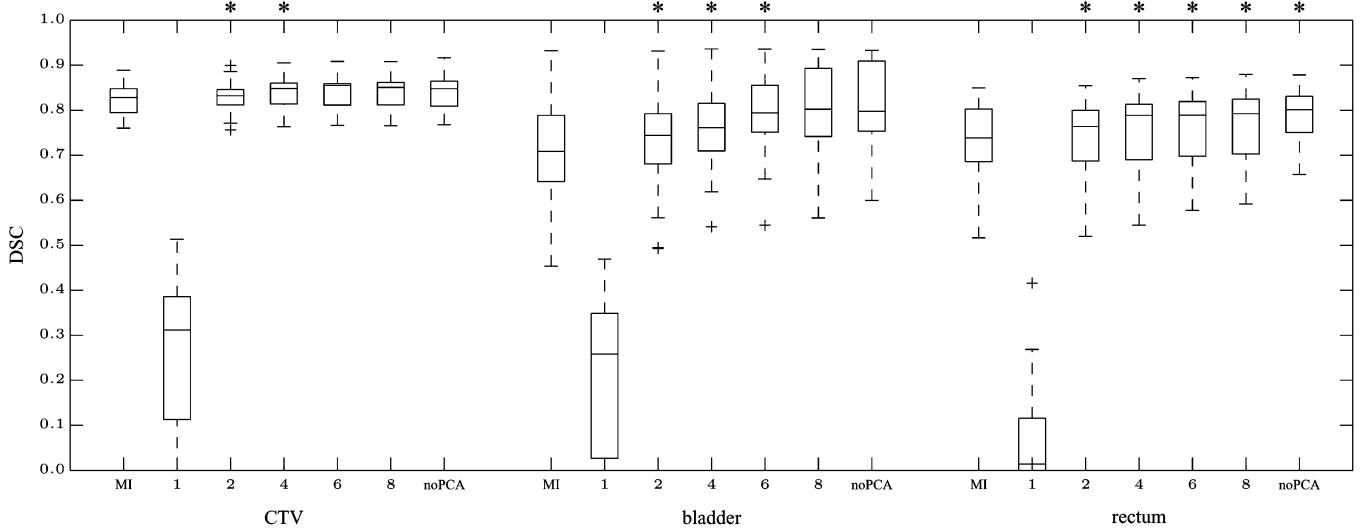


Fig. 3. The effect of feature selection and the number of principal components  $P$ . For each anatomical structure, the leftmost column shows the result for MI, the right-most for  $\alpha$ -MI without PCA ( $d = 15$ ), and the columns in between for  $\alpha$ -MI with PCA. The number below the graph refers to  $P$ . A star indicates a statistical significant difference of the median overlap compared to the previous column, starting from  $P = 2$ .

of the moving image to the fixed image domain, using the transformation  $T_{\hat{\mu}}$  found by the registration. This transformed segmentation  $A$  was compared to the manual segmentation  $B$  of the fixed image, using the Dice similarity coefficient (DSC) as a measure of overlap [35]

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (15)$$

where  $| \cdot |$  denotes the number of voxels within the segmentation. A value of 0 indicates that  $A$  and  $B$  are disjoint, 1 indicates perfect agreement. The DSCs are presented by box-and-whisker plots.

To compare two experiments paired, two-sided Wilcoxon tests [36] were performed on the corresponding overlap values. A value of  $p < 0.05$  was considered to indicate a statistically significant difference.

The DSC does not provide insight into the spatial distribution of the registration errors. To visualize the registration accuracy the shortest Euclidean distance between the manual and automatic segmentation boundaries is computed for every point on the boundary of the manual segmentation [7], [37]. A cartographic “Mollweide equal area” spherical coordinate projection is used to display the result, as proposed in [7].

## V. RESULTS

### A. Results: Data Set 1 (Parameter Selection)

Several aspects of the proposed algorithm were investigated. All experiments described in this section were performed on the first data set.

The selection of the scales of the features was addressed in a first experiment. Compared to using only  $\sigma = 1$ , including features at scale  $\sigma = 1$  and 2 gave slightly better results. The effect was most noticeable for the bladder, where the median overlap increased from 0.76 to 0.98. The additional use of  $\sigma = 4$  did not improve the results (no significant differences in median overlap). Therefore, for  $\alpha$ -MI, scales  $\sigma = 1$  and 2 were chosen.

TABLE II  
RUNTIME FOR THE VARIOUS REGISTRATION EXPERIMENTS. EXPERIMENTS WERE PERFORMED ON A SINGLE CORE OF A STANDARD PC (INTEL Q6600 RUNNING AT 2.4 GHZ.), USING 300 ITERATIONS.  $d = 15$  REFERS TO THE METHOD THAT DOES NOT USE PCA

method		time (min.)	method		time (min.)
MI		0.8			
$\alpha$ -MI	$P = 1$	31		$k = 1$	13
	$P = 2$	41		$k = 2$	17
	$P = 4$	63		$k = 5$	28
	$P = 6$	84		$k = 10$	47
	$P = 8$	106		$k = 20$	84
	$d = 15$	186		$k = 50$	198

In a second experiment, the effect of the number of principal components  $P$  that is used during the registration, was investigated. The results are shown in Fig. 3. If only one principal component was taken into account, most of the registrations failed; the result was worse than MI. Results similar to  $P = 1$  were obtained when using the original intensity images as the single feature in  $\alpha$ -MI. For each of the 19 image pairs the total explained variance resulting from the PCA analysis can be computed with  $\sum_{i=1}^P \lambda_i^2 / \sum_{i=1}^d \lambda_i^2$ . The median over the 19 image pairs was 0.57, 0.80, 0.93, 0.98, and 0.99 for  $P = 1, 2, 4, 6$ , and 8, respectively. Increasing  $P$  improved the results up to approximately  $P = 6$  or  $P = 8$ .

In a third experiment the influence of the use of PCA was examined.  $\alpha$ -MI was performed without the feature space reduction, i.e., with all local image structure features at scale 1 and 2 and the original intensity data (in total 15 features). For each anatomical structure, the right-most column of Fig. 3 shows the result for this experiment. Inclusion of all features did not improve registration performance considerably, except for the rectum where the difference with PCA ( $P = 8$ ) is significant. However, the registration time was substantially increased and no gain was obtained for the CTV and the bladder. Therefore, the best trade-off of computation time and registration accuracy

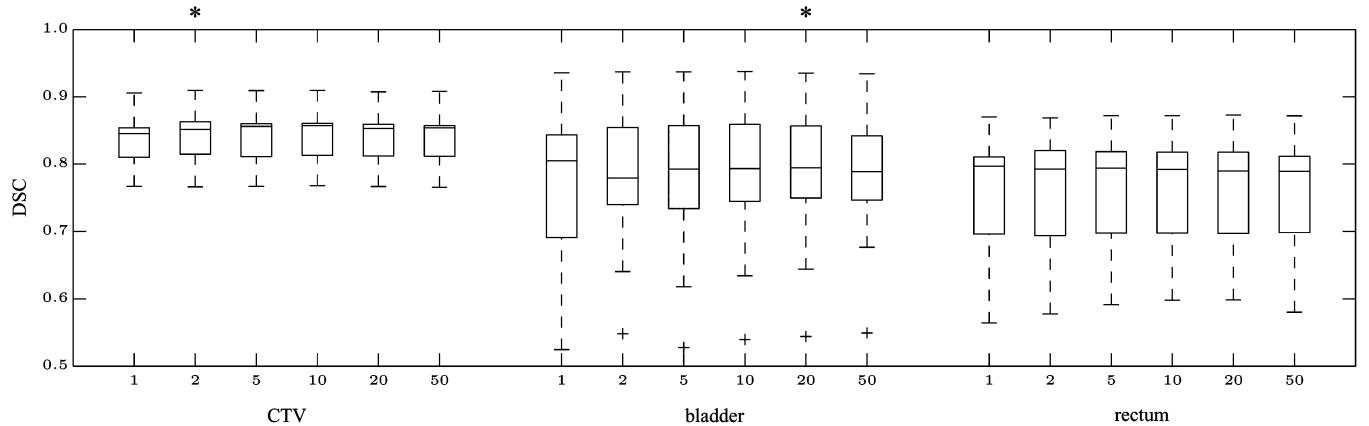


Fig. 4. The effect of  $k$ . The number below the graph refers to  $k$ . A star indicates a statistical significant difference of the median overlap compared to the previous column.

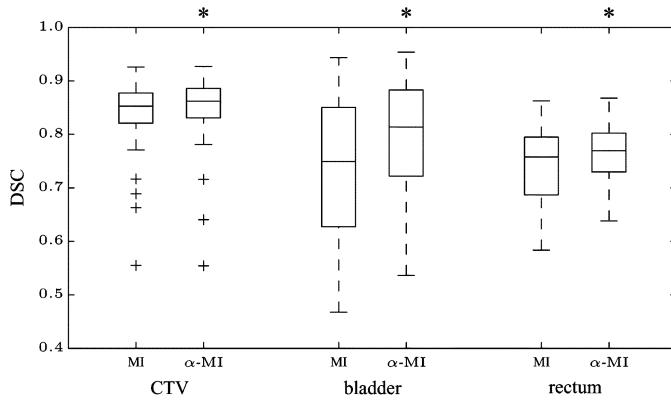


Fig. 5. MI versus  $\alpha$ -MI. For each anatomical structure, the leftmost column shows the result for MI. The second column shows the result for  $\alpha$ -MI ( $P = 6$ ). A star indicates a statistical significant difference of the median overlaps of the two methods.

was with the use of PCA with  $P = 6$ . With this setting the registration time was 84 min, a reduction with a factor of 2.4 compared to using the full feature set, see the left part of Table II. Results similar to standard histogram-based MI were obtained when using the first principal component of the features instead of the intensity; The median overlap decreased by 0.01 for all structures compared to standard histogram-based MI.

A fourth experiment was performed to analyze the influence of the number of nearest neighbors  $k$ . For  $P = 6$ , we varied  $k \in \{1, 2, 5, 10, 20, 50\}$ . The results are depicted in Fig. 4 and are not very sensitive to  $k$ , as long as  $k > 1$ . The registration time depends heavily on  $k$ , see the right part of Table II. We choose  $k = 5$  for the remainder of this paper. With this setting the registration time is 28 min, using 300 iterations.

#### B. Results: Data Set 2 (Evaluation)

MI is compared to  $\alpha$ -MI (PCA,  $P = 6$ ,  $k = 5$ ) on data set 2. For each of the two similarity measures, 36 registrations were performed. The DSC results are shown in Fig. 5. Not much difference between the two methods is seen at the CTV, although the difference is significant (from 0.85 to 0.86,  $p = 2 \cdot 10^{-2}$ ). At the bladder and the rectum, however, the median overlap increases significantly from 0.75 to 0.81 ( $p = 8 \cdot 10^{-6}$ ) and from 0.76 to 0.77 ( $p = 2 \cdot 10^{-4}$ ), respectively. Also note the increase

of the first quartile, meaning that if the technique were to be used in the clinic, less manual correction of the treatment plans is needed when using  $\alpha$ -MI.

In Fig. 6, a typical example of the result of registration is given, for MI (c) and  $\alpha$ -MI (d). Much less deformation is expected at fatty tissue and near bony anatomy, see for example the bottom and right side of Fig. 6(c) and (d). The examples show accurate registrations at those positions, both for MI and  $\alpha$ -MI. A large difference in bladder filling is hard to recover for MI. Although the bladders are not perfectly aligned by  $\alpha$ -MI, the result is better. Registration problems sometimes also occur at the GTV, where the tissue changes due to irradiation. Fig. 6(e) and (f) show an example of such a tissue change, which is difficult to handle for the registration.

For all 36 registrations the registration error was computed at every point of the surface of the CTV, bladder and rectum, for both registration methods. The median and third quartile of the spatial distribution of these errors are given in Fig. 7; for the third quartile 75% of the registrations have an error smaller than the distance depicted in the Fig. 8 shows a surface map to aid the interpretation of Fig. 8. For all structures it can be seen that the registration error of  $\alpha$ -MI is at most that of MI at all locations of the boundary. For the CTV,  $\alpha$ -MI reduces the registration error at the posterior side, i.e., at the interface with the bladder: for MI the third quartile errors are up to 5 mm., for  $\alpha$ -MI the error is at most 3.5 mm. At the bladder a large reduction is visible at the cranial side.  $\alpha$ -MI shows a reduction of the third quartile registration error at the anterior side of the rectum, i.e., at the interface with the uterus and cervix, from 2–3 mm. to 1–2 mm. The registration has also improved at the cranial side of the rectum.

#### VI. DISCUSSION AND CONCLUSION

In this paper, we have introduced a registration method for cervical MR images capable of taking into account multiple image features. Three aspects were addressed to be able to use the method in practice, also for nonrigid registration. First, a graph-based implementation of  $\alpha$ -mutual information was chosen, instead of a histogram-based implementation. Histogram-based approaches are currently not able to cope with a larger number of features, although efforts to do so have

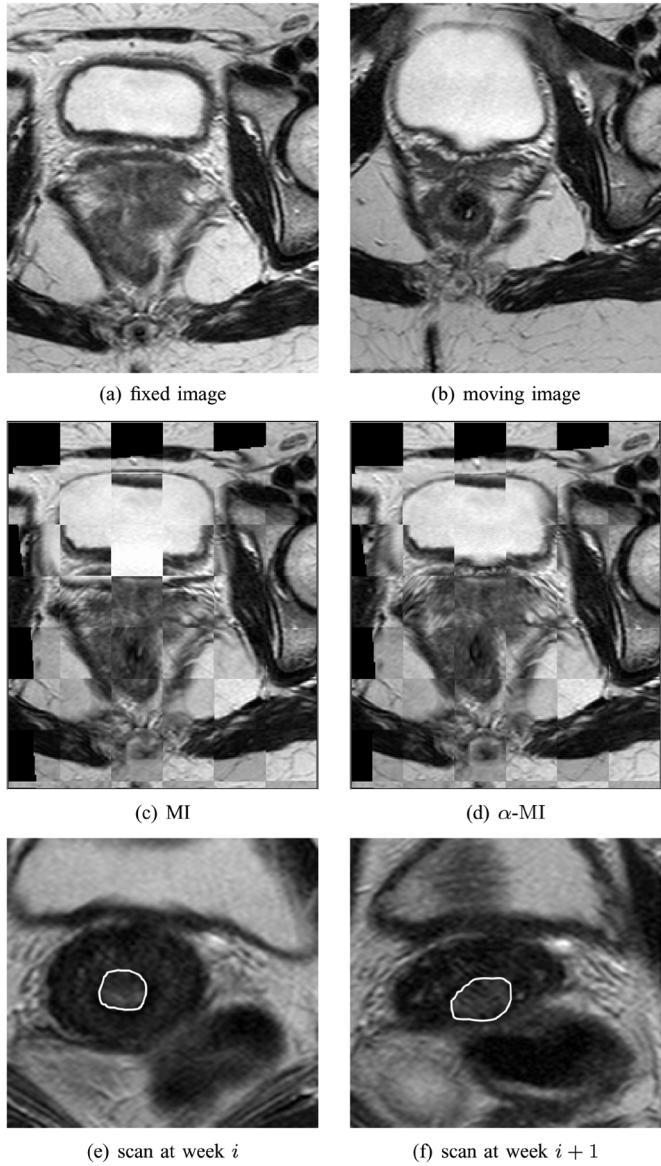


Fig. 6. Example results. The bladder filling of the fixed (a) and moving (b) image is very different; a large deformation is required at that position. The fixed image is combined with the deformed moving image, using a checkerboard pattern. (c) The result of MI clearly is not well aligned at the bladder. (d)  $\alpha$ -MI performs much better for this large deformation. (e) and (f) show the GTV (delineated) at two subsequent weeks. Note the change of the tissue around the GTV.

been undertaken [38]–[40]. The graph-based implementation makes it possible to take into account an enlarged feature space. Second, an analytical derivative was constructed, which enables the use of a fast stochastic gradient descent optimisation routine. A finite difference gradient descent approach would require  $2M$  evaluations of the cost function per iteration, which results in a runtime of  $\approx 170$  h, compared to 62 min when the analytical derivative is employed ( $d = 15, k = 5$ ). Finally, the feature space was reduced by means of a PCA algorithm. This reduces the number of features from 15 to 6, which leads to another reduction in registration time by a factor of  $\approx 2$ . This work is the first to report feasible registration times for multidimensional mutual information with the number of parameters of the transformation in the thousands ( $M = 3300$ ).

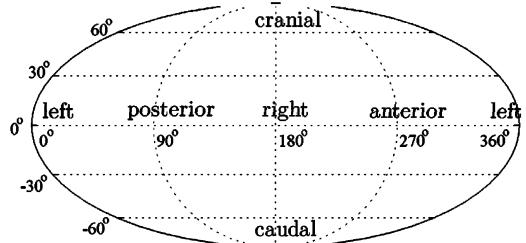


Fig. 7. Mollweide projection of the boundary of structures.

In Fig. 3 it can be seen that the use of a single feature gives bad results for  $\alpha$ -MI. A possible explanation for this behavior can be found by closer inspection of the definition of  $\alpha$ -MI in (5). The expression contains the division by  $\Gamma_i^f$ . In the case  $d = 1$  this can be a very small number when  $k$  out of the  $N$  samples have similar intensity. Division by a small number can make the estimation of  $\alpha$ -MI unstable. For  $d > 1$  the probability that samples have a small distance to their nearest neighbors is in general reduced, except in the unlikely event that the second feature is constant. This observation coincides with our experience that the problem is reduced for  $k > 1$ . The instability of  $\alpha$ -MI due to small distances in the graph-based estimators is addressed in several papers. Kybic [41] proposes to only take neighbors into account with distance greater than some constant  $R$ . Other possible solutions are to add a small amount of noise to the features [42] or to switch to a histogram-like estimator for small distances [43].

It was shown on clinical data of patients with cervical cancer that the proposed method outperforms a standard approach based on the mutual information of intensity only. For use in the clinic, further improvements are still required. One option is the use of a localized version of mutual information [10]. On MR images acquired for the radiation therapy of the prostate this was shown to improve the registration, at no additional computational cost. Another possibility with  $\alpha$ -MI is the inclusion of other imaging data, acquired at the same time point, such as an MR image acquired in another scanning direction, or acquired with a different protocol. For the purpose of updating treatment plans, it is possible to use a multiatlas matching approach [44], once two or more previous scans are available. Rohlfing *et al.* [44] reported improved results for this technique, compared to using a single atlas image.

Propagation of the GTV segmentation is probably not possible with a registration approach only. Even if anatomical correspondence is found, tumorous tissue may disappear in time, due to successful treatment. In this case it might be necessary to employ a tissue classification technique, after the registration.

Although registration using multiple image features is feasible with the proposed method, in our implementation it still requires much more time than the histogram-based approach. For clinical use the computational burden of  $\alpha$ -MI probably needs to be decreased, depending on the specific application. From (5) it is clear that the major part of the computation can be parallelized by distributing the summation over multiple processors. Another interesting approach is to use a minimal spanning tree (MST) instead of a  $k$ NN graph. With the MST,  $\alpha$ -MI is computed as a sum over  $N - 1$  edges; with the  $k$ NN graph it is a sum

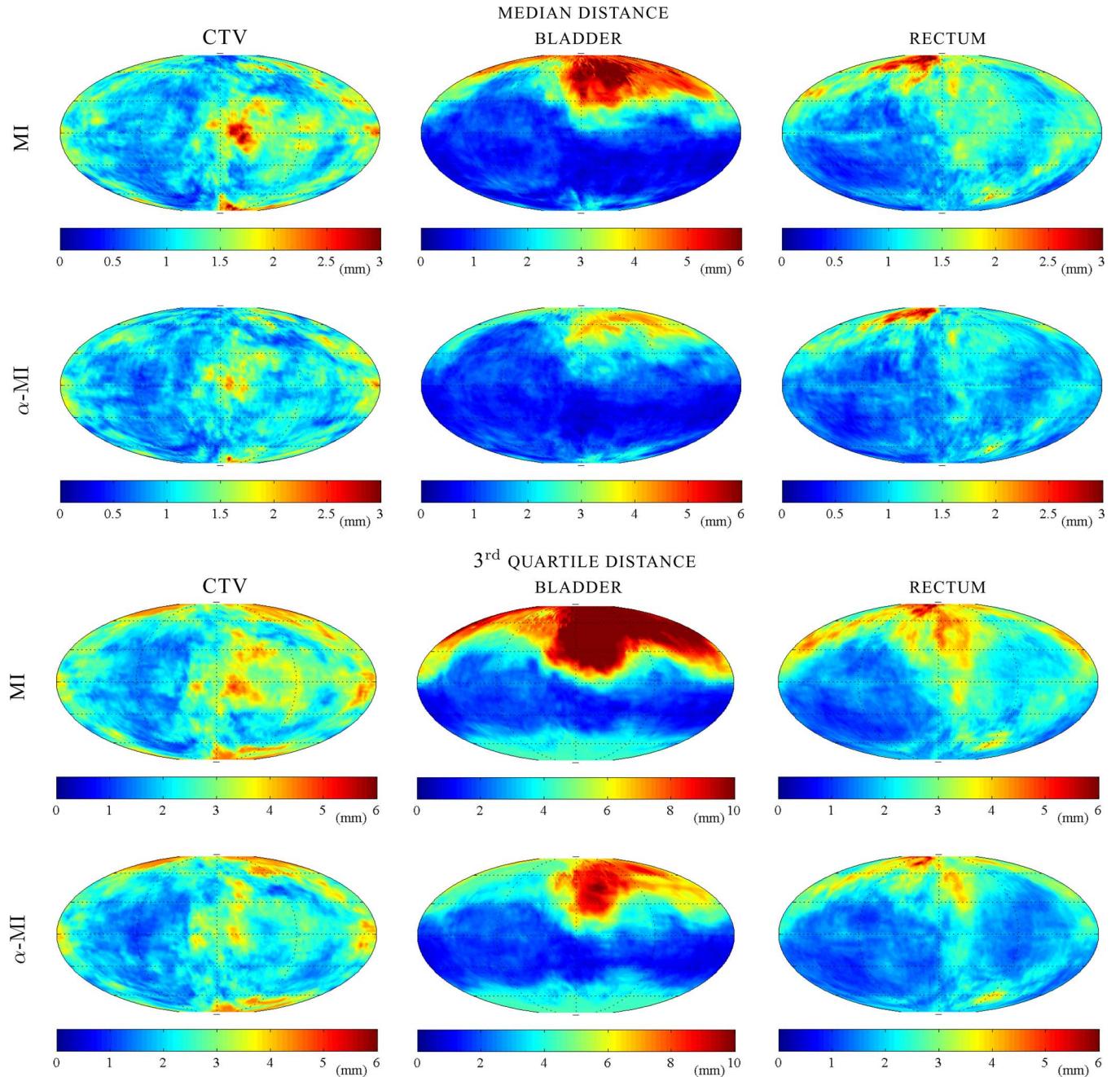


Fig. 8. The spatial distribution of the registration errors. The graphs at the top two rows show the median distances between the manual and automatic segmentation. The bottom two rows show the third quartile of the distances. Note that the graphs have different color scales.

over  $kN$  edges, see (15). During the experiments we noticed that most of the computation time is spent in calculating (14). The construction of the graph and searching for the  $k$  nearest neighbors appeared to be of minor influence. Therefore, although it takes more time to construct the MST, it is probably faster to compute the analytical derivative with it. The influence of this change on the rate of convergence of the registration remains to be investigated.

In conclusion, compared to a standard approach, the proposed method accomplishes improved nonrigid registration for a challenging registration problem, by inclusion of multiple features in the registration cost function.

## VII. ACKNOWLEDGEMENTS

The authors would like to thank L. van de Bunt, M.D. and G. A. P. de Kort, M.D. for providing the manual segmentations. This work also benefited from the use of the Insight Segmentation and Registration Toolkit (ITK), an open source software package developed as an initiative of the U.S. National Library of Medicine and available at <http://www.itk.org>.

## REFERENCES

- [1] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002," *Cancer J. Clinicians*, vol. 55, no. 2, pp. 74–108, 2005.

- [2] R. Pötter, C. Haie-Meder, E. V. Limbergen, I. Barillot, M. D. Brabandere, J. Dimopoulos, I. Dumas, B. Erickson, S. Lang, A. Nulens, P. Petrow, J. Rownd, and C. Kirisits, "Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology," *Radiotherapy Oncol.*, vol. 78, no. 1, pp. 67–77, 2006.
- [3] R. Pötter, J. Dimopoulos, P. Georg, S. Lang, C. Waldhäusl, N. Wachter-Gerstner, H. Weitmann, A. Reinthaller, T. H. Knocke, S. Wachter, and C. Kirisits, "Clinical impact of MRI assisted dose volume adaptation and dose escalation in brachytherapy of locally advanced cervix cancer," *Radiotherapy Oncol.*, vol. 83, no. 2, pp. 148–155, 2007.
- [4] L. van de Bunt, U. A. van de Heide, M. Ketelaars, G. A. de Kort, and I. M. Jürgenliemk-Schulz, "Conventional, conformal, and intensity-modulated radiation therapy treatment planning of external beam radiotherapy for cervical cancer: The impact of tumor regression," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 1, pp. 189–196, 2006.
- [5] L. van de Bunt, I. M. Jürgenliemk-Schulz, G. A. de Kort, J. M. Roesink, R. J. Tersteeg, and U. A. van de Heide, "Motion and deformation of the target volumes during IMRT for cervical cancer: What margins do we need?" *Radiotherapy Oncol.*, 2008.
- [6] M. R. Kaus, K. K. Brock, V. Pekar, L. A. Dawson, A. M. Nichol, and D. A. Jaffray, "Assessment of a model-based deformable image registration approach for radiation therapy planning," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 2, pp. 572–580, 2007.
- [7] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi, "Large deformation 3D image registration in image-guided radiation therapy," *Phys. Med. Biol.*, vol. 50, no. 24, pp. 5869–5892, 2005.
- [8] K. K. Brock, M. B. Sharpe, L. A. Dawson, S. M. Kim, and D. A. Jaffray, "Accuracy of finite element model-based multiorgan deformable image registration," *Med. Phys.*, vol. 32, no. 6, pp. 1647–1659, 2005.
- [9] K. K. Brock, L. A. Dawson, M. B. Sharpe, D. J. Moseley, and D. A. Jaffray, "Feasibility of a novel deformable image registration technique to facilitate classification, targeting, and monitoring of tumor and normal tissue," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 4, pp. 1245–1254, 2006.
- [10] S. Klein, U. A. van de Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med. Phys.*, vol. 35, no. 4, pp. 1407–1417, 2008.
- [11] K. S. C. Chao, S. Bhide, H. Chen, J. Asper, S. Bush, G. Franklin, V. Kavadi, V. Liengswangwong, W. Gordon, A. Raben, J. Strasser, C. Koprowski, S. Frank, G. Chronowski, A. Ahamad, R. Malyapa, L. Zhang, and L. Dong, "Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, no. 5, pp. 1512–1521, 2007.
- [12] K. K. Brock, D. J. Moseley, J. Stewart, K. Lim, P. Chan, R. Dinniwell, Y. Cho, A. Fyles, M. Milosevic, and D. A. Jaffray, "Deformable modeling of physiological motion and treatment response of cervical cancer patients," in *Int. Conf. Use Comput. Radiat. Therapy*, 2007.
- [13] C. López and S. Chakravarti, "Imaging of cervical cancer," *Imaging*, vol. 18, no. 1, pp. 10–19, 2006.
- [14] D. G. Mitchell, B. Snyder, F. Coakley, C. Reinhold, G. Thomas, M. Amendola, L. H. Schwartz, P. Woodward, H. Pannu, and H. Hricak, "Early invasive cervical cancer: Tumor delineation by magnetic resonance imaging, computed tomography, and clinical examination, verified by pathologic results in the ACRIN 6651/GOG 183 intergroup study," *J. Clin. Oncol.*, vol. 24, no. 36, pp. 5687–5693, 2006.
- [15] J. Dimopoulos, G. Schard, D. Berger, S. Lang, G. Goldner, T. Helbich, and R. Pötter, "Systematic evaluation of MRI findings in different stages of treatment of cervical cancer: Potential of MRI on delineation of target, pathoanatomic structures, and organs at risk," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 5, pp. 1380–1388, 2006.
- [16] A. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 85–95, Sep. 2002.
- [17] H. Neemuchwala and A. Hero, *Entropic Graphs for Registration*, ser. Signal Processing and Communications. Boca Raton, FL: CRC Press, 2005, ch. 6, pp. 185–235.
- [18] H. F. Neemuchwala, A. Hero, and P. Carson, "Image matching using alpha-entropy measures and entropic graphs," *Signal Process.*, vol. 85, no. 2, pp. 277–296, 2005.
- [19] H. F. Neemuchwala and A. O. Hero, "Image registration in high dimensional feature space," in *SPIE Electronic Imaging: Computational Imaging III*, ser. Proceedings of SPIE, C. A. Bouman and E. L. Miller, Eds. Bellingham, WA: SPIE Press, 2005, vol. 5674, pp. 99–113.
- [20] E. Oubel, A. F. Frangi, and A. Hero, "Complex wavelets for registration of tagged MRI sequences," in *IEEE Int. Symp. Biomed. Imag.*, 2006, pp. 622–625.
- [21] B. Ma, R. Narayanan, H. Park, A. O. Hero, P. H. Bland, and C. R. Meyer, "Comparing pairwise and simultaneous joint registrations of decorrelating interval exams using entropic graphs," in *Information Processing in Medical Imaging*. New York: Springer, 2007, vol. 4584, Lecture Notes in Computer Science, pp. 270–282.
- [22] M. R. Sabuncu and P. J. Ramadge, "Using spanning graphs for efficient image registration," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 788–797, May 2008.
- [23] E. Oubel, M. de Craene, M. Gazzola, A. Hero, and A. Frangi, "Multi-view registration of cardiac tagging MRI sequences," in *IEEE Int. Symp. Biomed. Imag.*, Arlington, VA, 2007, pp. 388–391.
- [24] A. Collignon, "Multi-modality medical image registration by maximization of mutual information," Ph.D. dissertation, Catholic Univ. Leuven, Leuven, Belgium, 1998.
- [25] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [26] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.
- [27] J. Beardwood, J. H. Halton, and J. M. Hammersley, "The shortest path through many points," in *Proc. Cambridge Phil. Soc.*, 1959, vol. 55, pp. 299–327.
- [28] A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal k-point random graphs," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1921–1939, Sep. 1999.
- [29] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, Berkeley, CA, 1961, vol. 1, pp. 547–561.
- [30] S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.
- [31] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Trans. Image Process.*, vol. 16, pp. 2879–2890, 2007.
- [32] L. Florack, "The syntactical structure of scalar images," Ph.D. dissertation, Univ. Utrecht, Utrecht, The Netherlands, 1993.
- [33] L. Ibáñez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide* Kitware, 2005, ISBN 1-930934-15-7.
- [34] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.
- [35] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [36] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.
- [37] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque, "Definition of the prostate in CT and MRI: A multi-observer study," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 43, no. 1, pp. 57–66, 1999.
- [38] D. Rueckert, M. J. Clarkson, D. L. G. Hill, D. J. Hawkes, M. Sonka, and J. M. F. Eds., "Non-rigid registration using higher-order mutual information," in *SPIE Medical Imaging: Image Processing*, ser. Proceedings of SPIE. San Diego, CA: SPIE Press, 2000, vol. 3979, pp. 438–447.
- [39] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, Jr., T. Pajdla and J. Matas, Eds., "Image similarity using mutual information of regions," in *Eur. Conf. Comput. Vis. (ECCV)*, Prague, Czech Republic, 2004, vol. 3023, pp. 596–607.
- [40] D. Tomažević, B. Likar, and F. Pernuš, "Multi-feature mutual information," in *SPIE Medical Imaging: Image Processing*, ser. Proceedings of SPIE, J. M. Fitzpatrick and M. Sonka, Eds. San Diego, CA: SPIE, 2004, vol. 5370, pp. 143–154.
- [41] J. Kybic, "High-dimensional entropy estimation for finite accuracy data: R-nn entropy estimator," in *Information Processing in Medical Imaging (IPMI)*, ser. Lecture Notes in Computer Science. Kerkrade, The Netherlands: Springer Verlag, 2007, vol. 4584, pp. 569–580.
- [42] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, pp. 066138–066138, 2004.
- [43] J. Kybic, "Incremental updating of nearest neighbor-based high-dimensional entropy estimation," in *IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, Toulouse, France, 2006, vol. 3, pp. III–804.
- [44] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.