Nonrigid Image Registration Using Conditional Mutual Information

Dirk Loeckx*, *Member, IEEE*, Pieter Slagmolen, Frederik Maes, Dirk Vandermeulen, and Paul Suetens, *Member, IEEE*

***Abstract*—Maximization of mutual information (MMI) is a popular similarity measure for medical image registration. Although its accuracy and robustness has been demonstrated for rigid body image registration, extending MMI to nonrigid image registration is not trivial and an active field of research. We propose conditional mutual information (cMI) as a new similarity measure for nonrigid image registration. cMI starts from a 3-D joint histogram incorporating, besides the intensity dimensions, also a spatial dimension expressing the location of the joint intensity pair. cMI is calculated as the expected value of the cMI between the image intensities given the spatial distribution. The cMI measure was incorporated in a tensor-product B-spline nonrigid registration method, using either a Parzen window or generalized partial volume kernel for histogram construction. cMI was compared to the classical global mutual information (gMI) approach in theoretical, phantom, and clinical settings. We show that cMI significantly outperforms gMI for all applications.**

***Index Terms*—B-splines, biomedical image processing, conditional mutual information, free-form deformation, image matching, image motion analysis, image processing, image registration, mutual information, nonrigid registration, spline functions, voxel-based similarity measure.**

I. Introduction

SINCE its introduction for medical image registration in 1995 [2], [3], mutual information (MI) has gained wide interest in the field [4]. MI is a basic concept from information theory that measures the amount of information one image contains about the other. Starting from a reference image R and floating image F with intensity bins r and f, the mutual information $I(R,F)$ is calculated from the joint and marginal probabilities $p(r,f), p(r) = \sum_f p(r,f)$, and $p(f) = \sum_r p(r,f)$

$$I(R,F) = H(R) + H(F) - H(R,F) \tag{1}$$

$$= \sum_{r\in R}\sum_{f\in F} p(r,f)\ln\left(\frac{p(r,f)}{p(r)p(f)}\right) \tag{2}$$

with $H(A,B) = -\sum_{a,b} p(a,b)\ln(p(a,b))$ and $H(A) = -\sum_a p(a)\ln(p(a))$ the joint and marginal entropy of random variables A and B.

The use of maximization of mutual information (MMI) as a registration criterion starts from the hypothesis that the images are correctly aligned when the MI between the intensities of corresponding voxels is maximal. Because MMI assumes only a statistical relationship between both images, without making any assumption about the nature of this relationship, image registration using MMI is widely applicable. MMI has been applied to a wide range of registration problems [5], covering monomodal as well as multimodal applications. Its accuracy and robustness has been demonstrated for rigid body image registration of computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET) brain images [6].

The calculation of MI is typically based upon a global joint histogram, expressing the joint intensity probabilities over the whole image. The underlying assumption is that the statistical relationship between both images is the same over the whole image domain. Often, this is only approximately true, e.g., when the images are distorted with a bias field or when structures with different intensities in one image have similar intensities in the other image, as e.g., bone and background in CT and MR.

When MI is applied to rigid and affine image registration, information from the whole image is taken into account to steer a limited set of parameters. Each of these parameters influences the transformation over the whole image field (e.g., scale, translation). Therefore, global image registration hardly suffers from local inconsistencies. Only strong bias fields are known to cause problems [4].

Extending MMI to nonrigid image registration is not trivial and an active field of research. A nonrigid transformation model, with a higher number of degrees-of-freedom, will be able to adapt for the local inconsistencies incurred using a global joint histogram. As we will show in this paper, global MI-based nonrigid registration might reduce smaller image details or have problems with spatially-varying intensity inhomogeneity due to a bias field.

These problems can be reduced using a local estimation of the joint histogram. This can be obtained by progressively subdividing the image and performing a set of local registrations [7]–[9]. However, when the subdivided image parts become too small, the small number of samples can limit the estimation

Manuscript received February 07, 2009; revised April 16, 2009; accepted April 17, 2009. First published May 12, 2009; current version published January 04, 2010. This work was partially supported by a grant from the K. U. Leuven Research Fund, from Varian Medical Systems and from the Research Foundation—Flanders (FWO). D. Loeckx is Postdoctoral Fellow of the Research Foundation—Flanders (FWO). This paper extends the work presented at the 2007 Information Processing in Medical Imaging Conference, Rolduc, The Netherlands. *Asterisk indicates corresponding author.*

*D. Loeckx is with the Group of Medical Image Computing, Center for Processing Speech and Images, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, 3000 Leuven, Belgium (e-mail: dirk.loeckx@uz.kuleuven.ac.be).

F. Maes, D. Vandermeulen, and P. Suetens are with the Group of Medical Image Computing, Center for Processing Speech and Images, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, 3000 Leuven, Belgium.

P. Slagmolen is with the Department of Radiation Oncology, Faculty of Medicine, Katholieke Universiteit Leuven, 3000 Leuven, Belgium.

Digital Object Identifier 10.1109/TMI.2009.2021843

0278-0062/$26.00 © 2009 IEEE

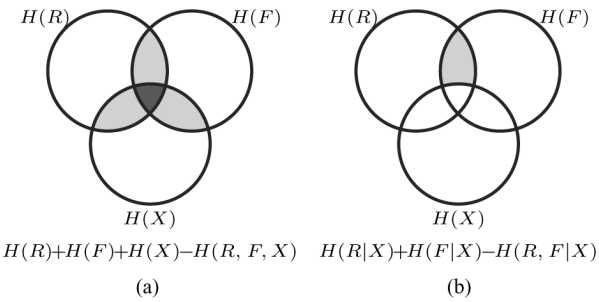

Fig. 1. Comparison of entropies included in the total correlation (a) and cMI (b). The dark color in (a) signifies that the area is counted twice.

performance of the local joint histogram. Several adaptations have been proposed to overcome this. Likar and Pernuš [8] combine the local and global intensity distribution. Andronache *et al.* [9] present a local intensity remapping to allow for the use of cross correlation as similarity measure in the smaller subimages. Weese *et al.* [10] argue that, when the image parts are sufficiently small, they will likely contain not more than two different structures that are rather homogeneous. Therefore, they propose to use cross correlation straightaway, without the need for an intensity remapping.

An alternative approach has been proposed by Studholme *et al.* [11]. They present a nonrigid viscous fluid registration scheme, using a similarity measure that is calculated over a set of overlapping subregions of the image. This is achieved by extending the intensity joint histogram with a third channel representing a spatial label. Several extensions of mutual information for more than two variables exist; Studholme *et al.* choose the total correlation (TC)

$$C(R,F,X) = H(R) + H(F) + H(X) - H(R,F,X) \quad (3)$$
$$= H(R) + H(F) - H(R,F\,|\,X) \quad (4)$$
$$= \sum_{r\in R}\sum_{f\in F}\sum_{\mathbf{x}\in X} p(r,f,\mathbf{x}) \ln\left(\frac{p(r,f,\mathbf{x})}{p(r)p(f)p(\mathbf{x})}\right) \quad (5)$$

as similarity measure (see Fig. 1), with X expressing the spatial position in the reference image. The image is subdivided into cubic regions of 10^3–50^3 voxels, overlapping 50% in each dimension (i.e., in 3-D, every voxel belongs to eight regions). The joint intensity histogram within each region is constructed using *B*-spline Parzen windows [12].

Within this paper, we propose conditional mutual information (cMI) [13] as similarity measure. The cMI $I(R,F\,|\,X)$ is calculated between the reference and floating intensity distributions R and F, given a certain spatial distribution X

$$I(R,F\,|\,X) = H(R\,|\,X) + H(F\,|\,X) - H(R,F\,|\,X) \quad (6)$$
$$= \sum_{\mathbf{x}\in X} p(\mathbf{x}) \sum_{r\in R}\sum_{f\in F} p(r,f\,|\,\mathbf{x}) \ln\left(\frac{p(r,f\,|\,\mathbf{x})}{p(r\,|\,\mathbf{x})p(f\,|\,\mathbf{x})}\right). \quad (7)$$

Whereas the bi-variate gMI $I(R,F)$ expresses the reduction in the uncertainty of R due to the knowledge of F (and vice–versa), cMI expresses the reduction in the uncertainty of R due to knowledge of F (and vice–versa) when the spatial location X is known[1].

Total correlation calculates the sum of the pairwise mutual information between different channels. Thus, as shown in Fig. 1 it includes not only the information shared by (R,F), but also the information shared by (R,X) and (F,X). cMI calculates the information that is shared between (R,F) but not available in X. We believe cMI corresponds to the actual situation in medical image registration, where the spatial location in the reference image of each joint intensity pair is indeed known *a priori*. The transformation will only alter the floating intensities in each joint intensity pair.

Similar to [11], we calculate cMI by extending the joint histogram with a third dimension representing the spatial distribution of the joint intensities. We incorporate cMI in a tensor-product *B*-spline registration algorithm [14], using the same kernel to calculate the weighted average of the *B*-spline deformation coefficients and for the spatial distribution of the joint intensities over the joint histogram. Thus, the similarity measure is calculated at the same scale as the transformation field.

We extensively compare cMI, gMI, and TC similarity measures for image registration. We first show that the proposed method works on a theoretical example and on artificial images representing either a CT/MR registration or a registration distorted by a bias field. Next, we start from a set of clinical images and the BrainWeb database [15] to create a validation dataset containing realistic transformation fields with a known ground-truth. Finally, we use a real clinical problem of CT/MR registration. In this case, the ground truth transformation is unknown. For the validation, we resort to indirect measures such as the Dice similarity coefficient and surface distance between manually delineated corresponding regions in both images.

In the next section, we give a detailed explanation of our method. In Section III, the validation results are presented. Section IV discusses the proposed method and results. We conclude in Section V.

II. Methods

Consider a reference image R with intensity $\mathcal{I}_R(\mathbf{x}_R)$, a floating image F with intensity $\mathcal{I}_F(\mathbf{x}_F)$ and a transformation $\mathbf{g}(\mathbf{x}_R;\mu)$ that maps every reference position $\mathbf{x}_R$ to the corresponding floating position $\mathbf{x}_F = \mathbf{g}(\mathbf{x}_R;\mu)$ given a set of parameters μ. In the next paragraphs we first describe the transformation model $\mathbf{g}(\mathbf{x}_R;\mu)$ that defines the space in which the best solution is sought and then the similarity measure that calculates how well the reference and deformed floating image match. Next, we give some details about the spatial resolution at which both the transformation model and proposed similarity measure are calculated. To conclude, we describe the calculation of the derivatives and the optimization and discuss the computational complexity of the algorithm.

[1]Although total correlation (5) and cMI (6) are clearly different, (7) is equivalent to (11) in [11]. We think this is due to a miscalculation of Studholme *et al.*: in [11, Eq. (6) and (7)], the authors state that $p(m_1) = p(m_1\,|\,r)p(r)$ and $p(m_2) = p(m_2\,|\,r)p(r)$. However, this should be $p(m_1) = \sum_{r\in R} p(m_1\,|\,r)p(r)$ and analogous. Note that Studholme uses a slightly different notation, i.e., $p_r(m_1) \equiv p(m_1\,|\,r)$.

A. Transformation Model

Several transformation models have been proposed for nonrigid image registration. We adopt a tensor-product *B*-spline model, as proposed by Rueckert *et al.* [14]. The *B*-spline model is situated between a global rigid registration model and a local nonrigid model at voxel-scale. Its locality or nonrigidity can be adopted to a specific registration problem by varying the mesh spacing and thus the number of degrees-of-freedom.

The 3-D transformation field is given by

$$\mathbf{g}(\mathbf{x}_R;\mu) = \sum_{ijk} \mu_{ijk} \beta^{n_x}_{\Delta_x}\left(x - k^x_i\right) \beta^{n_y}_{\Delta_y}\left(y - k^y_j\right) \times \beta^{n_z}_{\Delta_z}\left(z - k^z_k\right) \quad (8)$$

with $\Delta_{x,y,z}$ the mesh spacing and $n_{x,y,z}$ the *B*-spline degree. The transformation is governed by the displacement vectors $\boldsymbol{\mu}_{ijk}$ associated with the tensor-product knots $\mathbf{k}_{ijk} = (k^x_i, k^y_j, k^z_k)$. Note that the transformation parameters μ are 3-D vectors, and thus (8) models a different function for each dimension: $g_x(\mathbf{x}_R;\boldsymbol{\mu}^x), g_y(\mathbf{x}_R;\boldsymbol{\mu}^y)$, and $g_z(\mathbf{x}_R;\boldsymbol{\mu}^z)$.

B. Similarity Measure

We describe three different similarity measures: standard or global mutual information (gMI), conditional mutual information (cMI), and total correlation (TC). Each similarity measure is based upon the (conditional) joint histogram, which is estimated using either Parzen window (PW) interpolation [12] or generalized partial volume (PV) estimation [16], [17]. We will first detail the formulation of the different similarity measures using Parzen window interpolation. Next the equivalent formulas for the partial volume approach are given.

1) Global Mutual Information: To calculate the gMI between R and F, we start from the joint histogram $\mathcal{H}(r,f;\mu)$ using a set of reference and floating bin centres r and f. The Parzen window joint histogram is given by

$$\mathcal{H}^{\mathrm{pw}}(r,f;\mu) = \sum_{\mathbf{x}_R \in R} w_r(\mathcal{I}_R(\mathbf{x}_R) - r) w_f(\mathcal{I}_F(\mathbf{g}(\mathbf{x}_R;\mu)) - f) \quad (9)$$

where w_r and w_f are the Parzen window kernels used to distribute an intensity over the neighboring bins. Because of their attractive mathematical properties, we have chosen to use *B*-splines for the Parzen window kernels $w_r(x) = w_f(x) = \beta^{n_{\mathrm{pw}}}_{\epsilon}(x)$ with ϵ the (constant) spacing between neighboring bins and $\beta^n_\epsilon(x) = \beta^n(x/\epsilon)/\epsilon$ the expanded *B*-spline of degree n. Throughout this paper, second degree ($n_{\mathrm{pw}} = 2$) *B*-splines are used to construct the histogram. Thus, each joint intensity pair contributes to a 3×3 region in the joint histogram. As the floating intensity $\mathcal{I}_F(\mathbf{x}_F)$ is sought at a noninteger position $\mathbf{x}_F = \mathbf{g}(\mathbf{x}_R;\mu)$, it is obtained using n_ith degree *B*-spline intensity interpolation, choosing again $n_i = 2$.

The joint probability distribution can now be calculated as

$$p(r,f;\mu) = \frac{\mathcal{H}^{\mathrm{pw}}(r,f;\mu)}{\sum_{r,f} \mathcal{H}^{\mathrm{pw}}(r,f;\mu)} \quad (10)$$

which immediately leads to the mutual information as described in (2).

2) Conditional Mutual Information: To extend the joint histogram with spatial information, we overlay a regular lattice with knots $\mathbf{x}_{ijk} = (x^x_i, x^y_j, x^z_k)$ over the reference image. The joint histogram (9) is extended with a spatial kernel $w_{\mathbf{x}}$

$$\mathcal{H}^{\mathrm{pw}}_s(r,f,\mathbf{x};\mu) = \sum_{\mathbf{x}_R \in R} w_r(\mathcal{I}_R(\mathbf{x}_R) - r) \times w_f(\mathcal{I}_F(\mathbf{g}(\mathbf{x}_R;\mu)) - f) w_{\mathbf{x}}(\mathbf{x}_R - \mathbf{x}) \quad (11)$$

with $\mathbf{x} \in \{\mathbf{x}_{ijk}\}$. $\mathbf{x}$ can be considered as spatial bins, as their role in the above equation is analogous to the role of r and f. Using $n'_{x,y,z}$th degree *B*-spline kernels for the spatial kernel in each dimension, and mesh spacing $\Delta'_{x,y,z}, w_{\mathbf{x}}$ is given by

$$w_{\mathbf{x}}(\mathbf{x}_R - \mathbf{x}) = \beta^{n'_x}_{\Delta'_x}(x_{R,x} - x_x) \beta^{n'_y}_{\Delta'_y} \times (x_{R,y} - x_y) \beta^{n'_z}_{\Delta'_z}(x_{R,z} - x_z) \quad (12)$$

with $\mathbf{x}_R = (x_{R,x}, x_{R,y}, x_{R,z})$ and $\mathbf{x} = (x_x, x_y, x_z)$.

Starting from (11), the conditional probability can be obtained similar to (10), replacing $p(r,f;\mu)$ with

$$p(r,f \mid \mathbf{x};\mu) = \frac{\mathcal{H}^{\mathrm{pw}}_s(r,f,\mathbf{x};\mu)}{p(\mathbf{x};\mu)} \quad (13)$$

using

$$p(\mathbf{x};\mu) = \frac{\sum_{r,f} \mathcal{H}^{\mathrm{pw}}_s(r,f,\mathbf{x})}{\sum_{r,f,\chi} \mathcal{H}^{\mathrm{pw}}_s(r,f,\chi)}. \quad (14)$$

Finally, the above equations can be substituted in (7) to obtain the cMI.

3) Total Correlation: As can be seen from (5), the total correlation is a mixture of global marginal entropies and conditional local entropies. Therefore, it can be easily calculated combining the formulas mentioned above.

4) Generalized Partial Volume: Compared to the Parzen window approach, the generalized partial volume approach [17] does not interpolate the floating image at the target position. Instead, it partially hits the histogram with the intensities in the floating image surrounding the target position, weighted with a *B*-spline kernel. Thus, (9) becomes

$$\mathcal{H}^{\mathrm{pv}}(r,f;\mu) = \sum_{\mathbf{x}_R \in R} \sum_{\mathbf{x}_F \in F} w_r(\mathcal{I}_R(\mathbf{x}_R) - r) \times w_f(\mathcal{I}_F(\mathbf{x}_F) - f) w_v(\mathbf{x}_F; \mathbf{g}(\mathbf{x}_R;\mu)) \quad (15)$$

with w_v the partial volume kernel defining the neighborhood surrounding the target position to include. For (15) to be differentiable with respect to the registration parameters μ, it is sufficient that w_v is differentiable with respect to μ; hence, we can choose nearest-neighbor interpolation for the kernels w_r and w_f. As w_v is spatially limited, the sum over the floating positions can be reduced to the nonzero values of the volume kernel. The partial volume kernel is chosen equivalent to the interpolation kernel above, thus we adopt a multidimensional *B*-spline

kernel, i.e., the product of a simple n_ith degree *B*-spline kernel in each dimension.

The conditional joint histogram (11) becomes

$$\mathcal{H}_s^{\mathrm{pv}}(r,f,\mathbf{x};\mu) = \sum_{\mathbf{x}_R\in R}\sum_{\mathbf{x}_F\in F} w_r(\mathcal{I}_R(\mathbf{x}_R)-r) \times w_f(\mathcal{I}_F(\mathbf{x}_F)-f)w_v(\mathbf{x}_F;\mathbf{g}(\mathbf{x}_R;\mu))w_{\mathbf{x}}(\mathbf{x}_R-\mathbf{x}). \quad (16)$$

C. Spatial Resolution

Equations (8) and (12) both stipulate the spatial resolution of the algorithm. The former governs the region of influence of a registration parameter and the latter the scale at which the cMI and thus similarity measure is calculated. We will use the same settings for the *B*-spline degree, mesh knots, and spacing in both formulas. Thus, the local transformation around a certain displacement vector is guided by the conditional joint histogram, both using the same concept and scale of locality. Therefore, we also keep the *B*-spline degree $n_{x,y,z} = n'_{x,y,z}$ fixed over a single registration: using *B*-spline subdivision formula's, (8) can be refined exactly for constant n when $\Delta_{x,y,z} = \Delta'_{x,y,z}$ is divided by two.

The extent of the image region contributing to the joint histogram for a given spatial bin $\mathbf{x}$ is the same as the region of influence of the corresponding displacement vector $\boldsymbol{\mu}_{\mathbf{x}}$. Due to the limited span properties of *B*-splines, this extent is limited to a $(n_x+1)\Delta_x$ by $(n_y+1)\Delta_y$ by $(n_z+1)\Delta_z$ cuboid centred around $\mathbf{x}$, with a higher contribution for the voxels closer to $\mathbf{x}$. The mesh spacing $\Delta_{x,y,z}$ regulates the locality of the cMI. To represent a more global histogram, a large $\Delta_{x,y,z}$ is chosen, yielding a limited number of spatial bins or displacement vectors containing contributions from a large image area. A fine mesh with small $\Delta_{x,y,z}$ and many spatial bins is used to calculate a more local histogram and transformation field. In practice, a multiresolution scheme is adopted, starting from a coarse mesh and gradually refining it. For most applications we choose $\Delta_x = \Delta_y = \Delta_z$, although we deviate from this for non-isotropic data.

D. Derivatives and Optimization

A limited memory quasi Newton method [18] is adopted for the optimization, using analytical derivatives to avoid discretization errors. The derivative of (11) with respect to a transformation parameter μ_{ijk}^{ξ} is given by

$$\frac{\partial\mathcal{H}^{\mathrm{pw}}(r,f,\mathbf{x};\mu)}{\partial\mu_{ijk}^{\xi}} = \sum_{\mathbf{x}_R\in R} w_r(\mathcal{I}_r(\mathbf{x}_R)-r) \times\frac{\partial w_f(\mathcal{I}_F(\mathbf{g}(\mathbf{x}_R;\mu))-f)}{\partial\mu_{ijk}^{\xi}} w_x(\mathbf{x}_R-\mathbf{x}) \quad (17)$$

and similar for gMI. The central factor is calculated using

$$\frac{\partial w_f(\mathcal{I}_F(\mathbf{g}(\mathbf{x}_R;\mu))-f)}{\partial\mu_{ijk}^{\xi}} = \left.\frac{\partial w_f(i-f)}{\partial i}\right|_{i=\mathcal{I}_F(\mathbf{g}(\mathbf{x}_R;\mu))} \left.\frac{\partial\mathcal{I}_F(\mathbf{r})}{\partial\mathbf{r}}\right|^T_{\mathbf{r}=\mathbf{g}(\mathbf{x}_R;\mu)} \frac{\partial\mathbf{g}(\mathbf{x}_R;\mu)}{\partial\mu_{ijk}^{\xi}} \quad (18)$$

and

$$\frac{\partial\mathbf{g}(\mathbf{x};\mu)}{\partial\mu_{ijk}^{\xi}} = [\delta_{\xi x},\delta_{\xi y},\delta_{\xi z}]^T \beta_{\Delta_x}^{n_i}(x-k_i^x)\,\beta_{\Delta_y}^{n_i}\left(y-k_j^y\right)\beta_{\Delta_z}^{n_i}(z-k_k^z). \quad (19)$$

The image derivative $\partial\mathcal{I}_f(\mathbf{r})/\partial\mathbf{r}$ depends on the image interpolation scheme. We use n_ith degree *B*-spline image interpolation throughout this work, thus the calculation of the image derivative requires only *B*-spline derivatives.

For PV, the derivative of (16) with respect to a transformation parameter μ_{ijk}^{ξ} is given by

$$\frac{\partial\mathcal{H}_s^{\mathrm{pv}}(r,f,\mathbf{x};\mu)}{\partial\mu_{ijk}^{\xi}} = \sum_{\mathbf{x}_R\in R}\sum_{\mathbf{x}_F\in F} w_r(\mathcal{I}_r(\mathbf{x}_R)-r) \times w_f(\mathcal{I}_F(\mathbf{x}_F)-f)\frac{\partial w_v(\mathbf{x}_F;\mathbf{g}(\mathbf{x}_R;\mu))}{\partial\mu_{ijk}^{\xi}} w_{\mathbf{x}}(\mathbf{x}_R-\mathbf{x}) \quad (20)$$

requiring no image derivative.

E. Computational Complexity

Taking the limited-span properties of the *B*-splines into account, the construction of the PW joint histogram (9) for 3-D images requires $v\times 2(n_{\mathrm{pw}}+1)\times 3(n_i+1)$ *B*-spline evaluations per iteration, with v the number of voxels, $2(n_{\mathrm{pw}}+1)$ accounting for the Parzen window histogram update and $3(n_i+1)$ for the floating image interpolation. The *B*-spline evaluations needed to calculate the transformation field at each voxel are performed only once for each multiresolution stage, which is possible if the mesh spacing $\Delta_{x,y,z}$ equals an integer number of voxels. For the PV histogram, the image interpolation is replaced by the partial volume kernel and the Parzen window histogramming is no longer used, leading to $v\times 3(n_i+1)$ *B*-spline evaluations. Note that the number of evaluations is independent of the mesh spacing. The number of evaluations does not change for cMI, as the spatial kernel for cMI is the same as used to calculate the transformation field. If we take the same *B*-spline degree for all aspects of the method, the ratios of computational complexity for $n=1,2,3$ are $1:2.25:4$ for PW and $1:1.5:2$ for PV.

However, the number of multiplications required for cMI will increase, as a different joint histogram has to be constructed for each mesh control point μ_{ijk}. Whereas each joint intensity pair contributes only to 1 (i.e., the global) joint histogram for gMI, it will contribute to $(n_x+1)(n_y+1)(n_z+1)$ joint histograms for cMI. This factor amounts to 8, 27, and 64 for $n=1,2,3$, respectively. The number of joint histogram derivatives will be multiplied with the same factor. Moreover, for gMI, the memory requirement for the joint histogram and its derivatives is $b_r\times b_f\times(\#\mu+1)$ floating-point numbers, with b_r and b_f the number of reference and floating bins and $\#\mu$ the number of displacement vectors. For cMI, each spatial bin is influenced by $(2n_x+1)(2n_y+1)(2n_z+1)$ displacement vectors, thus we need $b_r\times b_f\times((2n_x+1)(2n_y+1)(2n_z+1)+1)\times\#\mu$ floating-point numbers to store the conditional joint histograms and their derivatives, leading to excessive memory requirements. Luckily, we do not need to store all cMI joint

histograms simultaneously in memory. In theory, we could calculate each joint histogram independently, requiring only $b_r \times b_f \times ((2n_x+1)(2n_y+1)(2n_z+1)+1)$ floating-point numbers in memory at the cost of extra calculation time as we will have to recalculate the transformation field and re-interpolate the floating image for each joint histogram. In practice, we run over the voxels of the reference image, and create new cMI joint histogram and joint histogram derivative structures as long as we have sufficient memory. When we run out of memory, we stop allocating memory and perform a second run to calculate the cMI for the skipped spatial bins. If necessary, further runs are performed as well. This way, we limit the number of cMI joint histograms calculated simultaneously at the cost of extra transformation field calculations and image interpolations.

To conclude, the calculation time ratio between conditional and global MI will be somewhere between 1 if the execution time is determined by *B*-spline evaluations and without memory limits and 27 or 64 if we have to calculate all cMI joint histograms independently. In practice, the number of iterations depends on the method used, also influencing the calculation time.

III. Experiments and Results

Accurate validation of nonrigid registration requires a set of reference and floating image pairs and, for each pair, the ground truth deformation. A registration is performed on each image pair. Comparison of the obtained deformation fields with the ground truth yields an estimate of the accuracy of the registration algorithm. However, ground truth deformations for nonrigid registration are hard or impossible to obtain for real clinical cases.

In this section, we used several experiments to compare cMI, gMI, and TC, ranging from theoretical over phantom to clinical data. First, we used a theoretical example to illustrate mathematically the difference between the different similarity measures. In the second and third experiment, we applied a random transformation to pairs of artificial images, representing multimodal or bias field registration. In the fourth experiment, we deduced realistic artificial deformation fields and corresponding images starting from the BrainWeb [15] database and real patient images, similar to our work performed in [19]. Finally, we evaluated the performance of the proposed registration algorithms on clinical images used for radiotherapy planning. As in this case the ground-truth transformation is unknown, we used overlap measures on manual segmentations for the validation.

A. Experiment 1: Theoretical Example

Before we start the actual experiments, we calculate gMI, cMI, and TC on a parameterised image pair. Starting from two square images picturing two perfectly aligned concentric circles, as shown in Fig. 2, we apply a perturbation to the area A_3^F of the inner circle S_3^F in the floating image. See the Appendix for the mathematical details.

The influence of the perturbation on the similarity measures is shown in Fig. 3(a). For cMI and the $H(R,F\,|\,X)$ term in TC, we calculate the conditional entropy only in the central region, as the conditional entropy of other regions remains constant. All three measures have a maximum for $A_3^F = A_3^R$. However, the maximum of gMI and TC is only local. These measures

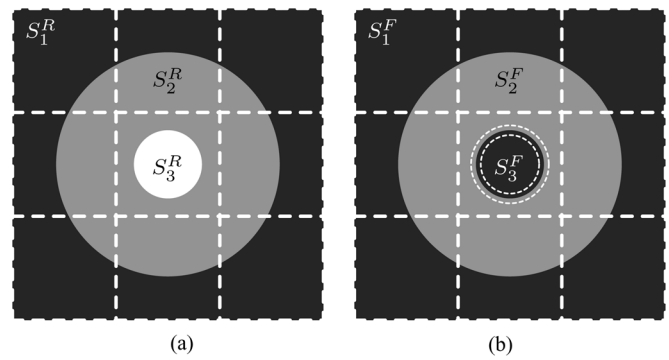

Fig. 2. Images used in the theoretical example. Starting from aligned images, we apply a perturbation to the area of the inner circle S_3^F in the floating image. (a) Reference. (b) Floating.

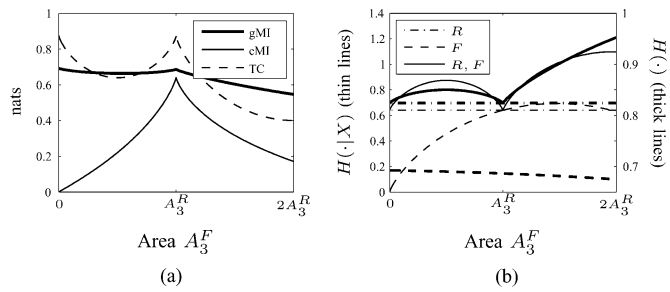

Fig. 3. Evolution of (a) the MI and (b) the joint and marginal entropies for a perturbation on the area A_3^F of S_3^F. cMI shows a global maximum for $A_3^F = A_3^R$, whereas the gMI and TC maximum are only local. (a) Mutual Information. (b) Entropy.

reach their global maximum when $A_3^F = 0$. cMI, on the other hand, has its global maximum as expected when the inner circles are aligned. This can be explained using the entropy graphs in Fig. 3(b). The global and conditional joint entropy show a local optimum for both $A_3^F = 0$ and $A_3^F = A_3^R$. However, for cMI, the optimum at $A_3^F = 0$ is more than compensated by the strong decrease in $H(F\,|\,X)$ when going to zero. For gMI and TC, the optimum at zero is the global optimum, as $H(F)$ slightly increases when going to zero.

B. Experiment 2: Multimodal Registration

For the second experiment, we created 200 artificial 2-D multimodality image pairs, as shown in Fig. 4. They are inspired by a slice through a lower limb in CT and MR, picturing background, muscle, and bone. Each image, measuring 256×256 pixels, consists of a dark background $(I = 0)$ and two concentric polygons. The larger polygon is a hexagon with a medium intensity $(I = 200)$, representing soft tissue. The smaller polygon is a pentagon, high intense $(I = 400)$ in the CT image and dark $(I = 0)$ in the MR image, representing bone. We have added uniform Gaussian noise $(\sigma = 50)$ to the images. A mesh spacing of $\Delta_x = \Delta_y = 32$ voxels was used for the *B*-spline transformation and the conditional joint histogram estimation; 30 bins were used in the joint histogram for the floating and the reference intensity. The experiment was performed using either second or third degree *B*-splines for the mesh and spatial kernel.

For each image pair, the transformation field was initialised choosing μ from a uniform distribution with a maximum

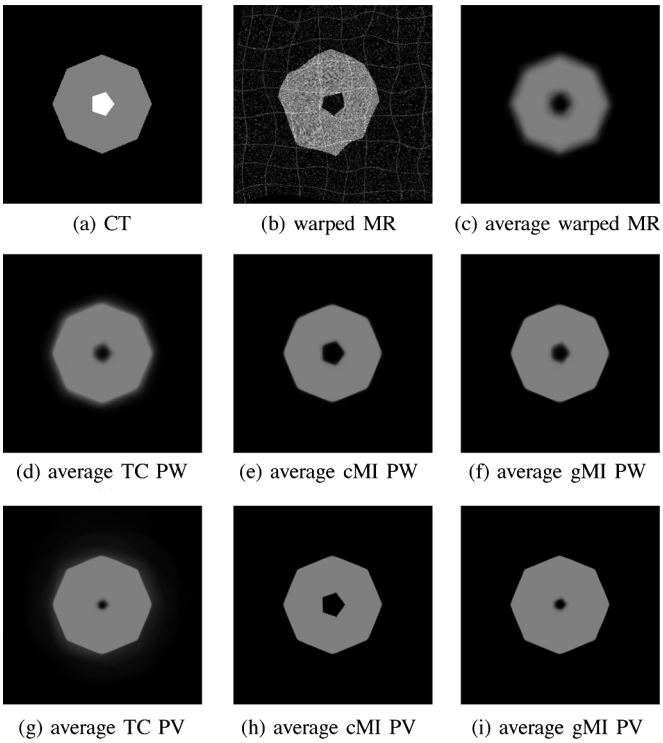

Fig. 4. Original images and average of the results obtained over all 200 multimodal registrations using second degree *B*-splines. The best results (sharpest average image) are obtained using conditional MI (cMI) and Partial volume (PV) interpolation.

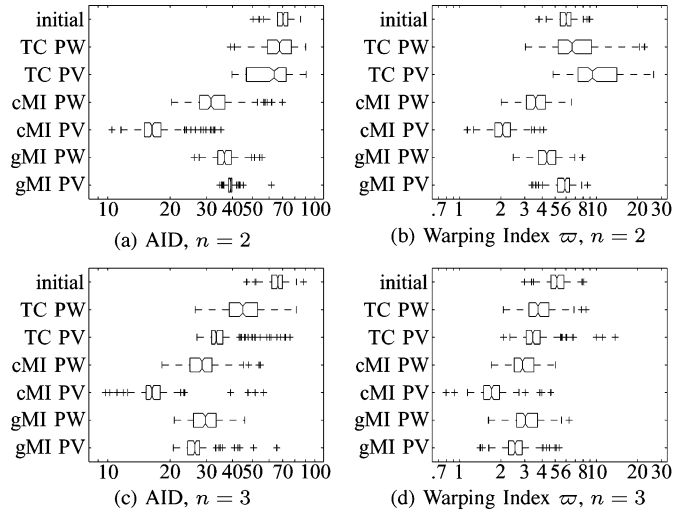

Fig. 5. Boxplot of the registration results obtained for 200 artificial multimodality image pairs. The graphs show the average intensity difference (AID) and warping index ϖ (in voxels) of the registration result compared to the ground truth. Registrations were performed using global MI (gMI), conditional MI, and total correlation (TC), using Parzen window (PW) and partial volume (PV) histogram estimation, and using *B*-spline degree $n = 2$ and $n = 3$.

TABLE I
RESULTS OF THE EXPERIMENTS SHOWN IN FIG. 5

		AID	ϖ	time (s)
$n = 2$	gMI PV	39.16 ± 2.10	5.80 ± 0.88	2.36 ± 0.45
	gMI PW	37.13 ± 4.71	4.45 ± 0.94	2.43 ± 0.61
	cMI PV	**17.17 ± 3.81**	**2.12 ± 0.46**	11.94 ± 2.16
	cMI PW	33.53 ± 8.32	3.69 ± 0.83	13.13 ± 4.13
	TC PV	61.59 ± 13.95	10.76 ± 4.23	20.24 ± 14.25
	TC PW	68.04 ± 11.08	7.87 ± 3.80	11.33 ± 4.62
	initial	70.06 ± 6.30	6.06 ± 0.89	
$n = 3$	gMI PV	26.83 ± 4.63	2.61 ± 0.61	2.54 ± 0.53
	gMI PW	30.34 ± 5.48	3.15 ± 0.77	3.26 ± 0.91
	cMI PV	**17.25 ± 5.09**	**1.77 ± 0.46**	34.66 ± 6.74
	cMI PW	29.40 ± 6.22	3.00 ± 0.68	42.37 ± 13.75
	TC PV	35.56 ± 7.84	3.68 ± 1.29	41.14 ± 16.27
	TC PW	46.45 ± 10.57	3.95 ± 1.03	42.96 ± 18.63
	initial	65.80 ± 6.17	5.23 ± 0.89	

amplitude of 30 pixels. Starting from this transformation, the MR images were registered to the CT image. The registration quality was measured as the average intensity difference (AID) between the original (noise-free) CT image and the final transformation applied to this image. We also calculated the warping index ϖ, which is the root mean square of the local registration error in each voxel. Note that, even for a perfect registration (according to the similarity measure), the final transformation might contain nonzero components within homogeneous regions. Although this has no influence on the final transformed image, it leads to an increased warping index. Both validation measures were calculated over a region of interest 20% larger than the outer polygon.

Fig. 4 shows the original images, the average of the transformed CT images and the averages of the registered images obtained using the different methods and second degree *B*-splines. Sharp average images indicate accurate registration results. The best results are obtained using cMI and PV interpolation. The average image of the registrations using cMI combined with PW has smoother pentagon and—to a lesser extent—octagon corners. The gMI and TC registrations show a clear shrinkage of the pentagon, whereas the octagon is again sharper in the average image using PV interpolation. The worst results are obtained using TC and PW. The shrinkage could be expected from Section III-A. Similar results have been obtained for third degree *B*-splines.

The AID and warping index using second and third degree *B*-splines are summarised in Fig. 5 and Table I; the table also contains the calculation time. Within both the second and third degree *B*-splines experiments, cMI PV is significantly better $(p < 0.01)$ then all other methods for the AID and warping index. Still for cMI PV, the difference in AID between $n = 2$ and $n = 3$ is not significant $(p = 0.86)$ although a significant difference is found in the warping index. The latter can probably be explained by the smoother deformation field for $n = 3$. Calculations take about 3 times longer for $n = 3$ compared to $n = 2$ for cMI.

C. Experiment 3: Bias Field Registration

In the third experiment, we want to compare the robustness of the different similarity measures to a bias field. To generate the pairs of images, we start from the Lena image (8 bit, 256 × 256 voxels). It was used unmodified for the reference image, for the floating images it was distorted with a second-degree multiplicative bias field $ax^2 + bxy + cx + dy^2 + e$, with a to e uniformly sampled from $\{[-0.75, -0.25], [0.25, 0.75]\}$. Next, the floating image was deformed and registered and the results were validated similar to the previous section.

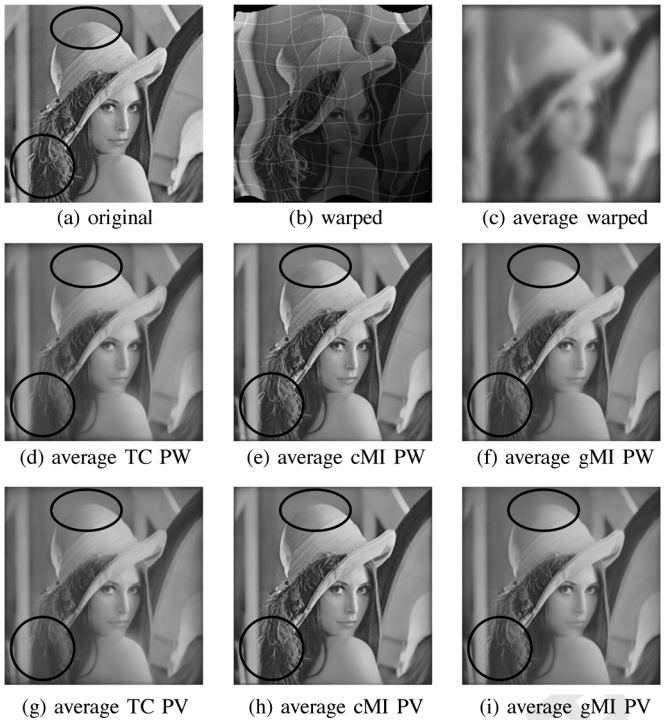

Fig. 6. Average of the results obtained over all 200 registrations after bias field distortion.

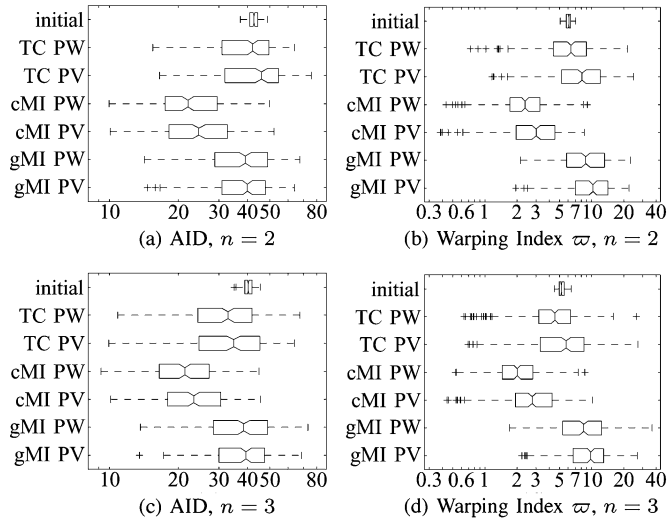

Fig. 7. Boxplot of the registration results obtained for 200 bias field distortion pairs. The graphs show the average intensity difference (AID) and warping index ϖ (in voxels) of the registration result compared to the ground truth. Registrations were performed using global MI (gMI), conditional MI, and total correlation (TC), using Parzen window (PW) and partial volume (PV) histogram estimation, and using *B*-spline degree $n = 2$ and $n = 3$.

TABLE II
RESULTS OF THE EXPERIMENTS SHOWN IN FIG. 7

		AID	ϖ	time (s)
n=2	gMI PV	39.61 ± 11.82	10.64 ± 4.80	6.41 ± 2.43
	gMI PW	38.73 ± 12.74	9.69 ± 5.01	7.52 ± 3.05
	cMI PV	25.38 ± 9.00	3.27 ± 1.82	40.17 ± 11.61
	cMI PW	**23.93 ± 8.39**	**2.58 ± 1.40**	51.60 ± 13.51
	TC PV	44.13 ± 13.60	9.38 ± 5.58	26.43 ± 12.86
	TC PW	40.60 ± 12.21	7.08 ± 4.06	30.74 ± 18.10
	initial	42.63 ± 2.29	6.10 ± 0.39	
n=3	gMI PV	39.16 ± 11.82	10.27 ± 4.91	8.48 ± 3.08
	gMI PW	38.54 ± 12.76	9.31 ± 5.24	10.56 ± 4.08
	cMI PV	24.29 ± 8.39	3.25 ± 1.97	113.56 ± 34.48
	cMI PW	**22.21 ± 7.62**	**2.36 ± 1.44**	158.30 ± 37.30
	TC PV	35.55 ± 13.02	6.57 ± 4.52	77.03 ± 31.86
	TC PW	33.37 ± 11.52	5.00 ± 3.33	104.69 ± 47.69
	initial	40.34 ± 2.09	5.27 ± 0.44	

The original Lena image, an example of a warped and bias field distorted image, and the average images after transformation and after registration are shown in Fig. 6. The average images obtained using cMI are clearly sharper than the ones obtained using gMI, indicating a more accurate registration. This can be seen e.g., at the top of the hat, in the details of the hair, but also in the face. Similar results have been obtained for third degree *B*-splines. The AID, warping index, and calculation time using second and third degree *B*-splines are summarised in Fig. 7 and Table II.

For $n = 2$, both cMI PV and cMI PW are significantly better than all other methods for the AID as well as for the warping index. cMI PW is significantly better than cMI PV for the warping index, whereas for the AID $p = 0.096$. Also for $n = 3$, cMI PV and cMI PW significantly outperform all the others, with cMI PW being significantly better than cMI PV for both criteria. Comparing $n = 2$ and $n = 3$ shows no significant difference between both cMI PV or cMI PW results. The $n = 3$ calculation time for conditional measures is almost 3 times longer than the time for $n = 2$.

D. Experiment 4: BrainWeb

For the next experiment, we move to clinical images and realistic deformations. The images are obtained from the BrainWeb database [15]. This database consists of simulated t1-weighted (T1), t2-weighted (T2), and proton density (PD) brain MR images of $181 \times 217 \times 181$ voxels with a 1 mm voxel size in each dimension, calculated from a single phantom. Therefore, the images are *a priori* in perfect alignment. BrainWeb also provides multiplicative bias fields that can be applied to the images. To obtain a set of realistic deformation fields, we start from nine patient T1 brain images with dimension $256 \times 182 \times 256$ and various voxel sizes of about $1 \times 1.2 \times 1$ mm^3. The BrainWeb T1 image is registered to each patient T1 image. The obtained deformations are applied to the BrainWeb T1 and PD images to produce deformed $\mathrm{T1}'$ and PD' images for each patient. Prior to deformation, a bias field is applied to the BrainWeb images. This way, a set of deformed images with known deformation fields is obtained. This procedure is repeated for each mesh spline degree, so each time a slightly different transformation field is obtained. Next, a different bias field is applied to the original BrainWeb T1 and PD images and they are registered to the deformed T1' and PD' BrainWeb images. Thus, for each patient and each parameter set, four nonrigid BrainWeb-patient validation registrations can be performed: two intramodal ($T1 \rightarrow T1', \mathrm{PD} \rightarrow \mathrm{PD}'$) and two intermodal ($T1 \rightarrow \mathrm{PD}', \mathrm{PD} \rightarrow T1'$).

Experiments were performed with 6, 14, 30, 62, 126, and 254 intensity bins; a mesh *B*-spline degree of 1, 2, and 3, and global and conditional MI. The registration settings can be found in Table III. We learned from previous experiments [19] that the

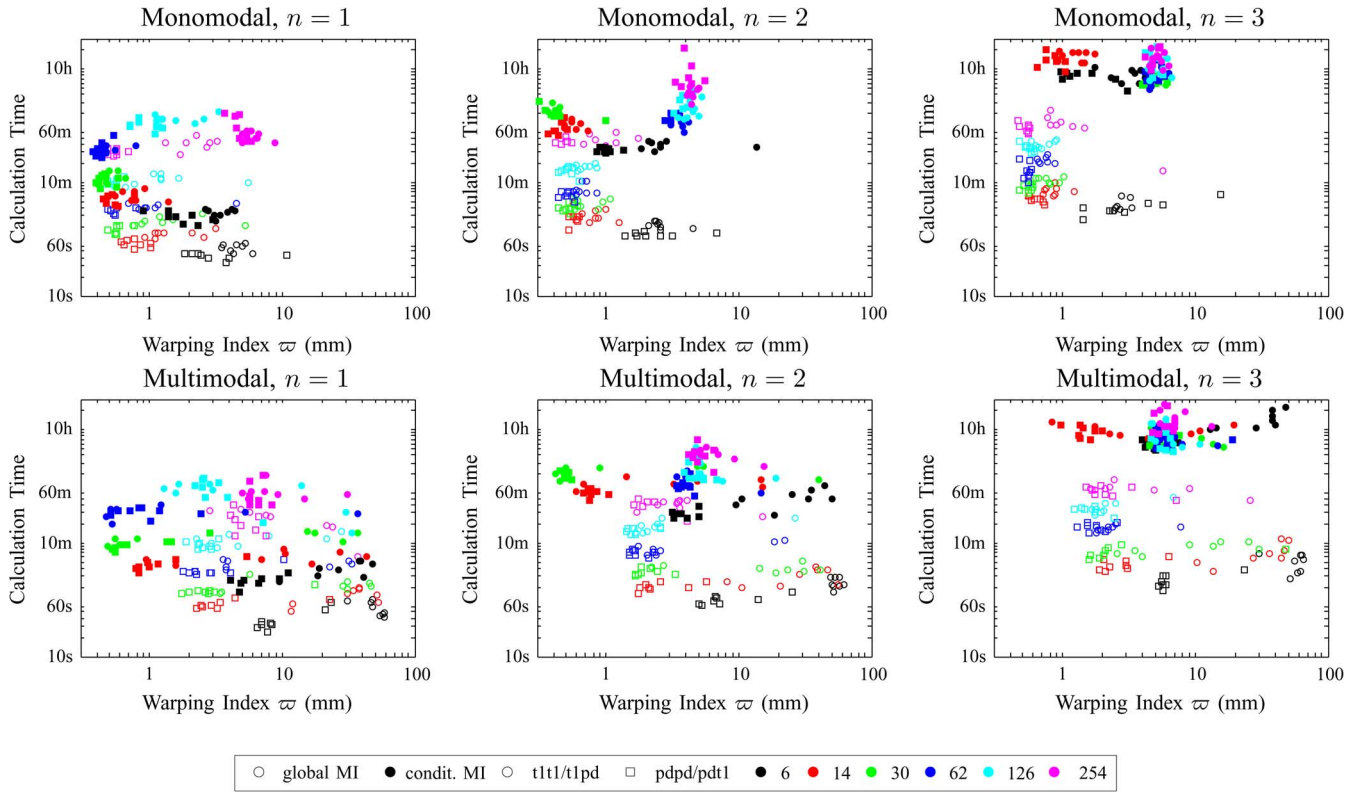

Fig. 8. BrainWeb registration results for (left) monomodal and (right) multimodal and (top to bottom) mesh spline degree $n = 1, 2$, and 3. Each graph represents the results for 6–254 bins and conditional and global MI. The best results are obtained with $n = 2$ and 30 bins. The difference is clearest in the multimodal case.

TABLE III
REGISTRATION SETTINGS FOR BRAINWEB IMAGES

Stage	Scale	Mesh (voxels)	Interpolation
1	1/4	256	partial volume
2	1/4	128	partial volume
3	1/4	64	partial volume
4	1/4	32	Parzen window
5	1/2	32	Parzen window
6	1/2	16	Parzen window

best results are obtained using PV interpolation for the coarse registration and PW interpolation for the finer stages. Therefore, we used PV interpolation for stages 1–3 and PW for the remaining stages. For the validation, we calculated the warping index between the ground-truth field and the field obtained by the validation registrations. The warping index was evaluated in the foreground only, defined by a ROI covering the brain region.

The results are summarized in Fig. 8, showing the calculation time and warping index for different registration settings. In most cases, best results are achieved using cMI, however at the cost of an increased calculation time. The biggest advantage of cMI over gMI is seen for multimodal registration with first and second degree *B*-splines. High registration errors (up to 60 mm) can occur when registration in the final stages diverges from the ground-truth optimum.

E. Experiment 5: Clinical CT/MR

In our last experiment, we applied global and conditional MI to MR/CT registration for colorectal cancer treatment [20]. The clinical goal of the registration is to transfer expert delineations of structures of interest from the MR scan to the CT scan. Most structures (and the tumor) are better visible in the MR scan, yet the CT scan is required to perform the actual planning and dose distribution calculation. This registration is a challenging problem, as MR and CT images differ significantly due to differences in rectum position and filling and the CT contrast of the region of interest is rather limited. Moreover, the MR image is slightly distorted with a bias field.

We dispose of 41 pairs of patient MR and CT scans. The pairs are obtained from 14 patients. For each patient, at three different time points during treatment, an MR and a CT image were recorded (one CT/MR pair is missing). A trained radiotherapist manually segmented the rectum in each CT and MR image. A multiresolution registration scheme is adopted to register the images within each pair to each other, using the CT image as reference image. The mesh spacing, as well for the cMI calculation as for the tensor product *B*-spline field, is gradually decreased, starting from $512 \times 512 \times 64$ voxels (about $400 \times 400 \times 313$ mm^3) in the first stage to $32 \times 32 \times 8$ voxels (about $25 \times 25 \times 40$ mm^3) in the fifth and last stage. Based upon initial experiments, we have chosen to use 30 bins for the floating and reference intensities in the joint histogram and PW interpolation.

The results obtained using cMI and gMI are compared to rigid alignment. As validation measure, the Dice similarity criterium (DSC) and surface distance (SD) between corresponding rectum segmentations are evaluated (see Fig. 9). Both for DSC and SD, conditional MI significantly ($p < 0.01$) outperformed global MI and rigid registration. An example comparing segmentations obtained by global and conditional MI is shown in Fig. 10. The cMI registrations took on average 10610 s (8050–13022 s) and

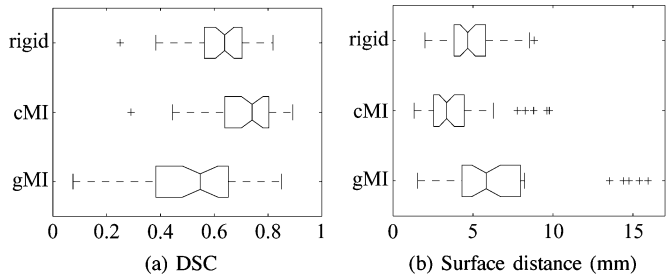

Fig. 9. Validation results for global MI (gMI) and conditional (cMI) MI registration of clinical MR and CT data sets for the manual delineation of the rectum. (a) Dice similarity overlap criterium (DSC) and (b) the average surface distance. Better results have a higher DSC and lower surface distance.

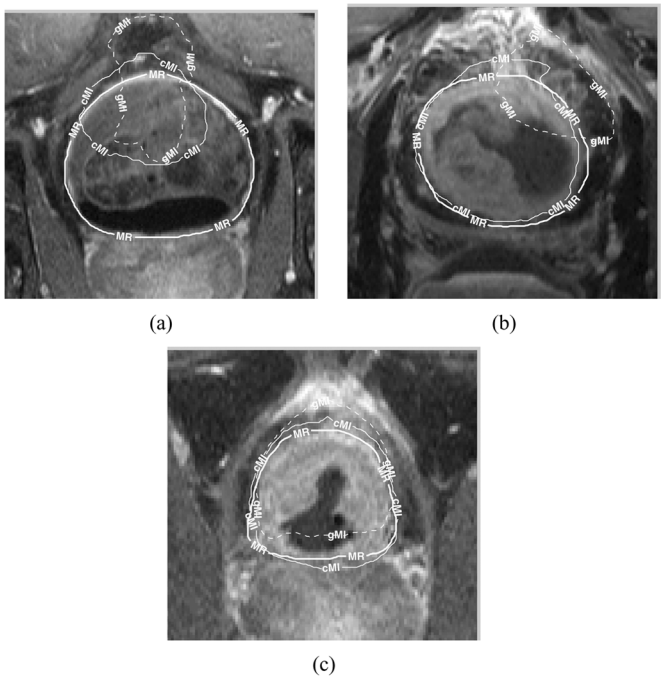

Fig. 10. Some rectum delineations obtained using gMI and cMI based image registration, compared to the manual (MR) segmentation.

the gMI registrations 723 s (466–960 s), with the fifth stage accounting for 56% and 66% of the total registration time.

IV. Discussion

In the context of nonrigid registration, registration by maximization of mutual information may lead to undesired results. Our experiments have shown that global MI, an established similarity measure for rigid registration, can be outperformed by conditional MI for different applications of nonrigid registration.

In experiment 1, we have shown that global MI has a spurious optimum when the size A_3^F of the inner circle S_3^F in the floating image is reduced to zero. Mutual information tries to maximize the marginal entropies while simultaneously minimizing the joint entropy. If A_2^R would grow while A_1^R shrinks proportionally, the depth of the local minimum for gMI and TC between $A_3^F = 0$ and $A_3^F = A_3^R$ will increase and the optimum at $A_3^F = 0$ will decrease compared to the one at $A_3^F = A_3^R$. At a certain point, the optimum at $A_3^F = A_3^R$ will become the global optimum, although a local optimum at $A_3^F = 0$ will remain. Looking at Fig. 3(b), the relative weight of both local optima depends only on $H(F)$. $H(F)$ reaches a maximum if both intensities have a probability of 0.5, thus the sign of the slope of $H(F)$ depends on the relative size of $A_1^R + A_3^R$ compared to A_2^R. As the slope remains limited compared to the slope of $H(R, F)$, it will never eliminate the local optimum as $A_3^F = 0$.

Total correlation, as proposed by Studholme [11] behaves similar to gMI. The steeper slope in the conditional joint entropy compared to the global joint entropy will make the local optima at both $A_3^F = 0$ and $A_3^F = A_3^R$ more attractive. The relative weight of both is determined by the global marginal floating entropy, which is unable to compensate for the local optimum at $A_3^F = 0$.

$H(R, F \mid X)$ also shows two local optima. However, the optimum at $A_3^F = 0$ is removed from the mutual information due to the large change in $H(F \mid X)$ for diminishing A_3^F. Even if the optimum of $H(F \mid X)$, which is currently around $A_3^F \approx 1.5 A_3^R$, shifts to the left due to a different relative size of A_3^R compared to $A_2^{R\prime}$, $H(F \mid X)$ will always be 0 for $A_3^F = 0$ as at this position the central region is completely homogeneous. Therefore, the optimum at $A_3^F = A_3^R$ will always be the global optimum. The conditional joint histogram is no longer distorted by features somewhere else in the image, and mutual information can do its job: comparing the joint and marginal entropies.

Experiment 2 validates the results obtained in experiment 1. As expected, the experiment clearly shows that global MI and total correlation can reduce small image features, whereas conditional MI encounters no problems.

In experiment 3, we compare global MI and conditional MI for bias-field distorted images. gMI estimates the joint probability by combining contributions from spatial locations all over the images, thus implicitly assuming the joint intensity distribution for a given structure is spatially homogeneous. However, a nonuniform bias field will provoke an artificial smoothing of the histogram causing spurious optima. Therefore, nonrigid registration using global MI might look for a transformation that sharpens this histogram without aligning corresponding structures. Using cMI, the extent of the assumed spatial homogeneity of the joint intensity distribution is limited to a single spatial bin. This reduces the relative influence of the bias field on the joint histogram compared to the features themselves. Thus the registration will more often register the features, as shown in experiment 3.

Experiment 4 moves from the artificial images in the previous experiments to more realistic images and deformations, yet sufficiently controlled to enable calculation of the ground-truth transformation. When the course of the similarity measure is influenced by a combination of noise, bias fields and intensity disagreements between the reference and floating image, cMI can increase both the robustness and accuracy of the registration.

Experiment 5, finally, shows that also for real clinical applications, cMI can provide significantly better results than gMI, however at the cost of extra computation time.

Although we did not experimentally validate it, we think cMI might also be beneficial for rigid image registration, especially for images distorted with a strong bias field or for images showing a lot of structures with similar intensities in one of both images. We think one should avoid calculating a similarity measure, and especially mutual information, at a larger

TABLE IV
% NONZERO BINS IN FINAL STAGE FOR BRAINWEB IMAGES

bins	6^2	14^2	30^2	62^2	126^2	254^2
gMI, $n=1$	100%	99%	96%	94%	90%	81%
gMI, $n=2$	100%	99%	96%	94%	90%	84%
gMI, $n=3$	100%	99%	96%	94%	90%	84%
cMI, $n=1$	47%	38%	32%	27%	20%	13%
cMI, $n=2$	54%	45%	40%	35%	29%	20%
cMI, $n=3$	58%	49%	46%	42%	35%	25%

scale than the scale of the transformation field. Otherwise, the transformation field might introduce spurious transformations. However, when the similarity measure is calculated at the same scale (as in this work) or at a smaller scale (e.g., when using cMI for rigid registration), the similarity measure will not be able to introduce this kind of spurious transformations.

This scale problem is not encountered when using e.g., sum of squared differences (SSD) as similarity measure. SSD calculates the image similarity on a voxel-scale by evaluating the voxel-by-voxel intensity difference. Global SSD is the integral of the local intensity difference over the whole image; in the case of nonrigid registration, the local similarity (i.e., the similarity within the span of a transformation parameter) is not influenced by intensities outside its span. The peculiarity about global MI is that a local change in one part of the image will change the joint histogram, and thus also the measured similarity in another, completely different part.

An important choice to make, both for cMI and tensor-product *B*-spline registration, is the mesh size. In the multiresolution registration scheme, the mesh size decreases at each refinement stage. For a given mesh size, the spatial bin size depends on the *B*-spline degree n. In 3-D, each conditional joint histogram receives contributions from $(n+1)^3$ mesh elements. The *B*-spline degree also modifies the smoothness of the transformation, as $\mathbf{g} \in C^{(n-1)}$. Finally, a higher *B*-spline degree will also exponentially increase the cMI calculation time, as for each conditional joint histogram more mesh elements have to be taken into account.

When the size of a spatial bin decreases, the number of joint intensity pairs contributing to a single bin in the joint histogram diminishes, reducing the statistical power of the histogram. This effect is countered because the number of features present in a spatial bin will decrease also. As the intensity bins are defined such that they span the intensity range of the whole image, the number of nonzero bins will decrease for fewer features. The available information is distributed over a smaller number of bins, again increasing the statistical power of the joint histogram. Therefore, in a small mesh element a significant amount of bins is expected to be equal to zero. The average percentage of nonzero bins for the BrainWeb experiments is given in Table IV. The number of voxels contributing to each nonzero bin further increases due to the Parzen window or generalized partial volume interpolation, both using second degree *B*-splines. Because of this interpolation, each joint intensity pair actually contributes to 3×3 or $3 \times 3 \times 3$ bins.

In the BrainWeb case, as can be seen in Fig. 8, we obtain cMI results comparable to gMI with first-degree *B*-splines, up to 126×126 bins in the joint histogram and a voxel mesh of $8 \times 8 \times 8$ voxels. In theory, this would lead to on average 0.258 voxels/bin in the final stage; after correction for nonzero bins and interpolation we end up with about 11.6 voxels/bin. Best results are obtained with $n = 2$ and 30^2 bins, using on average 346 voxels/bin.

It is difficult to define *a priori* an ideal size for the spatial bins and mesh size. In our experiments, we did not encounter problems using the smallest mesh size used in previous experiments using only gMI [19]. We choose the smallest mesh size that still gives a reasonable increase in similarity for a given result for gMI, and do not notice problems with cMI for the same settings.

When the spatial bins become too small to allow for a reliable MI calculation, an alternative could be to use local cross-correlation, as proposed by Weese *et al.* [10]. They argue that, for sufficiently small regions, the number of distinct structures present in each region is almost everywhere either 1 or 2 as most regions will be either completely contained in a single homogeneous structure or located at the boundary of two adjacent structures, with a negligible number of regions containing three or more structures. For one or two distinct structures, cross-correlation is a sufficiently strong measure, as argued by Weese *et al.*. They obtain a good agreement between registrations using local CC and MI. Based on these results, local CC might be a good measure for nonrigid registration as well. For this to be true, the assumptions made above for local CC should not only hold on average over the whole image (as for affine registration), but also in each individual region.

Within this work, we have chosen to use a *B*-spline mesh as input for the spatial bins. It is straightforward to use any other (fuzzy) image segmentation available. In this way, the work is related to the work of D'Agostino *et al.* who uses classes obtained from an atlas to build multiple joint histograms [21].

The main disadvantage of cMI is its increased computation time. Whereas the average multimodal gMI registration times in experiment 4 for first, second, and third degree *B*-splines are 100 s, 190 s, and 484 s, respectively, the corresponding cMI registrations take on average 436 s, 5966 s, and 17675 s, which is an increase with a factor of 4, 31, and 37. However, as for cMI the conditional histograms can be calculated separately, it can easily be parallelized.

V. CONCLUSION

We have proposed cMI as a new similarity measure for nonrigid image registration. We have shown that cMI can overcome several problems inherent to the use of global MI, using artificial and clinical images, and theoretically explained the results.

APPENDIX

Consider Fig. 2. We use A_i^R and I_i^R to describe the area and intensity of structure S_i^R and analogous. As $I_3^F = I_1^F$, the joint histogram for perfectly aligned images, is given by

$$\mathcal{H}_g = \begin{bmatrix} 0 & A_2^R & 0 \\ A_1^R & 0 & A_3^R \end{bmatrix} = \begin{bmatrix} 0 & A_2^F & 0 \\ A_1^F & 0 & A_3^F \end{bmatrix} \tag{21}$$

as initially $A_i^R = A_i^F$. To calculate the conditional mutual information, we assume a zeroth degree *B*-spline mesh overlaid over the reference image, as indicated by the dashed line in Fig. 2.

The conditional mutual information will be the sum of the mutual information of each of the spatial bins. Using $A_2^{R\prime}$ as the area of the part of S_2^R contained within the central spatial bin, the joint histogram for this central spatial bin is

$$\mathcal{H}_c = \begin{bmatrix} 0 & A_2^{R\prime} & 0 \\ 0 & 0 & A_3^R \end{bmatrix} = \begin{bmatrix} 0 & A_2^{F\prime} & 0 \\ 0 & 0 & A_3^F \end{bmatrix}. \quad (22)$$

Now, we apply a perturbation ϵ on S_3^F such that $A_3^F = A_3^R + \epsilon$. The global and conditional joint histograms now become

$$\mathcal{H}_g(\epsilon) = \begin{cases} \begin{bmatrix} 0 & A_2^R - \epsilon & 0 \\ A_1^R & \epsilon & A_3^R \end{bmatrix}, & \text{for } \epsilon \geq 0 \\ \begin{bmatrix} 0 & A_2^R & -\epsilon \\ A_1^R & 0 & A_3^R + \epsilon \end{bmatrix}, & \text{for } \epsilon \leq 0. \end{cases} \quad (23)$$

$$\mathcal{H}_c(\epsilon) = \begin{cases} \begin{bmatrix} 0 & A_2^{R\prime} - \epsilon & 0 \\ 0 & \epsilon & A_3^R \end{bmatrix}, & \text{for } \epsilon \geq 0 \\ \begin{bmatrix} 0 & A_2^{R\prime} & -\epsilon \\ 0 & 0 & A_3^R + \epsilon \end{bmatrix}, & \text{for } \epsilon \leq 0. \end{cases}. \quad (24)$$

The perturbation will only influence the size of the structures contained within the central spatial bin, so the contribution of the peripheral spatial bins to cMI and TC remains constant.

From (23) and (24), the mutual information can be calculated easily using (2). To create the graph in Fig. 3, we chose a radius of 3.5 and 12.5 for the inner and outer circle and a square edge of 32. The local optimum of the gMI is situated at $\epsilon = -A_2^R A_3^R / (A_1^R + A_2^R)$.

REFERENCES

[1] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," in *Proc. IPMI*, 2007, pp. 725–737.

[2] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Proc. IPMI*, 1995, pp. 263–274.

[3] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Proc. ICCV*, 1995, pp. 16–23.

[4] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proc. IEEE*, vol. 91, no. 10, pp. 1699–1722, Oct. 2003.

[5] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[6] J. West *et al.*, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, no. 4, pp. 554–566, 1997.

[7] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid multimodal image registration using mutual information," in *Proc. MICCAI*, 1998, pp. 1099–1106.

[8] B. Likar and F. Pernuš, "A hierarchical approach to elastic registration based on mutual information," *Image Vis. Comput.*, vol. 19, no. 1–2, pp. 33–44, Jan. 2001.

[9] A. Andronache, P. Cattin, and G. Székely, "Local intensity mapping for hierarchical non-rigid registration of multi-modal images using the cross-correlation coefficient," in *Proc. WBIR*, 2006, pp. 26–33.

[10] J. Weese, P. Rösch, T. Netsch, T. Blaffert, and M. Quist, "Gray-value based registration of CT and MR images by maximization of local correlation," in *Proc. MICCAI*, 1999, pp. 656–663.

[11] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 626–639, May 2006.

[12] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Signal Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[14] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. 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.

[15] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, Jun. 1998.

[16] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[17] H.-M. Chen and P. K. Varshney, "Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1111–1119, Sep. 2003.

[18] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995.

[19] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens, "Comparison between Parzen window interpolation and generalised partial volume estimation for non-rigid image registration using mutual information," in *Proc. WBIR*, 2006, pp. 206–213.

[20] P. Slagmolen, D. Loeckx, S. Roels, X. Geets, F. Maes, K. Haustermans, and P. Suetens, "Nonrigid registration of multitemporal CT and MR images for radiotherapy treatment planning," in *Proc. WBIR*, 2006, pp. 297–305.

[21] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "Atlas-to-image non-rigid registration by minimization of conditional local entropy," in *Proc. IPMI*, 2007, pp. 320–332.