Edgeworth-based approximation of Mutual Information for medical image registration

Mathieu Rubeaux[12], Jean-Claude Nunes[12], Laurent Albera[12] and Mireille Garreau[12]
[1] INSERM, UMR 642, Rennes, F-35000, France
[2] Université de Rennes 1, LTSI, F-35000, France
e-mail: jean-claude.nunes@univ-rennes1.fr

***Abstract*— Mutual Information (MI) has been extensively used as a similarity measure in image registration and motion estimation, and it is particularly robust for 3D multimodal medical image registration. However, MI estimators are known i) to have a high variance and ii) to be computationally costly. In order to overcome these drawbacks, we propose a new similarity measure based on an Edgeworth-based third order expansion of MI and named 3-EMI in the following. This kind of approximation is well known in signal processing, and especially in Independent Components Analysis (ICA), but its computation is easier since data can be prewhitened contrary to images in registration. The performance of affine and non-rigid registrations based on the 3-EMI metric is studied through computer results in the context of cardiac multislice computed tomography. In fact, an estimate of the 3-EMI metric using sample statistics is compared to a histogram-based estimate of the standard normalized MI, showing a better robustness of the 3-EMI measure with respect to the range of the searched deformation. In addition, in practice, one part of the floating image may be missing regarding the reference image. Computer results show that our approach is less sensitive to such a practical problem.**

***Keywords*— Image registration, mutual information, Edgeworth expansions, cumulants, MSCT, cardiac images.**

I. Introduction

Registration is a problem of big interest in medical image analysis. It consists in recovering a transformation or deformation field that optimizes some similarity measurements between two sets of data (images, volumes, sequences) and is a fundamental component towards computer assisted diagnosis and interventional gestures. The different imaging modalities that are used in clinical practice are most often complementary and provide data of various sizes (2D/3D or dynamic images) and various nature (morphological, functional, metabolic) that have to be combined to perform a complete analysis. These imaging modalities are always in constant technological evolution and their combination offer new tools for diagnosis and therapy. But the development of efficient registration techniques is needed in many types of clinical situations that can concern the registration of images acquired in i) mono-modality for the same patient but at different instants (in dynamic analysis for moving organs or in treatment following) ; ii) in multi-modality for fusion of complementary informations (issued from nuclear imaging and scanner or MRI imaging in oncology for example). The application of image registration that is presented in this paper concerns the motion estimation in Multi-Slice Computed Tomography (MSCT) for cardiac function assessment, this modality providing 3D image sequences and offering the means to analyse both anatomy and function. Many automatic medical image registration methods have been proposed in the literature (see [1] for a bibliographical survey). The basic principle of image registration consists in finding the geometric transformation between two images or 3D volumes which minimizes a cost or objective function. These registration techniques commonly use intensity correlation, mean square difference of the image intensity values, or MI as objective functions.

In the context of multidimensional non-rigid and multimodal medical image registration [2], the techniques based on MI are the most common since their first appearance [3], [4]. Indeed, they allow to use all the statistical information contained in the data. More particularly, built from the Kullback-Leibler divergence [5], MI provides a statistical measure of dependence between two random variables, say between two probability distributions. A variant of this metric, the Normalized MI (NMI), is nowadays intensively used in medical image registration [6]. In practice, MI cannot be computed exactly and need to be estimated from an estimation of marginal and joint Probability Density Functions (PDF's). There exists essentially three different techniques to estimate MI: histogram-based estimators [7], kernel-based estimators [8], [9] and parametric methods. Parametric methods are clearly not appropriate to the context of medical image registration due to the statistical complexity of processed images. Regarding kernel-based estimators, they have too many adjustable parameters such as the optimal kernel width and the optimal kernel form. The latter parameters can be iteratively determined [8], but the resulting method becomes computationally intensive. Histogram-based estimators are more attractive due to their simplicity. However, they suffer from variance, bias caused by i) the finite number of observations, ii) quantization and iii) the finite histogram.

In order to overcome this drawback, a new similarity measure is proposed in this paper, which is an approximation of MI based on a third order Edgeworth expansion of marginal and joint PDF's. This idea has been introduced two decades ago in the context of Independent Component Analysis (ICA) [10]. Nevertheless, the different approximations derived for ICA can't be use for our purpose. Indeed, either they only use the Edgeworth expansion to approximate the marginal entropies without approximating the joint entropy [11] or they

978-1-4244-7249-9/10/$26.00 ©2010 IEEE

require to prewhiten the data in order to considerably simplify the computation of the global approximation [12]. Note that the prewhitening step is only possible if the random variables taken into account are uncorrelated enough. This makes sense in the context of ICA because the goal is to restore the statistical independence of the considered variables. However, in the registration task, this is inappropriate because the aim is to find the perfect match between random variables, the computation of the approximation becomes then more tedious, and it is given in this paper leading to the novel 3-EMI metric. The performance of affine and non-rigid registrations based on the 3-EMI metric is studied through computer results in the context of cardiac multislice computed tomography. In fact, an estimate of the 3-EMI metric using sample statistics is compared to a histogram-based estimate of the standard Normalized MI (NMI) [6], showing a better robustness of the 3-EMI estimator with respect to the amplitude of the searched deformation. In addition, in practice, one part of the floating image may be missing regarding the reference image. Computer results show that our approach is less sensitive to such a practical problem.

II. MI approximation based on Edgeworth expansion

The major components of a registration framework are basically: i) the feature space (the characteristics of the images taken into account), ii) the similarity measure used to compare these characteristics, iii) the type of geometrical transformation we consider, and iv) the chosen optimization method. In this section, we focus on the new similarity measure proposed to register two images or 3D volumes.

A. The classical MI

In information theory, MI of a vector $\boldsymbol{x}$ of two random variables $x^{(1)}$ and $x^{(2)}$ gives a measure of the statistical dependence of both variables. It can be expressed as a function of the marginal and joint entropies of $x^{(1)}$ and $x^{(2)}$:

$$MI(\boldsymbol{x}) = H(x^{(1)}) + H(x^{(2)}) - H(\boldsymbol{x}) \tag{1}$$

Shannon entropy [13] $H(x)$ of a random variable x is a measure of the average or expected information content of an event described x, given by:

$$H(x) = -\int_{\mathbb{R}} p_x(u) \log(p_x(u))\, du \tag{2}$$

where p_x is the marginal PDF of x. The joint entropy $H(\boldsymbol{x})$, measuring the dispersion of the joint PDF, $p_{\boldsymbol{x}}$, of the vector $\boldsymbol{x}$ is defined by:

$$H(\boldsymbol{x}) = -\int_{\mathbb{R}^2} p_{\boldsymbol{x}}(\boldsymbol{u}) \log(p_{\boldsymbol{x}}(\boldsymbol{u}))\, d\boldsymbol{u} \tag{3}$$

In medical image registration, NMI is often preferred because it is an overlap invariant similarity measure [6]. It is given by:

$$NMI(\boldsymbol{x}) = \frac{H(x^{(1)}) + H(x^{(2)})}{H(\boldsymbol{x})} \tag{4}$$

In our experiments, we will compare the 3-EMI measure to the NMI one due to its effectiveness in the context of medical image registration [2]. As shown in equation (4), the computation of NMI requires in practice the estimation of the marginal and joint PDF's of the couple $(x^{(1)}, x^{(2)})$. This is done in [6] by a histogram-based method. For the marginal PDF, it consists in counting the occurrence of values at points within an image. For the joint PDF, given several realizations of vector $\boldsymbol{x}$ provided by both images to register, a joint histogram is built that tells us how often pairs of values occur together.

B. Edgeworth-based third order approximation of MI (3-EMI)

Edgeworth expansion aims at approximating a PDF from its cumulants and Hermite polynomials. Given a zero-mean bidimensional random vector $\boldsymbol{x} = [x^{(1)}, x^{(2)}]$, second and third order cumulants of $\boldsymbol{x}$ are defined by:

$$\begin{aligned} \kappa^{i_1,i_2}(\boldsymbol{x}) &= \mathsf{E}[x^{(i_1)} x^{(i_2)}] \\ \kappa^{i_1,i_2,i_3}(\boldsymbol{x}) &= \mathsf{E}[x^{(i_1)} x^{(i_2)} x^{(i_3)}] \end{aligned} \tag{5}$$

where $\mathsf{E}[.]$ denotes the mathematical expectation operator. For instance, the third order marginal cumulant of the zero-mean random variable $x^{(i)}$ is given by $\kappa^{i,i,i}(\boldsymbol{x}) = \mathsf{E}[(x^{(i)})^3]$. Regarding Hermite polynomials, definitions and useful properties are recalled in appendix.

Edgeworth expansion is formally equivalent to the Gram-Charlier one. However, from a practical point of view, it has better asymptotic convergence properties due to the correction terms that are monotonely decreasing in N providing that the considered random variables can be decomposed as a sum of N independent variables. Hence our preference for Edgeworth expansion. We assume hereafter that both random variables of $\boldsymbol{x}$ have unit variance. In our context, it means that our images have been standardized which is less restrictive than the prewhitening step discussed in introduction. The univariate expansion up to order 3 of the marginal PDF of $x^{(i)}$ is given by [14]:

$$p_{x^{(i)}}(u^{(i)}) = \phi_{x^{(i)}}(u^{(i)}) \left(1 + \frac{\kappa^{i,i,i}(\boldsymbol{x}) H_3(u^{(i)})}{3!}\right) + \mathcal{O}(N^{-1}) \tag{6}$$

where $\phi_{x^{(i)}}$ is the marginal normal distribution given by (14) and H_3 is the third order Hermite polynomial (15). The bivariate expansion of the joint PDF $p_{\boldsymbol{x}}$ is defined by [14]:

$$\begin{aligned} p_{\boldsymbol{x}}(\boldsymbol{u}) = \phi_{\boldsymbol{x}}(\boldsymbol{u}) \Big(1 &+ \frac{1}{6}\kappa^{1,1,1}(\boldsymbol{x}) H_{1,1,1}(\boldsymbol{u}) \\ &+ \frac{1}{2}\kappa^{1,1,2}(\boldsymbol{x}) H_{1,1,2}(\boldsymbol{u}) + \frac{1}{2}\kappa^{1,2,2}(\boldsymbol{x}) H_{1,2,2}(\boldsymbol{u}) \\ &+ \frac{1}{6}\kappa^{2,2,2}(\boldsymbol{x}) H_{2,2,2}(\boldsymbol{u})\Big) + \mathcal{O}(N^{-1}) \end{aligned} \tag{7}$$

where $\phi_{\boldsymbol{x}}$ is the bivariate normal distribution defined by (18) and H_{i_1,i_2,i_3} is a third-order bivariate Hermite polynomial (20).

Using these expansions, we can approximate the marginals (2) and joint (3) entropies. Note that, we focus on third order approximation, because they showed to be sufficient for our registration task. Further improvements will include an increase of the order of the approximation in order to have a more precise estimation of the entropies. The marginal entropy up to third order is well-known [14] and is given by:

$$H(x^{(i)}) \approx \frac{1}{2}\log(2\pi\mathrm{e}) - \frac{1}{2}\frac{\kappa^{i,i,i}(\boldsymbol{x})^2}{3!} \tag{8}$$

One contribution of this paper is to give the approximation of the joint entropy of correlated random variables. Its computation, a bit tedious, will be given in a longer paper. We just give here the main ideas of such a computation. We start by inserting (7) into (3). Then, we simpify the expression using i) the Taylor expansion $(1+x)\log(1+x) \approx x + \frac{x^2}{2}$ and ii) the orthogonality property of Hermite polynomials (20). We get :

$$\begin{aligned}
H(\boldsymbol{x}) \approx\ & 1 + \log(2\pi) + 0.5\log(1-\rho^2) - \frac{1}{12(1-\rho^2)^3}\Bigg(\\
& \kappa^{1,1,1}(\boldsymbol{x})^2 + \kappa^{2,2,2}(\boldsymbol{x})^2 + 3\kappa^{1,1,2}(\boldsymbol{x})^2 + 3\kappa^{1,2,2}(\boldsymbol{x})^2 \\
& - 6\rho\Big(\kappa^{1,1,1}(\boldsymbol{x})\kappa^{1,1,2}(\boldsymbol{x}) + \kappa^{1,2,2}(\boldsymbol{x})\kappa^{2,2,2}(\boldsymbol{x}) \\
& + 2\kappa^{1,1,2}(\boldsymbol{x})\kappa^{1,2,2}(\boldsymbol{x})\Big) + 6\rho^2\Big(\kappa^{1,1,2}(\boldsymbol{x})^2 + \kappa^{1,2,2}(\boldsymbol{x})^2 \\
& + \kappa^{1,1,1}(\boldsymbol{x})\kappa^{1,2,2}(\boldsymbol{x}) + \kappa^{1,1,2}(\boldsymbol{x})\kappa^{2,2,2}(\boldsymbol{x})\Big) \\
& - 2\rho^3\Big(\kappa^{1,1,1}(\boldsymbol{x})\kappa^{2,2,2}(\boldsymbol{x}) + 3\kappa^{1,1,2}(\boldsymbol{x})\kappa^{1,2,2}(\boldsymbol{x})\Big)\Bigg)
\end{aligned} \tag{9}$$

where $\rho = \kappa^{1,2}(\boldsymbol{x})$ is the covariance between $x^{(1)}$ and $x^{(2)}$. We see that this expression is mainly composed of multivariate cumulants. In practice, cumulants cannot be exactly computed and require to be estimated from several realizations of the involved random variables. More particularly, let $\left\{x^{(i)}[\ell]\right\}_{1\leq\ell\leq L}$ be a set of L independent realizations of the zero-mean variable $x^{(i)}$, we get:

$$\kappa^{i_1,i_2}(\boldsymbol{x}) = \frac{1}{L}\sum_{\ell=1}^{L} x^{(i_1)}[\ell]\, x^{(i_2)}[\ell] \tag{10}$$

$$\kappa^{i_1,i_2,i_3}(\boldsymbol{x}) = \frac{1}{L}\sum_{\ell=1}^{L} x^{(i_1)}[\ell]\, x^{(i_2)}[\ell]\, x^{(i_3)}[\ell]$$

With formulas (8) and (9), the MI expression (1) can be approximated. It is natural to ask oneself if this third order approximation is good enough. In fact, assuming that $x^{(1)}$ and $x^{(2)}$ can be decomposed as a sum of N independent variables it is possible to deduce that the approximation error is about N^{-1}. The relevance of the decomposition assumption will be studied in details in a longer paper. We have now to incorporate our 3-EMI approximation in a global registration scheme, as explained in the next section.

III. The Registration Scheme

As mentioned above, a registration method is usually defined by its feature space, deformation model, similarity measure and optimization scheme. In the following, we will explain the choices we made to build up our experiments. We consider here an iconic method, so the features used are the pixel values and cumulants are estimated from these values.

A. Motion model

One crucial choice to make is the motion model to take into account. For our purpose, two main types of transformation have to be considered. First, we focus on affine transformations, with seven parameters in two dimensions (translations t_x and t_y, scalings s_x and s_y, shearings h_x and h_y, and a rotation angle θ). Considering both images to register, I_F and I_R will denote the floating and the reference images, respectively. The coordinates of a pixel in these images are then given by $\boldsymbol{v}_R = (x_R, y_R)^\intercal$ and $\boldsymbol{v}_F = (x_F, y_F)^\intercal$, respectively. Thus, the affine transformation linking a pixel of I_F according to a pixel of I_R is given by:

$$\boldsymbol{v}_R = \boldsymbol{A}\,\boldsymbol{v}_F + (t_x, t_y)^\intercal \tag{11}$$

where $\boldsymbol{A}$ is the matrix defined as:

$$\boldsymbol{A} = \begin{pmatrix} s_x\sin(\theta)h_y + s_x\cos(\theta) & s_x\sin(\theta) + s_x\cos(\theta)h_x \\ s_y\cos(\theta)h_y - s_y\sin(\theta) & s_y\cos(\theta) - s_y\sin(\theta)h_x \end{pmatrix}$$

However, the motion of the heart is highly non-rigid so that affine transformations alone are not sufficient. To describe a realistic deformation, one need to perform a non-rigid registration. For that purpose, we use a Free-Form Deformation (FFD) [15] model based on B-Splines. B-Splines are very popular in medical image registration due to their pleasant properties: compacity of their support, separability in each dimension, derivability and a relative straightforward implementation. But they are above all a powerful tool for modeling deformable objects. The FFD model warps an image by moving an underlying set of control points distributed over a regular grid. The displacement of a point $\boldsymbol{v}$ of the image can be written as a linear combination of B-spline functions $\beta^{(k)}$, weighted by the parameters $\boldsymbol{\xi}^{(k)}$ in the neighborhood $\mathcal{K}(\boldsymbol{v})$ of this point:

$$\phi(\boldsymbol{v}, \boldsymbol{\xi}) = \sum_{k\in\mathcal{K}(\boldsymbol{v})} \boldsymbol{\xi}^{(k)}\beta^k(\boldsymbol{v}) \tag{12}$$

B. Similarity measure

As explained in the previous section, the main contribution of this paper is to present a new similarity measure based on an Edgeworth-based third order expansion of MI. The aim of our test procedures is to compare the 3-EMI metric with the NMI one (4). Therefore in our experiments, we will use both metrics.

C. Optimization procedure

As a first approach, we choose a steepest gradient descent optimization scheme. It allows us to make a direct comparison between EMI and NMI metrics, in term of effectiveness and sensitivity to the amplitude of the deformation. The gradient is computed numerically, but our future work will be to write it analytically.

Moreover, we choose a coarse to fine (or pyramidal) approach, which brings a speed-up in the registration. During an affine registration procedure, the pyramidal approach consists in downsampling our images and performing the affine registration on the coarser level. When the optimum parameters are found, they are used to initialize the next level of the pyramid. During the non-rigid registration, the downsampling is done directly on the control grid of the B-Splines. We start with a coarse grid which is refined after the optimum is reached. This approach gives faster and better results, because it allows to register global as well as local motion.

IV. Test procedure and Results

We test the 3-EMI similarity measure on both affine and non-rigid registration tasks, and compare it to the NMI metric. We use real 2D MSCT images that are deformed using controlled parameters. Then, we perform the registration using the same optimizer for both similarity measures. When the optimum in the sense of each measure is reached, the Frobenius norm of the difference between the reference and registered images is calculated, in order to evaluate the quality of our registration: the Frobenius norm equals zero when the registered image fits perfectly the reference one.

A. Affine registration

We construct our synthetic datasets making the seven affine parameters vary. The importance of the deformation is controlled using an extra parameter α. Two test procedures are led. First, we fix the parameters such as $t_x = t_y = \theta = \alpha$, $s_x = s_y = 1 + 0.1\alpha$ and $h_x = h_y = 0.1\alpha$ with α varying from 0 to 5. The resulting images are highly deformed because of the high scaling and shearing, but due to the small translations, almost all the information contained in the reference image remains in the floating image. We perform the registration on two hundred sets of images. The computer results are good for both similarity measures.

Then, we use different parameters to deform our images: $t_x = t_y = \theta = \alpha$, $s_x = s_y = 1 + 0.01\alpha$ and $h_x = h_y = 0.01\alpha$, α varying from 0 to 40. This time, important translations and rotations occur, so that part of the reference image is missing in the floating one. Again, the tests are led on two hundred sets of images, and the results are given below: as soon as the missing part in the deformed image becomes too important

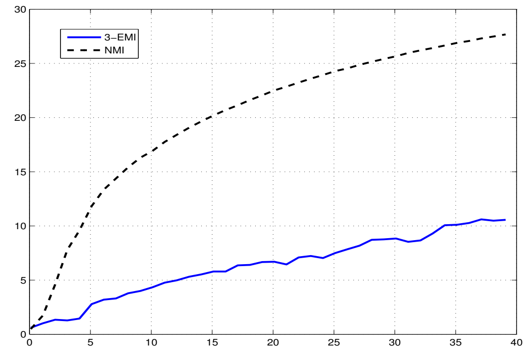
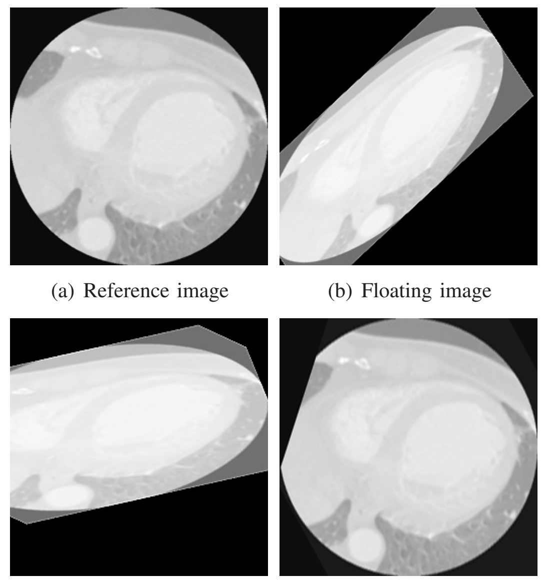


Fig. 1. Frobenius norm of the difference between the reference and registered images regarding the amplitude of the affine deformation.



(a) Reference image (b) Floating image

(c) NMI registered image (d) 3-EMI registered image

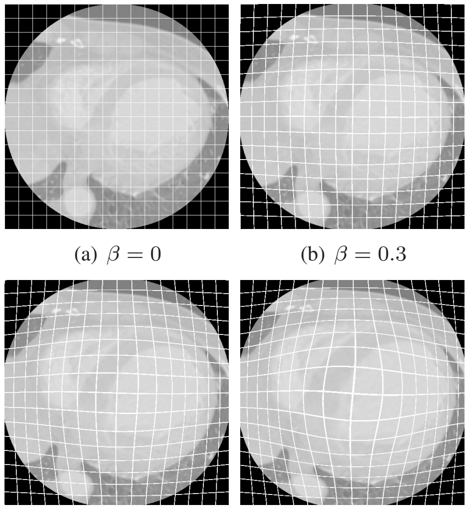
Fig. 2. Example of an affine registration for MSCT slices. The initial affine transformation is given by $\alpha = 35$. The 3-EMI-based registered image (d) nicely looks like the reference image (a), excepted on the left-hand corner which is missing in the floating image. In this extreme example, we see that NMI performs poorly (c).

(for α around 10), the NMI metric fails to register well. On the opposite, the 3-EMI measure still registers our images, even when a large part of the reference image is missing. An illustration is given in Fig. 2, with $\alpha = 35$. In Fig. 1, we see that 3-EMI measure outperforms the NMI metric.

The affine tests are conclusive, but non-rigid registration also has to be tested, since it's the final goal in our application.

B. Non-rigid registration

In the same way, we deform non-rigidly our reference image to obtain the floating image. To do so, we move the B-Splines control points of the grid over the images, increasing the amplitude of the distortion via a parameter β varying from 0.1 to 1. Then, the floating image is deformed thanks to these control points, which ensure a regular deformation. Examples



(a) $\beta = 0$ (b) $\beta = 0.3$

(c) $\beta = 0.6$ (d) $\beta = 1$

Fig. 3. Examples of B-Spline transformed images, with an increasing distortion coefficient β.

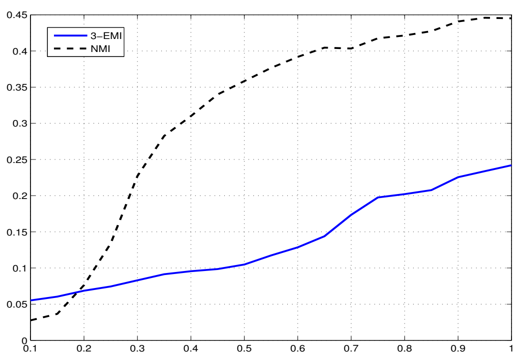
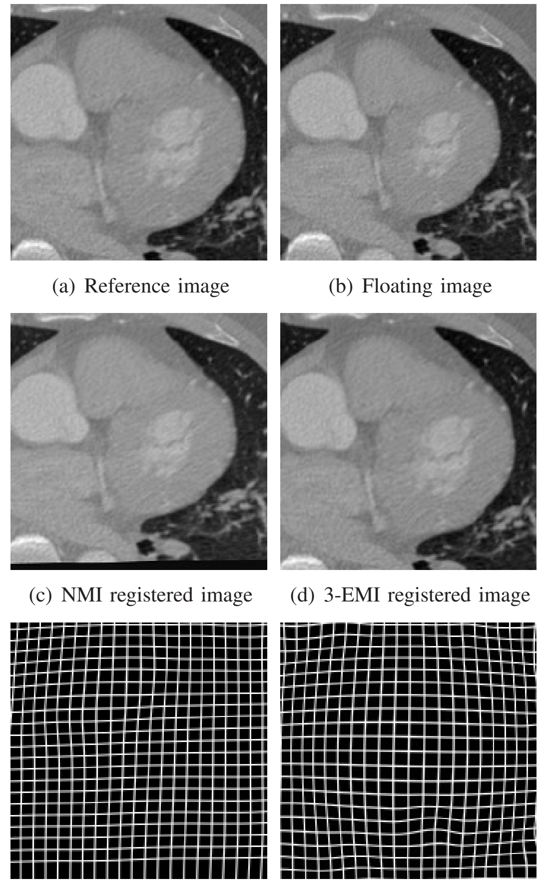


Fig. 4. Frobenius norm of the difference between the reference and registered images regarding the amplitude of the non-rigid deformation.



(a) Reference image (b) Floating image

(c) NMI registered image (d) 3-EMI registered image

(e) NMI final B-Spline grid (f) 3-EMI final B-Spline grid

Fig. 5. Example of a non-rigid registration for MSCT slices, with $\beta = 0.35$. The 3-EMI-based registered image (d) fits the reference image (a) well, while the NMI-based registration starts to diverge (c).

of the deformed images regarding the β parameter are given in Fig. 3. The registration is performed using the floating image as the reference and vice versa, which allows to make a direct comparison between the synthetic control grids of B-Splines and the ones estimated during the registration process. Once again, we see that 3-EMI outperforms NMI (Fig. 4), particularly when important deformations are taken in account. A registration example is given in Fig. 5 with $\beta = 0.35$.

V. Conclusion

In this paper, we present a new similarity measure based on a third order approximation of MI using an Edgeworth expansion. The underlying mathematical tools of this expansion are the Hermite polynomials and the cumulants. The orthogonality properties of Hermite polynomials are used to simplify the expressions, while the cumulants contain the relevant statistical information of the images to register. In fact, cumulants allow a good description of the image, without any prior information. They are a powerful tool which already showed its performance in ICA, but maybe not enough in image registration. A lot of work still has to be done. First, we would like to write analytically the gradient of our cost function, in order to increase the performance of our registration scheme. We would like also to use higher order Edgeworth expansions to use more statistical information about the images. Moreover, the number N of independent images that yield the MSCT images in a linear combination will be studied in order to quantify the approximation error of our metric. Finally, a 3-dimensional registration method has to be set-up, but it will be straightforward due to the properties of the different components of our registration scheme.

Appendix I
Univariate Hermite polynomials

The univariate Hermite polynomials are defined by:

$$\phi_x(u) H_k(u) = (-1)^k \, \frac{d^k \phi_x(u)}{du^k} \tag{13}$$

where:

$$\phi_x(u) = \frac{1}{\sqrt{2\pi Var(x)}} \mathrm{e}^{-(u-\mathsf{E}[x])^2/(2Var(x))} \tag{14}$$

is the marginal normal distribution. The first six univariate Hermite polynomials are given by:

$$\begin{aligned} H_0(u) &= 1 \\ H_1(u) &= u \\ H_2(u) &= u^2 - 1 \\ H_3(u) &= u^3 - 3u \end{aligned} \tag{15}$$

They enjoy the following orthogonality property:

$$\int_{\mathbb{R}} \phi_x(u) H_i(u) H_j(u) du = \begin{cases} i! \text{ if } \ i = j \\ 0 \text{ otherwise} \end{cases} \tag{16}$$

Appendix II
Bivariate Hermite polynomials

In the same way, one can define the bivariate Hermite polynomials as:

$$\phi_{\boldsymbol{x}}(\boldsymbol{u}) H_{i_1,\ldots,i_k}(\boldsymbol{u}) = (-1)^k \frac{\partial^k \phi_{\boldsymbol{x}}(\boldsymbol{u})}{\partial u^{(i_1)} \ldots \partial u^{(i_k)}} \tag{17}$$

with:

$$\phi_{\boldsymbol{x}}(\boldsymbol{u}) = \frac{1}{(2\pi)|\boldsymbol{\kappa}_{\boldsymbol{x}}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{u} - \mathsf{E}[\boldsymbol{x}])^T \boldsymbol{\kappa}_{\boldsymbol{x}}^{-1} (\boldsymbol{u} - \mathsf{E}[\boldsymbol{x}])} \tag{18}$$

the bidimensionnal normal distribution, where $\boldsymbol{\kappa}_{\boldsymbol{x}} = \mathsf{E}[\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}] - \mathsf{E}[\boldsymbol{x}]\mathsf{E}[\boldsymbol{x}^{\mathsf{T}}]$ is the covariance matrix of $\boldsymbol{x}$ whose elements are denoted by $\kappa^{i,j}(\boldsymbol{x})$ and where $\boldsymbol{\kappa}_{\boldsymbol{x}}^{-1}$ is the inverse of this covariance matrix whose elements are written $\kappa_{i,j}(\boldsymbol{x})$. The first three multivariate covariant Hermite polynomials are given by:

$$\begin{aligned} H_{i_1}(\boldsymbol{u}) &= \kappa_{i_1,1}(\boldsymbol{x}) u^{(1)} + \kappa_{i_1,2}(\boldsymbol{x}) u^{(2)} \\ H_{i_1,i_2}(\boldsymbol{u}) &= \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x})(u^{(1)})^2 \\ &+ (\kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x}) + \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})) u^{(1)} u^{(2)} \\ &+ \kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})(u^{(2)})^2 - \kappa_{i_1,i_2}(\boldsymbol{x}) \\ H_{i_1,i_2,i_3}(\boldsymbol{u}) &= \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x})\kappa_{i_3,1}(\boldsymbol{x})(u^{(1)})^3 \\ &+ \big(\kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x})\kappa_{i_3,1}(\boldsymbol{x}) \\ &+ \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})\kappa_{i_3,1}(\boldsymbol{x}) \\ &+ \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x})\kappa_{i_3,2}(\boldsymbol{x})\big)(u^{(1)})^2 u^{(2)} \\ &+ \big(\kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x})\kappa_{i_3,2}(\boldsymbol{x}) \\ &+ \kappa_{i_1,1}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})\kappa_{i_3,2}(\boldsymbol{x}) \\ &+ \kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})\kappa_{i_3,1}(\boldsymbol{x})\big) u^{(1)} (u^{(2)})^2 \\ &+ \kappa_{i_1,2}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x})\kappa_{i_3,2}(\boldsymbol{x})(u^{(2)})^3 \\ &- \big(\kappa_{i_1,i_2}(\boldsymbol{x})\kappa_{i_3,1}(\boldsymbol{x}) + \kappa_{i_1,i_3}(\boldsymbol{x})\kappa_{i_2,1}(\boldsymbol{x}) \\ &+ \kappa_{i_2,i_3}(\boldsymbol{x})\kappa_{i_1,1}(\boldsymbol{x})\big) u^{(1)} \\ &- \big(\kappa_{i_1,i_2}(\boldsymbol{x})\kappa_{i_3,2}(\boldsymbol{x}) + \kappa_{i_1,i_3}(\boldsymbol{x})\kappa_{i_2,2}(\boldsymbol{x}) \\ &+ \kappa_{i_2,i_3}(\boldsymbol{x})\kappa_{i_1,2}(\boldsymbol{x})\big) u^{(2)} \end{aligned} \tag{19}$$

Their orthogonality property can be express as:

$$\int H_{i_1,\ldots,i_k}(\boldsymbol{u}) H_{j_1,\ldots,j_\ell}(\boldsymbol{u}) \phi_{\boldsymbol{x}}(\boldsymbol{u}) d\boldsymbol{u} \tag{20}$$

$$= \begin{cases} [k!]\, \kappa(\boldsymbol{x})_{i_1,j_1} \ldots \kappa(\boldsymbol{x})_{i_k,j_k} & \text{if } \ell = k \\ 0 & \text{otherwise} \end{cases}$$

References

[1] J. Maintz and M. Viergever, “A survey of medical image registration,” *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[2] J. Pluim, J. Maintz, and M. Viergever, “Mutual information based registration of medical images: a survey,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[3] P. Viola and W. Wells, “Alignement by maximization of mutual information,” in *International Conference on Computer Vision, IEEE Computer Society Press, Los Alamitos, CA*, 1995, pp. 16–23.

[4] A. Collignon, D. Vandermulen, P. Suetens, and G. Marchal, “3D Multi-Modality Medical Image Registration Using Feature Space Clustering,” *CVRMed '95: Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, vol. 905, pp. 195–204, 1995.

[5] M. Basseville, “Distance measures for signal processing and pattern recognition,” *Signal Processing*, vol. 18, no. 4, pp. 349–369, 1989.

[6] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An overlap invariant entropy measure of 3d medical image alignment,” *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[7] R. Moddemeijer, “On estimation of entropy and mutual information of continuous distributions,” *Signal Processing*, vol. 16, pp. 233–248, March 1989.

[8] G. Mars, N.J.I. Van Arragon, “Time delay estimation in non-linear systems using average amount of mutual information analysis,” *Signal Processing*, vol. 4, pp. 139–153, 1982.

[9] Y.-I. Moon, B. Rajagopalan, and U. Lall, “Estimation of mutual information using kernel density estimators,” *Phys. Rev. E*, vol. 52, no. 3, pp. 2318–2321, Sept 1995.

[10] P. Comon, “Independent Component Analysis, a new concept ?” *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, April 1994.

[11] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” *Advances in Neural Information Processing Systems, In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, The MIT Press*, vol. 8, pp. 757–763, 1996.

[12] M. M. Van Hulle, “Edgeworth approximation of multivariate differential entropy,” *Neural Comput.*, vol. 17, no. 9, pp. 1903–1910, 2005.

[13] C. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423/623–656, 1948.

[14] O. B.-N. Cox, *Asymptotic techniques for use in statistics*, ser. (Monographs on statistics and applied probability). Chapman and Hall, 1989.

[15] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, “Nonrigid, registration using free-form deformations: Application to breast MR images,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, Aug. 1999. [Online]. Available: 1999/rueckert1999.pdf