A STATISTICAL SIMILARITY MEASURE FOR NON-RIGID MULTI-MODAL IMAGE REGISTRATION

YUNMEI CHEN*, JIANGLI SHI†, MURALI RAO‡, AND JINSEOP LEE§

Abstract. A novel variational model for deformable multi-modal image registration is presented in this work. As an alternative to the models based on maximizing mutual information, the Renyi's statistical dependence measure of two random variables is proposed as a measure of the goodness of matching in our objective functional. The proposed model does not require an estimation of the continuous joint probability density function, which is sensitive to the quantization of the intensities. Instead, it only deals with observed independent samples. Moreover, the theory for reproducing kernel Hilbert spaces is used to simplify the computation. Experimental results with comparisons are provided to show the effectiveness of the model.

Key words. Multi-modal, deformable, image registration, statistical dependence, mutual information

AMS subject classifications. 94A08, 68U10, 65D18

1. Introduction. Image registration is a fundamental problem in computer vision and data analysis. It has been increasingly applied to medical image analysis to assist diagnosis and treatment. When the intensities of two images are linearly related, they can be realigned by direct comparison of the data, such as minimizing their difference under an optimized linear transform, or maximizing their cross correlation (CC). However, in many applications we encounter the problem of matching similar structures between several image modalities, for which intensity comparison is not possible. In particular, in the applications of medical imaging images with different modalities often need to be registered for an accurate fusion of complementary information. For instance, magnetic resonance (MR) images are realigned with CT images of the same subject to precisely localize the tumor to assist surgery planing. Also, functional MR brain images are often realigned to high resolution T1-weighted anatomical images to correct distortions for accurate localization of brain activation maps.

The challenge in matching multi-modal images comes from the fact that there is no direct comparison of intensities. To cope with this difficulty a number of similarity measures based on statistical dependence have been proposed. Viola and Wells [25] and Maes et al. [7] first proposed to use mutual information (MI) as a similarity measure in rigid and affine multi-modal image registration. Since then many variational models based on maximizing MI have been developed to estimate non-rigid deformation field for multimodal image matching, and showed promising results e.g. [11, 5, 4, 8, 12, 6]. Besides MI the correlation ratio (CR) [26, 5], Kullback-Leibler (KL) divergence and joint entropy have also been proposed as similarity/dissimilarity measures for multimodal image registration [2, 5, 3, 8]. The CC is a measure of the linear dependency between the intensities, hence it is constrained for matching images

*358 Little Hall, P. O. Box 118105, Department of Mathematics, University of Florida, Gainesville, FL, 32611 (yun@ufl.edu).

†358 Little Hall, P. O. Box 118105, Department of Mathematics, University of Florida, Gainesville, FL, 32611 (jlshi@ufl.edu).

‡358 Little Hall, P. O. Box 118105, Department of Mathematics, University of Florida, Gainesville, FL, 32611 (rao@ufl.edu).

§358 Little Hall, P. O. Box 118105, Department of Mathematics, University of Florida, Gainesville, FL, 32611 (lee@ufl.edu).

that are linearly related. However, local CC has been suggested in multi-modal image registration [16]. Using this approach, it is implicitly assumed that the pair of images to be aligned have affine dependency locally.

The MI between two random variable X and Y is defined as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \tag{1.1}$$

where

$$H(Z) = -\int p_Z(z) log p_Z(z) dz$$

is the Shannon entropy of a random variable Z with probability density function (pdf) $p_Z(z)$, and

$$H(X,Y) = -\int_{R^N}\int_{R^N} p_{X,Y}(x,y) log p_{X,Y}(x,y) dxdy$$

is the joint entropy of X and Y. Hence, $MI(X,Y)$ can also be written as

$$MI(X,Y) = \int\int p_{X,Y}(x,y) log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} dxdy, \tag{1.2}$$

where $p_{X,Y}(x,y)$ is the joint pdf of X and Y, and $p_X(x)$ and $p_Y(y)$ are the marginal pdfs. Therefore MI can be interpreted as the KL divergence between the joint pdf $p_{X,Y}(x,y)$ and the product of the marginal pdfs $p_X(x)p_Y(y)$. MI is zero if and only if two random variables/vectors are independent. On the other hand, mutual information is maximized when two random variables/vectors are functions of each other. Approaches based on maximizing MI assume implicitly that the more dependent the random variables are, the more information about one ought to be given by the other and vice versa. MI is considered the state-of the-art similarity measure for multi-modal image registration. This measure has been proven to be promising for multi-modal image registration although it has a number of well-known disadvantages. The main difficulty in using this measure is in the need of estimating the joint pdf. Methods for constructing discrete pdfs are histogram based and easy to implement, but their accuracy is limited by the quantization of the intensities. In many efficient optimization schemes the derivative of the pdfs are used to estimate the deformation field, and hence, the construction of continuous pdfs is necessary. Parzen-window density estimator is one of the kernel based methods to estimate continuous pdf. The estimation of continuous joint pdfs makes computations in the algorithms based on maximizing MI complex, and sensitive to the quantization of the intensities. Moreover, theoretically if one random variable is a function of the other, their continuous joint pdf is a delta function, and hence the MI is not well defined.

The aim of this work is to induce an alternative dependence measure to MI, which is capable to handle multi-modal image matching, but no need of actually estimating the joint pdf. In 1959 Rényi [9] proposed a set of postulates for an appropriate measure $Q(X,Y)$ of dependence for two random variables X and Y on a probability space $[\Omega, ß, P]$. This work has drawn much attention ever since. These postulates include

1. $Q(X,Y)$ is defined for any pair of random variable X and Y, neither of them being constant with probability one.
2. $Q(X,Y) = Q(Y,X)$

3. $0 \leq Q(X,Y) \leq 1$,

4. $Q(X,Y) = 0$ if and only if X, Y are independent,

5. $Q(X,Y) = 1$ if $Y = f(X)$ or $X = g(Y)$, where f and g are Borel measurable functions,

6. If the Borel measurable functions f and g map the real axis in one to one way onto itself, $Q(f(X), g(Y)) = Q(X,Y)$.

Then, Rényi showed in this work that one measure satisfying these conditions is the maximum correlation coefficient MCC defined as follows.

$$MCC(X,Y) = \sup_{f,g \in V} CC(f(X), g(Y)), \tag{1.3}$$

where V is the space of all Borel measurable functions with finite positive variance. $CC(f(X), g(Y))$ is the correlation coefficient of $f(X)$ and $g(Y)$, i.e.,

$$CC(f(X), g(Y)) = \frac{Cov(f(X), g(Y))}{\sqrt{Var(f(X))}\sqrt{Var(g(Y))}},$$

where Cov and Var stand for covariance and variance, respectively.

$$\begin{aligned} Cov(f(X), g(Y)) &= E[(f(X) - E[f(X)])(g(Y) - E[g(Y)])] \\ &= E[f(X)g(X)] - E[f(X)]E[g(Y)]. \end{aligned}$$

$$Var(f(X)) = E[(f(X) - E[f(X)])^2] = E[f(X)^2] - E[f(X)]^2,$$

and $Var(g(Y))$ is defined in the same way.

In this work we propose to use Rényi's statistical dependence measure MCC as an alternating measure to MI, CC or joint entropy, that require an estimation of joint pdf, to deal with multi-modal image registration. To use MCC as a statistical dependence measure in registration, we implicitly assume that the smaller 1-$MCC(X,Y)$ is, the more dependent of X and Y are. This is a similar implicit assumption as using MI as a similarity measure in multi-modal image matching. The larger MI is, the more dependent of the two random variables are. This assumption is based on the thought that because of no a clear and simple constructive definition of dependence of two random variables, statistical dependencies are most often defined merely as the absence of independence. However, there is no a clear and simple constructive definition of dependence of two random variables. Hence, statistical dependencies are often defined merely as the absence of independence. Consider the 4th postulate, we implicitly assume that the smaller 1-MMCC(X; Y) is, the more dependent of X and Y are. This is similar to the cases where MI is used as a dependence measure. The larger MI is, the more dependent of the two random variables are. The advantage of using Rényi's statistical dependence measure is that in practice, to compute $MCC(X,Y)$ in (1.3) we do not deal with the measure $P_{X,Y}$ itself, but instead observed samples drawn independently according to it. This avoids the difficulties of continuous joint pdf estimation encountered in the variation models based on maximizing MI.

The framework of variational models for image registration is estimating a displacement vector field $h(x) = x + u(x)$, where $u(x)$ is the corresponding deformation field, by minimizing an energy functional in a set of admissible functions. The energy functional enforces the regularity of the deformation field and statistical dependency

of the deformed image and target image. There have been many regularization methods developed in the literature, such as minimizing a function of the derivatives of u in a suitable space [23], linear elastic regularization [27], and tensor based smoothing [28], where the tensor is designed to prevent the transformation fields from being smoothed across the boundaries of features. Since the focus of this work is on an alternative statistical measure of dependence for image registration, the regularity term is just defined as $R(u) = \int |Du|^2 dx$. In our variational model the statistical dissimilarity measure is defined as $1 - MCC(S(x+u(x)), T(x))$, where $S(x)$ and $T(x)$ are the source and target images, respectively. We would like to point out that MCC is different from CC. $CC(X, Y)$ is restricted for linear dependency of X and Y, while $MCC(X, Y)$ measures the linear dependency for $f(X)$ and $g(Y)$, where f and g are searched from all Borel measurable functions with finite positive variance. Therefore, even X and Y are not lineally related, $MCC(X, Y)$ may still be able to apply.

However, to find the functions f and g over all Borel measurable functions with finite positive variance is difficult. We will show in the next section that the search for f and g can be restricted to a much smaller space in which the supremum is preserved, i.e. the MCC is the same. This smaller space can be a reproducing kernel Hilbert space (RKHS) associated with a reproducing kernel that is continuous, symmetric, positive definite and vanishing at infinity. In this work we choose to use the following Gaussian kernel

$$K(x, y) = K_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{|x-y|^2}{2\sigma^2}\}. \tag{1.4}$$

By the theory of RKHS (see section 2 below), any function f in the RKHS associated with the kernel in (1.4) can be approximated by the sum of finite many functions $\{K(\cdot, \xi)_{\xi \in R}\}$, that is

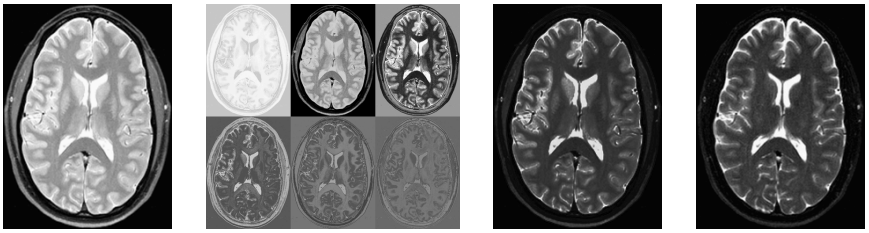
$$f(x) \approx \sum \frac{\alpha_i}{\sqrt{2\pi}\sigma} \exp\{-\frac{|x-\xi_i|^2}{2\sigma^2}\},$$

Therefore, the computation of finding f and g can be reduced to estimating the coefficients α_i and β_j in f and g. The number α_i (or β_j) to be estimated depends on how many ξ_i (or z_j) are chosen.

It is worth pointing out that this linear combination of Gaussian kernel functions plays a significant role in our registration task: (1) By transforming an original image $S(x)$ to $S^f(x) = \sum_{i=1}^m (\alpha_i/\sqrt{2\pi\sigma^2}) \exp\{-|S(x) - \xi_i|^2/2\sigma^2\}$, which is a function of the former, we obtain more control on the intensity distribution of $S(x)$ by manipulating the values of α and ξ. (2) If $T(x)$ and $S(x)$ are taken from same objects with different modalities, by Maximizing some similarity measure between $S^f(x)$ and $T(x)$) with respect to α we can cope with multi-modal images as if they were in the same modal environment. This is illustrated by Fig. 1.1 and the correctness is secured by the following theory. In practice, this process of modality assimilation and the task of image deforming are taking place alternatingly during our registration, as we can see from the following sections.

Without loss generality we can take $i = j$. Then, the MCC of two image I_1 and I_2 can be estimated by

$$MCC(I_1, I_2) = \sup_{\alpha_i, \beta_i} CC(\sum \frac{\alpha_i}{\sqrt{2\pi}\sigma} \exp\{-\frac{|I_1 - \xi_i|^2}{2\sigma^2}\}, \sum \frac{\beta_i}{\sqrt{2\pi}\sigma} \exp\{-\frac{|I_2 - \eta_i|^2}{2\sigma^2}\}).$$



(a) (b) (c) (d)

FIG. 1.1. *The effect of $f(x)$ applied on an image $S(x)$. (a) Image $S(x)$. (b) Images for each term in $S^f(x)$ where $m = 6$, sigma $= 0.4$ and α is optimized to maximize $MCC(S^f(x), T(x))$. (c) Image $S^f(x)$ as the sum of the previous six images. (d) Image $T(x)$*

Roughly speaking, when σ is suitably small, the term $\exp\{-\frac{|I_1-\xi_i|^2}{2\sigma^2}\}$ picks the main information of the voxel intensities of I_1, whose values are close to ξ_i, since this term decays fast outside the interval $(\xi_i - \sigma^2, \xi_i + \sigma^2)$. Then, minimizing $1 - MCC(I_1, I_2)$ aims to find all optimal α_i and β_i to make the CC of $\sum \frac{\alpha_i}{\sqrt{2\pi}} G_\sigma \exp\{-\frac{|I_1-\xi_i|^2}{2\sigma^2}\}$ and $\sum \frac{\beta_i}{\sqrt{2\pi}\sigma} \exp\{-\frac{|I_2-\eta_i|^2}{2\sigma^2}\}$ as close to one as possible.

In the next section, for convenience of readers, we briefly recall the definition and some properties of RKHS used in this work, and show that the supremum in (1.3) is preserved in RKHS. In the third section we describe our model. Then, we provide experimental results with comparisons of MI based algorithms in the fourth section.

2. RKHS and Dense Properties. In this section we show that the supreme in (1.3), in fact, can be taken over a RKHS associated with a symmetric positive definite kernel, such as a Gaussian kernel in (1.4).

2.1. Reproducing kernel Hilbert space. DEFINITION 2.1. *A RKHS is a Hilbert space in which all the point evaluations are bounded linear functionals ([29]).*

Let H be a Hilbert space of functions on a domain E. By the Riesz representation theorem, for every $x \in E$, there exists an element $h_x \in H$, such that the point evaluation $e_x(f)$ satisfying

$$e_x(f) = f(x) = \langle f, h_x \rangle_H, \quad \text{for } \forall f \in H. \tag{2.1}$$

Let $K(x, y) = \langle h_x, h_y \rangle_H$. Then $K(x, y)$ is symmetric and positive definite on $E \times E$. This means that for every n, any $x_1, x_2, ..., x_n \in E$ and any $\alpha_1, \alpha_2, ..., \alpha_n \in R$,

$$\sum_{i,j}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0. \tag{2.2}$$

The equality holds if and only if $\alpha_i's$ are zero. This is equivalent to

$$\int \int K(x, y) m(dx) m(dy) \geq 0, \tag{2.3}$$

for all bounded signed measures m with equality only if $m = 0$. The function K is called a reproducing kernel for H.

From the definition of K and the fact that $h_x \in H$, we have $h_x(y) = \langle h_x, h_y \rangle_H = K(x, y)$. Therefore, one can have

(a). For all $x \in E$, $K(x,\cdot) \in H$,

(b). For all $x \in E$, $\forall f \in H$, $\langle f, K(x,\cdot)\rangle = f(x) \in H$ (This condition is called "reproducing property".

(c). $K(x,y) = \langle K(x,\cdot), K(y,\cdot)\rangle_H$.

On the other hand, the Moore-Aronszajn theorem [30] states that for every symmetric, positive definite function $K : E \times E \to C$, there is a unique Hilbert space H of functions on E for which K is a reproducing kernel. In fact, define $h_x = K(x,\cdot)$. Let H_0 be the linear span of the functions $\{K(x,\cdot)_{x\in E}\}$, and define an inner product on it by

$$\langle \sum_1^n a_i K(x_i,\cdot), \sum_1^m b_j K(y_j,\cdot)\rangle_H = \sum_1^n \sum_1^m a_i \bar{b_j} K(x_i,y_j).$$

Let also H be the completion of H_0 with respect to this inner product. Then, it is not difficult to check that K is a reproducing kernel for H, and H is unique RKHS associated with this kernel.

2.2. Dense properties. Let E be a topological space (in our application $E = R$). Denote by $C_0(E)$ the space of all the continuous functions on E vanishing at infinity with the supreme norm. Suppose that K on $E \times E$ is a function satisfying

(a). K is symmetric, positive definite function (see (2.2 and 2.3).

(b). $K(x,\cdot) \in C_0(E)$ for all $x \in E$.

Then, as stated above, there is a unique Hilbert space, denoted by $H(E)$, associated with this K.

LEMMA 2.2. *$H(E) \cap C_0(E)$ is dense in $C_0(E)$.*

Proof: If $H(E) \cap C_0(E)$ is not dense in $C_0(E)$, by Hahn-Banach theorem there is a function $g \in C_0(E) \setminus H(E)$, and a bounded signed measure $m \in C_0(E)^*$, the dual space of $C_0(E)$, such that

$$\int_E g dm \neq 0, \tag{2.4}$$

but for all $f \in H(E)$

$$\int_E f dm = 0.$$

In particular, for any $x \in E$,

$$\int_E K(x,y) dm_y = 0,$$

and hence,

$$\int K(x,y) dm_x dm_y = 0.$$

Since K is positive definite, this implies $m = 0$, which contradicts (2.4).

LEMMA 2.3. *Let $V_B(E)$ be the space of bounded measurable functions on E. Let also X and Y be two random variable with joint pdf P_{XY}. Then,*

$$\sup_{f,g\in V_B(E)} \mathbf{CC}(f(X),g(Y)) = \sup_{f,g\in C_0(E)} \mathbf{CC}(f(X),g(Y)).$$

Proof: For any functions $f, g \in V_B(E)$ there are two sequences of functions $f_n, g_n \in C_0(E)$, such that for any $q > 0$, as $n \to \infty$,

$$E[|f_n(X) - f(X)|^q] \to 0, \quad E[|g_n(Y) - g(Y)|^q] \to 0, \tag{2.5}$$

By taking $q = 1, 2$ one can get that $E[|f_n(X)|]$ is uniformly bounded in n, then from (2.5) with $q = 2$

$$\begin{aligned} & E[f_n(X)g_n(Y) - f(X)g(Y)] \\ \le\ & E[|f_n(X) - f(X)||g(Y)|] + E[g_n(Y) - g(Y)||f_n(X)|] \\ \le\ & C(E[|f_n(X) - f(X)|^2]^{1/2} + E[g_n(Y) - g(Y)|^2]^{1/2}) \to 0, \end{aligned} \tag{2.6}$$

as $n \to \infty$. Furthermore, from (2.5) with $q = 1$ and (2.6)

$$\begin{aligned} E[(f_n(X) - E[f_n(X)])(g_n(Y) - E[g_n(Y)])] &= E[f_n(X)g_n(Y)] - E[f_n(X)]E[g_n(Y))] \\ \to E[f(X)g(Y)] - E[f(X)]E[g(Y))] &= E[(f(X) - E[f(X)])(g(Y) - E[g(Y)])]. \end{aligned}$$

Similarly, we can show that as $n \to \infty$,

$$\begin{aligned} E[(f_n(X) - E[f_n(X)])^2] &= E[f_n(X)^2] - E[f_n(X)]^2 \\ \to E[f(X)^2] - E[f(X)]^2 &= E[(f(X) - E[f(X)])^2], \end{aligned}$$

and

$$E[(g_n(Y) - E[g_n(Y)])^2] \to E[(g(Y) - E[g(Y)])^2].$$

Then, by the definition of CC, we get

$$CC(f_n(X), g_n(Y)) \to CC(f(X), g(Y)).$$

Thus,

$$\sup_{f,g \in V_B(E)} CC(f(X), g(Y)) = \sup_{f,g \in C_0(E)} \mathbf{CC}(f(X), g(Y)).$$

Lemma 2.4. *Let $V(E)$ be the space of all Borel measurable functions on E with finite positive variance, and X and Y be the same as in Lemma 2. Then*

$$\sup_{f,g \in V_B(E)} \mathbf{CC}(f(X), g(Y)) = \sup_{f,g \in V(E)} \mathbf{CC}(f(X), g(Y)).$$

Proof: For any function $f \in V(E)$, define f_n as

$$f_n = f, \ if \ |f| \le n, \ \ and \ \ f_n = nsign(f), \ if |f| > n.$$

Then, $f_n \in V_B(E)$, f_n converges to f pointwisely and as $n \to \infty$,

$$E[f_n(X)] \to E[f(X)]. \tag{2.7}$$

Moreover, $f_n^2 \le f^2$, and

$$E[f_n^2] \le E[f^2]. \tag{2.8}$$

From (2.8),

$$E[(f_n(X) - E[f_n(X)])^2] = E[f_n(X)^2] - E[f_n(X)]^2 \le E[f_n(X)^2] \le E[f(X)^2].$$

Now by using the dominated convergence theorem with (2.7) we conclude that as $n \to \infty$,

$$E[(f_n(X) - E(f_n(X)))^2] \to E[(f(X) - E(f(X)))^2].$$

Similarly, we can have

$$E[(g_n(Y) - E(g_n(Y)))^2] \to E[(g(Y) - E(g(Y)))^2].$$

$$E[(f_n(X)-E(f_n(X)))(g_n(Y)-E(g_n(Y)))] \to E[(f(X)-E(f(X)))(g(Y)-E(g(Y)))].$$

Then by the definition of CC, we get

$$CC(f_n(X), g_n(Y)) \to CC(f(X), g(Y)).$$

The lemma follows from this immediately.

The combination of these three lemmas gives the following theorem:

THEOREM 2.5.

$$\sup_{f,g \in H(E) \cap C_0(E)} CC(f(X), g(Y)) = \sup_{f,g \in V(E)} CC(f(X), g(Y)).$$

3. Proposed model. Consider that pixel intensities of each image involved in the process of registration are independent, and identically distributed data drawn from a particular distribution. Our proposed variational model estimates a deformation field by minimizing an energy functional consisting of two terms. One is a regularization term for the deformation field. The other term measures the goodness of matching, which is formulated by using the Rényi's statistical dependence measure MCC between the deformed image and the target image. Moreover, We introduce the reproducing kernel Hilbert spaces associated with Gaussian kernels to simplify the computation in finding the supremum in the definition of MCC.

Let $S(x)$ and $T(x)$ be the source image and target image on the image domain Ω, respectively, and $u(x)$ be the deformation field that deforms S to T. First we show how to estimate the CC between the deformed image $S_u(x) = S(x + u(x))$ and target image $T(x)$ using the theory of RKHS. Let $\mathcal{H}$ be the RKHS associated with the following Gaussian kernel:

$$K(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-y)^2}{2\sigma^2}\}.$$

By the theorem in the previous section the supremum of $CC(f(S_u), g(T))$ in (1.3) can be taken over $\mathcal{H}$. For $f, g \in \mathcal{H}$, $f(S_u(x))$ and $g(T(x))$, denoted by $S_u^f(x)$ and $T^g(x)$ respectively, can be approximated by taking

$$S_u^f(x) = \sum_{i=1}^{m} \alpha_i K(S_u(x), \xi_i), \qquad T^g(x) = \sum_{j=1}^{n} \beta_i K(T(x), z_j) \tag{3.1}$$

for some integers n and m, where ξ_i, z_j, α_i and β_j are real numbers. Hence we can estimate the mean of $S_u^f(x)$ and $T^g(x)$ by

$$\mu_1(u,\alpha) = \frac{1}{|\Omega|}\int_\Omega \sum_{i=1}^{n} \alpha_i K(S_u(x),\xi_i)dx, \qquad \mu_2(\beta) = \frac{1}{|\Omega|}\int_\Omega \sum_{j=1}^{m} \beta_i K(T(x),z_j)dx,$$

respectively. The variance and the covariance of $S_u^f(x)$ and $T^g(x)$ are then estimated as:

$$\begin{aligned}
\nu_1(u,\alpha) &= \frac{1}{|\Omega|}\int_\Omega \{\sum_{i=1}^{n} \alpha_i K(S_u(x),\xi_i)\}^2 dx - \mu_1(u,\alpha)^2,\\
\nu_2(\beta) &= \frac{1}{|\Omega|}\int_\Omega \{\sum_{j=1}^{m} \beta_i K(T(x),z_j)\}^2 dx - \mu_2(\beta)^2,\\
\nu_{12}(u,\alpha,\beta) &= \frac{1}{|\Omega|}\int_\Omega \{\sum_{i=1}^{n} \alpha_i K(S_u(x),\xi_i) - \mu_1(u,\alpha)\}\cdot\{\sum_{j=1}^{m} \beta_i K(T(x),z_j) - \mu_2(\beta)\}dx,
\end{aligned} \tag{3.2}$$

respectively. Thus

$$CC(S_u^f, T^g) = \frac{\nu_{12}(u,\alpha,\beta)}{\sqrt{\nu_1(u,\alpha)}\cdot\sqrt{\nu_2(\beta)}}, \tag{3.3}$$

and

$$MCC(S_u,T) = \sup_{f,g} CC(S_u^f, T^g) = \max_{\alpha,\beta} \frac{\nu_{12}(u,\alpha,\beta)}{\sqrt{\nu_1(u,\alpha)}\cdot\sqrt{\nu_2(\beta)}},$$

Now we register the image $S(x)$ to $T(x)$ by solving the following minimization problem with respect to the deformation field $u(x)$:

$$\min_{u,\alpha,\beta} \quad \lambda\int_\Omega |\nabla u(x)|^2 dx + |\Omega|(1 - MCC(S_u,T))^p, \tag{3.4}$$

where parameter λ balances the regularization of the deformation field and goodness of the matching

$$\mathcal{R}(u) = \int_\Omega |\nabla u(x)|^2 dx$$

and the parameter p in the second term of (3.4) influences the speed of $MCC(S_u,T)$ increasing towards one. Fig. 3 depicts the graph of $\eta(x) = (1-x)^p$, $x \le 1$ for different values of p. It can be seen that when x approaches 1, the derivative of $\eta(x)$ at x is increasing if $0 < p < 1$. This increment of derivative provide us with efficiency for obtaining the optimized $u(x)$ in the minimization problem. This can also be seen from the Euler-Lagrange equation for $u(x)$. The constant $|\Omega|$ balances the scale level between those two terms.

By minimizing this energy we can get a regularized deformation field, which forces the deformed image and the target image to have maximum dependence in terms of Rényi's statistical dependence measure MCC. In the extreme case that the deformed image is a function of the target image or vise versa, $MCC(S,T) = 1$, and hence, the second term is zero.

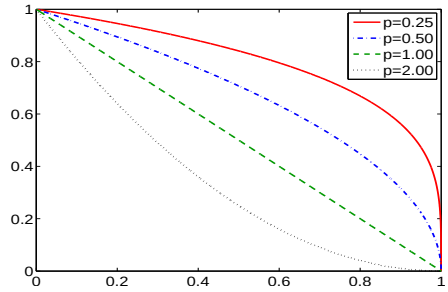


FIG. 3.1. *Graph of* $\eta(x) = (1-x)^p, 0 \le p \le 2$ *for different values of* p.

Using notation above, the minimization problem (3.4) can be rewritten as:

$$\min_u \{\lambda \mathcal{R}(u) + |\Omega| \max_{\alpha,\beta} \mathcal{J}(u,\alpha,\beta)\}. \tag{3.5}$$

where

$$\mathcal{J}(u,\alpha,\beta) = (1 - \frac{\nu_{12}(u,\alpha,\beta)}{\sqrt{\nu_1(u,\alpha,\beta)} \cdot \sqrt{\nu_2(\beta)}})^p. \tag{3.6}$$

4. The iterative algorithm. An alternating algorithm is used here to find the minimizer $u(x)$, α, β. Starting from an initial $u_0(x)$, for instance, $u_0(x) = 0$, this method computes a sequence of:

$$\left[\alpha^{(1)}, \beta^{(1)}\right], u^{(1)}, \left[\alpha^{(2)}, \beta^{(2)}\right], u^{(2)}, \cdots, \left[\alpha^{(k)}, \alpha^{(k)}\right], u^{(k)}, \cdots$$

such that

$$\left[\alpha^{(k)}, \beta^{(k)}\right] = \arg\min_{\alpha,\beta} \mathcal{J}(u^{(k-1)}, \alpha, \beta), \qquad u^{(k)} = \arg\min_u \lambda \mathcal{R}(u) + \mathcal{J}(u, \alpha^{(k)}, \beta^{(k)}).$$

First, from (3.3) we note that searching for α and β which minimize $\mathcal{J}(u,\alpha,\beta)$ is equivalent to searching for α and β which maximize the following:

$$CC(S_u^f, T^g) = \frac{\nu_{12}(u,\alpha,\beta)}{\sqrt{\nu_1(u,\alpha)} \cdot \sqrt{\nu_2(\beta)}},$$

where

$$f(t) = \sum_{i=1}^{m} \alpha_i K(t, \xi_i), \qquad g(t) = \sum_{j=1}^{n} \beta_i K(t, \eta_i).$$

In the discrete version, arranging $S_u^f(x), T^g(x), p_i = K(S_u(x), \xi_i)$ and $q_j = K(T(x), z_j)$ column-wisely in (3.1) and (3.2) as vectors of length L, which is equal to the image size of $S(x)$ and $T(x)$. Then $S_u^f = P\alpha$ and $T^g = Q\beta$, where $\alpha = (\alpha_1, \ldots, \alpha_m)$, $\beta = (\beta_1, \ldots, \beta_n)$, $P = (p^{ij})_{L\times m} = (p_1, p_2, \cdots, p_m)$ and $Q = (q^{ij})_{L\times n} = (q_1, q_2, \cdots, q_n)$. Now it's readily to check that

$$\frac{\nu_{12}(u,\alpha,\beta)}{\sqrt{\nu_1(u,\alpha)} \cdot \sqrt{\nu_2(\beta)}} = \frac{< P_0\alpha, Q_0\beta >}{|P_0\alpha||Q_0\beta|},$$

where $P_0 = (p_0^{ij})_{L\times m} = (p^{ij} - \frac{1}{L}\cdot\sum_{i=1}^{L} p^{ij})_{L\times m}$ and $Q_0 = (q_0^{ij})_{L\times n} = (q^{ij} - \frac{1}{L}\cdot\sum_{i=1}^{L} q^{ij})_{L\times n}$. Let $\hat{P}$ be a matrix whose columns are the orthonormal basis for the vector space spanned by the columns of P_0 and $\hat{Q}$ be a matrix obtained similarly from Q_0. Since maximizing $CC(S_u^f, T^g)$ is equivalent to maximizing $< P_0\alpha, Q_0\beta >$ with respect to α and β under the condition $|P_0\alpha| = |Q_0\beta| = 1$, it is also equivalent to maximizing $< \hat{P}\hat{\alpha}, \hat{Q}\hat{\beta} >$ with respect to $\hat{\alpha}$ and $\hat{\beta}$ under the condition $|\hat{\alpha}| = |\hat{\beta}| = 1$. Since $< \hat{P}\hat{\alpha}, \hat{Q}\hat{\beta} >= \hat{\alpha}^T\hat{P}^T\hat{Q}\hat{\beta}$, we carry out a singular value decomposition for $\hat{P}^T\hat{Q} = U\Lambda V^T$ where U and V are unitary matrices and Λ is a diagonal matrix whose diagonal entries are nonnegative and are listed in a descending order from the upper left to the lower right. Hence, we can choose $\hat{\alpha}$ and $\hat{\beta}$ such that $\hat{\alpha}^T U = (1, 0, 0, \cdots, 0)$ and $\hat{\beta}^T V = (1, 0, 0, \cdots, 0)$, which implies $\hat{\alpha} = U\cdot(1, 0, 0, \cdots, 0)^T$ and $\hat{\beta} = V\cdot(1, 0, 0, \cdots, 0)^T$. Finally, we can just set $S_u^{f^*} = P_0\alpha^* = \hat{P}U\cdot(1, 0, 0, \cdots, 0)^T$ and $T^{g^*} = Q_0\beta^* = \hat{Q}V\cdot(1, 0, 0, \cdots, 0)^T$ which will not affect the subsequent results. As have shown here that we did not explicitly get the solution for the minimizer α^* and β^*, but it implies the existence of those minimizers and we get $S_u^{f^*}(x)$ and $T^{g^*}(x)$ as we need.

Next, Let α^* and β^* be the minimizers from the previous step, and let the $\mu_1, \mu_2, \nu_1, \nu_2 and \nu_1 2$ in (3.2) be $\mu_1(u) = \mu_1(u, \alpha^*)$, $\mu_2 = \mu_2(\beta^*)$, $\nu_1(u) = \nu_1(u, \alpha^*)$, $\nu_2 = \nu_2(\beta^*)$, $\nu_{12}(u) = \nu_{12}(u, \alpha^*, \beta^*)$, and

$$\mathcal{J}^*(u) \triangleq (1 - \frac{\nu_{12}(u)}{\sqrt{\nu_1(u)}\cdot\sqrt{\nu_2}})^p.$$

By computing the first variation of $\lambda\mathcal{R}(u) + |\Omega|\mathcal{J}^*(u)$ with respect to $u(x)$, we get the following Euler-Lagrange (EL) equation:

$$-2\lambda\triangle u(x) - F(x, u) = 0$$

where

$$\begin{aligned} F(x,u) =& p\left(1 - \frac{\nu_{12}(u)}{\nu_1^{1/2}(u)\cdot\nu_2^{1/2}}\right)^{p-1} \\ & \cdot\left\{\frac{T^g(x) - \mu_2}{\nu_1^{1/2}(u)\cdot\nu_2^{1/2}} - \frac{S_u^f(x) - \mu_1(u)}{\nu_1^{3/2}(u)\cdot\nu_2^{1/2}}\cdot\nu_{12}(u)\right\}\cdot\nabla S_u^f(x). \end{aligned}$$

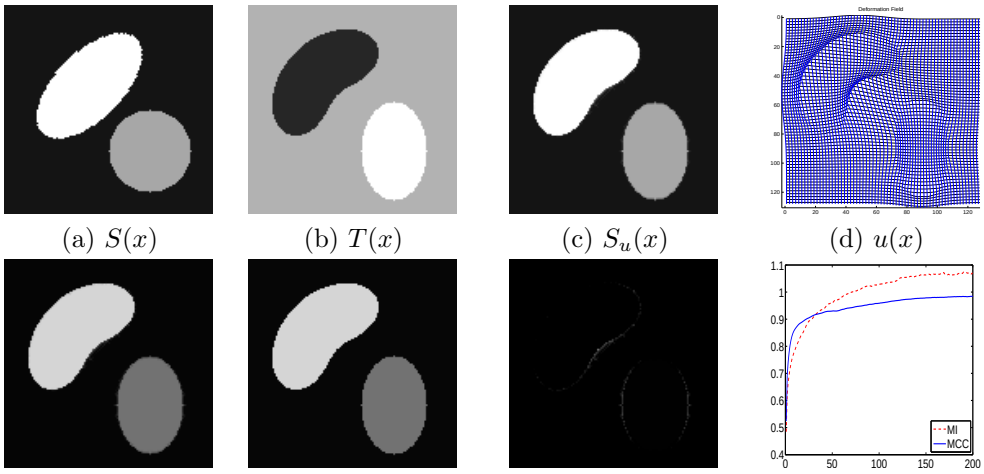
The corresponding evolution equation to the EL equation of $u(x, t)$ is

$$\frac{\partial u(x,t)}{\partial t} = 2\lambda\triangle u(x,t) + p[1 - MCC(S_{u_t}(x), T(x))]^{p-1}F(x,u).$$

5. Experimental Results. The algorithm described in Section 4 is carried out in the following experiments. The Efficiency of the algorithm is verified by applying our model on a variety of image data. The Robustness to noise and stableness to the choice of initial u_0 are also demonstrated. Since the objective functional is not convex, we use a multi-resolution approach and apply the gradient descent to a set of smoothed and subsampled images. This coarse-to-fine strategy helps avoiding the deformation field from getting stuck at local minima of the objective functional.

5.1. Efficiency Validation. To verify the effectiveness of the proposed model we first applied model (3.4) to a pair of synthetic images shown as the first two images

in Fig. 5.1. The source image $S(x)$ shown as the first image consists of two objects with their intensities 1.00 and 0.70, and a background with intensity 0.15. The target image $T(x)$ shown as the second image is generated in a similar manner. But the two corresponding objects in $T(x)$ have different shapes from those in $S(x)$, and their intensities are 0.08 and 1.0000, respectively. The background intensity in $T(x)$ is 0.65. Therefore, the intensities of $S(x)$ and $T(x)$ are far from linearly related.



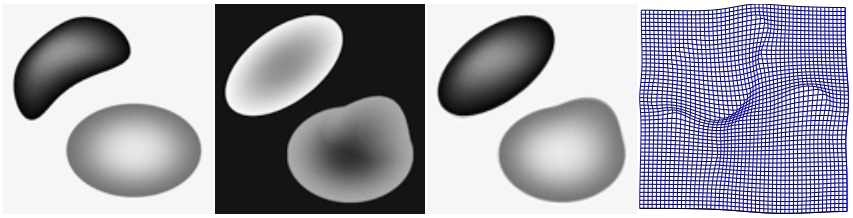
(a) $S(x)$ (b) $T(x)$ (c) $S_u(x)$ (d) $u(x)$

(e) $S_u^{f^*}(x)$ (f) $T^{g^*}(x)$ (g) $|S_u^{f^*}(x) - T^{g^*}(x)|$ (h) MCC and MI

FIG. 5.1. *Registration results for synthetic images with homogeneous regions. (a) Source image $S(x)$. (b) Target image $T(x)$. (c) Deformed image $S_u(x)$. (d) Deformation field $u(x)$ applied to a regular grid. (e) Image $S_u^{f^*}(x)$ with optimized f^*. (f) Image $T^{g^*}(x)$ with optimized g^*. (g) Image $|S_u^f(x) - T^g(x)|$. (h) $MCC(S_u(x), T(x))$ and $MI(S_u(x), T(x))$ vs. iterations.*

Applying the flow of the EL equation associated with model (3.4) on $S(x)$ and $T(x)$ with the parameters $\lambda = 0.06$, $dt = 15$, $\sigma = 0.10$, and $\xi_i = \eta_i = i/5$ ($i = 0, 1, \cdots, 5$) we get the deformed image $S_u(x)$ and the optimized deformation field $u(x)$. At the same time we obtain the optimal $S_u^f(x)$, $T^g(x)$ and $S_u^f(x) - T^g(x)$ associated with $u(x)$ and the optimized α and β as shown in (3.1). One can see that the images $S_u^f(x)$ and $T^g(x)$ can be considered as the same modality images although the deformed image and target image are not. The difference between $S_u^f(x)$ and $T^g(x)$ shows that their intensities are very close to each other. This is an indication of a good matching. We also plot the graph of MCC between $S_u(x)$ and $T(x)$ (or equivalently, CC between $S_u^{f^*}(x)$ and $T^{g^*}(x)$) vs. iterations , which shows that the MCC quickly increases to 0.999. Furthermore we studied the question: is the increasing of MCC consistent with the increasing of the MI between $S_u(x)$ and $T(x)$? To answer this question we also plot the graph of $MI(S_u(x), T(x))$ vs. iterations in the same figure as the graph of $MCC(S_u(x), T(x))$. Comparing these two graphs, one can see the consistency of maximizing Rényi's statistical dependence measure and mutual information approaches in multi-modal image registration.

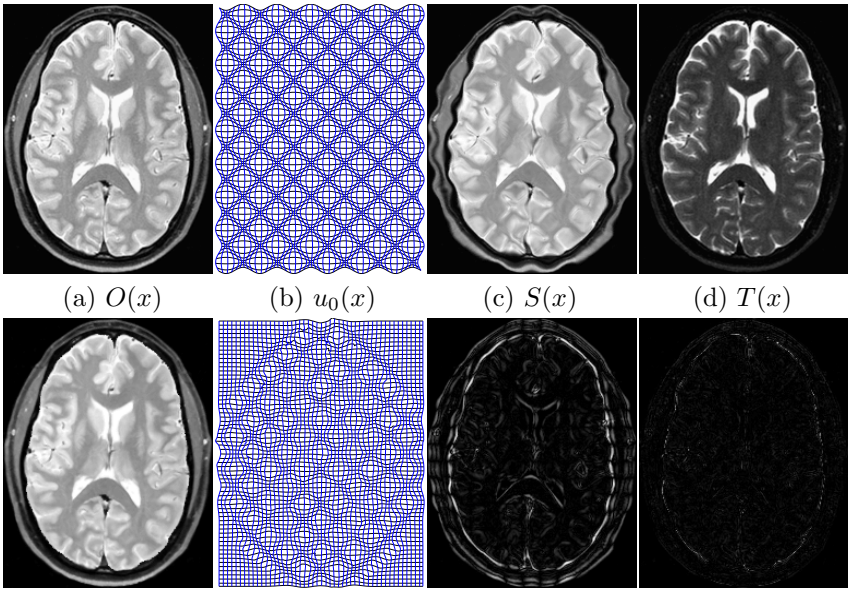
Next, we apply our model on Synthetic images with inhomogeneous regions shown as the first two images in Fig. 5.2. The parameters we choose here are the same as those in the previous experiment. The result shows almost perfectly matching for the objects in images, despite the fact that regional inhomogeneities increase the difficulty a lot in image registration, especially when the deformation is large.



(a) $S(x)$ (b) $T(x)$ (c) $S_u(x)$ (d) $u(x)$

FIG. 5.2. *Registration result for synthetic images with regional inhomogeneities. (a) Source image $S(x)$. (b) Target image $T(x)$. (c) Deformed image $S_u(x)$. (d) Deformation field $u(x)$ applied to a regular grid.*

Finally, we exam the efficiency of our algorithm on a set of PD and T2 brain images where the deformation field is known. The original T2 image $O(x)$ is perfectly aligned with a PD image $T(x)$. The image $S(x)$ to be registered here is obtain by applying a predetermined deformation field $u_0(x)$ on an original T2 brain image, and the target image $T(x)$ is an PD brain image which coincides with $O(x)$ accurately. We choose parameters $\lambda = 0.06$, $dt = 15$, $\sigma = 0.08$, and $\xi_i = \eta_i = i/10$ $(i = 0, 1, \cdots, 10)$.



(a) $O(x)$ (b) $u_0(x)$ (c) $S(x)$ (d) $T(x)$

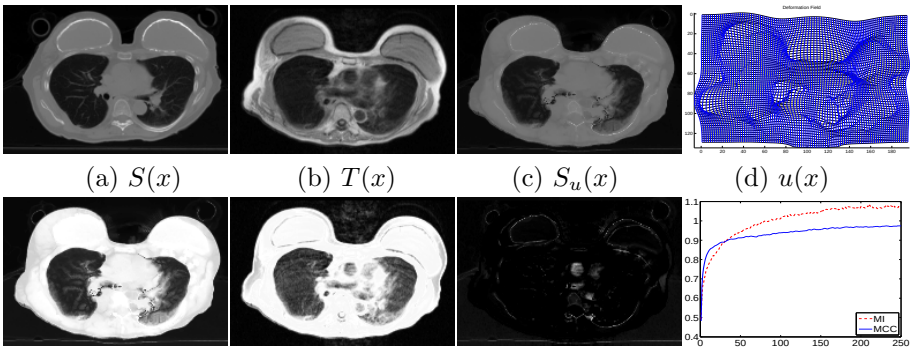
(e) $S_u(x)$ (f) $u(x)$ (g) $|O(x) - S(x)|$ (h) $|O(x) - S_u(x)|$

FIG. 5.3. *Registration results for T2 and PD brain images with predetermined deformation field $u_0(x)$. (a) Original T2 image $O(x)$. (b) Predetermined deformation field $u_0(x)$ applied to a regular grid. (c) Source image $S(x)$ obtained by applying $u_0(x)$ on $O(x)$. (d) PD Target image $T(x)$. (e) Deformed image $S_u(x)$. (f) Deformation field $u(x)$ applied to a regular grid. (g) Image $|S(x) - O(x)|$. (h) Image $|S_u(x) - O(x)|$.*

The results shown in Fig. 5.3 indicate that the deformed image $S_u(x)$ is almost perfectly matched with the original image $O(x)$, which can be seen from (c) as the difference image $|S(x) - O(x)|$. However, the deformation field $u(x)$ we get is as the same as the predetermined one $u_0(x)$. This is reasonable because the deformation

field for matching $S(x)$ and $T(x)$ may not be unique, and it depends on the regularity term in the objective functional. On the other hand, the predetermined $u_0(x)$ may not minimize the objective functional.

5.2. Robustness test on registering CT and MR images. This experiment aims to validate the effectiveness of the proposed model in real medical images and to test its robustness to noise. We apply our model to a CT lung image and MR lung image shown as the first two images in Fig.2. We use $\lambda = 0.06$, $dt = 15$, $\sigma = 0.18$, and $\xi_i = \eta_i = i/5$ $(i = 0, 1, \cdots, 10)$ as the parameters in this experiment.



(a) $S(x)$ (b) $T(x)$ (c) $S_u(x)$ (d) $u(x)$

(e) $S_u^{f^*}(x)$ (f) $T^{g^*}(x)$ (g) $|S_u^{f^*}(x) - T^{g^*}(x)|$ (h) MCC and MI

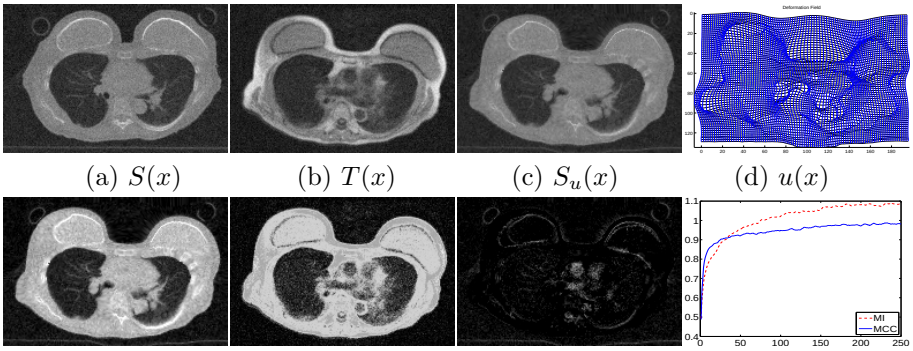
FIG. 5.4. *Registration results CT and MR images. (a) CT source image $S(x)$. (b) MR target image $T(x)$. (c) Deformed image $S_u(x)$. (d) Deformation field $u(x)$ applied to a regular grid. (e) Image $S_u^{f^*}(x)$ with optimized f^*. (f) Image $T^{g^*}(x)$ with optimized g^*. (g) Image $|S_u^f(x) - T^g(x)|$. (h) $MCC(S_u(x), T(x))$ and $MI(S_u(x), T(x))$ vs. iterations.*

In a pattern similar to Fig. 5.1 the difference between $S_u^f(x)$ and $T^g(x)$ in Fig. 5.4 shows that their intensities are close to each other and the structure of $S(x)$ is preserved in $S_u^{f^*J}(x)$. The graphs of MCC and MI between $S_u(x)$ and $T(x)$ are shown as in (h). From this figure one can see that the $MCC(S_u(x), T(x))$ rises to 0.98, and hence, the alignment is desirable. Moreover, the increasing trend in MCC is consistent with that in MI as observed in the first experiment.

The aim of the next experiment is to test the robustness of the proposed model to noise. For this purpose we applied model (3.4) with the same parameters as in the previous experiment to a pair of noisy CT and MR lung images shown as the first two images in Fig. 5.5. These two images are generated by adding a Gaussian noise of 0 mean and 0.1 variance to the images used in the previous experiment with the intensity range of [0,1]. Model (3.4) still works well in this case. The MCC between $S_u(x)$ and $T(x)$ is up to 0.97, and the difference between $S_u(x)$ and $T(x)$ is small. The consistency of the increasing trend for $MCC(S_u(x), T(x))$ and $MI(S_u(x), T(x))$ are still preserved.

6. Extension to local MCC. In this section, we extend our model to the local version, which provides us the flexibility to cope with non-stationarity in intensity distributions of the images to be registered. To do this, we estimate for the similarity measure between the neighborhoods of each point $x_0 \in \Omega$ in two images. This is achieved by weighing our previous estimates of the means, variances and covariances with a normalized Gaussian kernel of variance θ.

More precisely, the mean of $S_u(x)$ and $T^g(x)$ in the θ neighborhood of z can be



(a) $S(x)$ (b) $T(x)$ (c) $S_u(x)$ (d) $u(x)$

(e) $S_u^{f^*}(x)$ (f) $T^{g^*}(x)$ (g) $|S_u^{f^*}(x) - T^{g^*}(x)|$ (h) MCC and MI

FIG. 5.5. *Registration results CT and MR images. (a) CT source image $S(x)$. (b) MR target image $T(x)$. (c) Deformed image $S_u(x)$. (d) Deformation field $u(x)$ applied to a regular grid. (e) Image $S_u^{f^*}(x)$ with optimized f^*. (f) Image $T^{g^*}(x)$ with optimized g^*. (g) Image $|S_u^f(x) - T^g(x)|$. (h) $MCC(S_u(x), T(x))$ and $MI(S_u(x), T(x))$ vs. iterations.*

estimated by:

$$\mu_1^u(z) = \int_\Omega S_u^{f_z}(x) G_\theta(x-z)dx, \quad \mu_2(z) = \int_\Omega T^{g_z}(x) G_\theta(x-z)dx.$$

The variance of $S_u^f(x)$, $T(x)$ and the covariance of those two in the neighborhood of z are then computed by:

$$\begin{aligned}
\nu_1^u(z) &= \int_\Omega S_u^{f_z}(x)^2 G_\theta(x-z)dx - \mu_1^u(z)^2, \\
\nu_2(z) &= \int_\Omega T^{g_z}(x)^2 G_\theta(x-z)dx - \mu_2(z)^2, \\
\nu_{12}^u(z) &= \int_\Omega S_u^{f_z}(x) T^{g_z}(x) G_\theta(x-z)dx - \mu_1^u(z)\mu_2(z),
\end{aligned}$$

which leads to the local version of the model (3.4) based on Rényi's statistical measure:

$$\min_{u(x),\alpha(x),\beta(x)} \lambda \int_\Omega |\nabla u(x)|^2 dx + |\Omega| \int_\Omega (1 - \frac{\nu_{12}(x)}{\sqrt{\nu_1(x)} \cdot \sqrt{\nu_2(x)}})^p dx,$$

where $\alpha = (\alpha_x)_{m\times n}$ is a matrix whose entries are a collection of parameters α_x to be optimized:

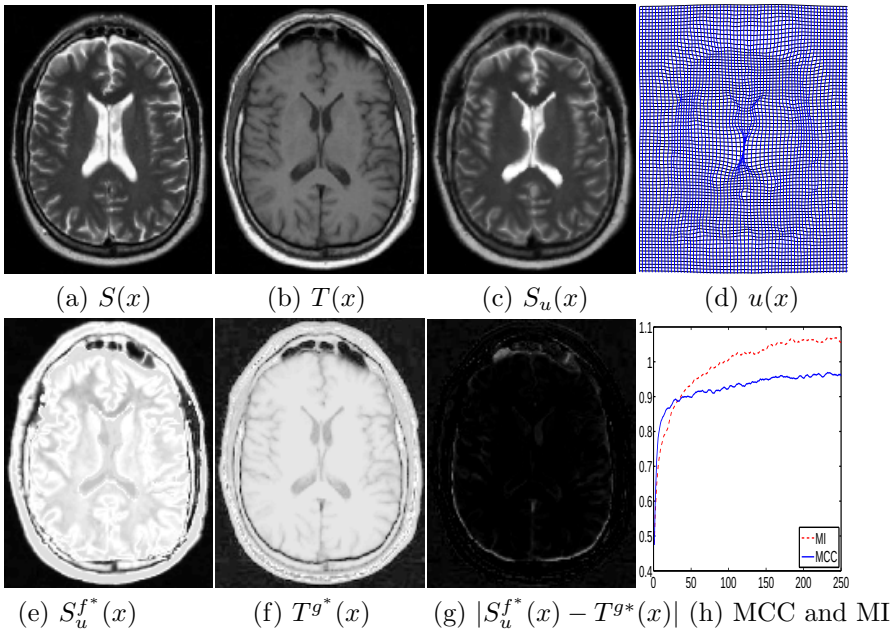
$$\alpha_x = \arg\min_{\alpha_x} \mathcal{J}_{Loc}(u, \alpha_x, x)$$

Hence, the E-L equation for the local version of the proposed model is

$$-2\lambda \triangle u(x) - p\left(1 - \int_\Omega \frac{\nu_{1,2}^u(z)}{[\nu_1^u(z)]^{1/2} \cdot [\nu_2(z)]^{1/2}} dz\right)^{p-1} \cdot \int_\Omega G_\theta(x-z) F(x,z,u) dz = 0$$

$$F(x,z,u) = \left\{ \frac{T^{g_z}(x) - \mu_2(z)}{[\nu_1^u(z)]^{1/2} \cdot [\nu_2(z)]^{1/2}} - \frac{S_u^{f_z}(x) - \mu_1^u(z)}{[\nu_1^u(z)]^{3/2} \cdot [\nu_2(z)]^{1/2}} \cdot \nu_{1,2}^u(z) \right\} \cdot \nabla S_u^{f_z}(x).$$

Fig. 6.1 shows the result of registering T1 and T2 images using the above local version (6.1). The aim of this experiment is to test if the local version results in a good registration for images with more complex intensities. We applied the local version to align a pair of T1 and T2 brain images, for which a relatively large deformation is required. The images shown as the first two images in Fig. 6.1 have complex and nonuniform intensities, and the deformation from one to the other can not be too small. The global version of model (3.4) can't provide a desirable registration result. We observed that MCC is not sensitive to the mismatching when the intensities and structures in the images are complex. Therefore we don't have a strong forcing term to drive the deformation field to get an optimal alignment.



(a) $S(x)$ (b) $T(x)$ (c) $S_u(x)$ (d) $u(x)$

(e) $S_u^{f^*}(x)$ (f) $T^{g^*}(x)$ (g) $|S_u^{f^*}(x) - T^{g*}(x)|$ (h) MCC and MI

FIG. 6.1. *Registration results T1 and T2 images. (a) T2 source image $S(x)$. (b) T1 target image $T(x)$. (c) Deformed image $S_u(x)$. (d) Deformation field $u(x)$ applied to a regular grid. (e) Image $S_u^{f^*}(x)$ with optimized f^*. (f) Image $T^{g^*}(x)$ with optimized g^*. (g) Image $|S_u^f(x) - T^g(x)|$. (h) $MCC(S_u(x), T(x))$ and $MI(S_u(x), T(x))$ vs. iterations.*

The parameters we used in this experiment are $\lambda = 0.2$, $dt = 1$, $\sigma = 0.15$, and $\xi_i = \eta_i = i/5$ $(i = 0, 1, \cdots, 10)$. As shown in Fig.4 the difference between $S_u^f(x)$ and $T^g(x)$ is small, and MCC can reach up to 0.98. The deformation is relatively large. The increasing of MCC vs. iterations is consistent with that of MI. All these indicate the effectiveness of the proposed model even for complex images.

6.1. Comparisons with other deformable models. In this section, we make comparisons with other two deformable models. One is to maximizing the cross correlation (CC) and the other is to maximizing the correlation ratio (CR). We implemented all those three models in the local version on a pair of T1 and T2 brain images (Fig. 6.2). We use mutual information between the deformed image and the target image as criterion to measure the goodness of matching. Table. 6.1 gives the parameters and resulting data.

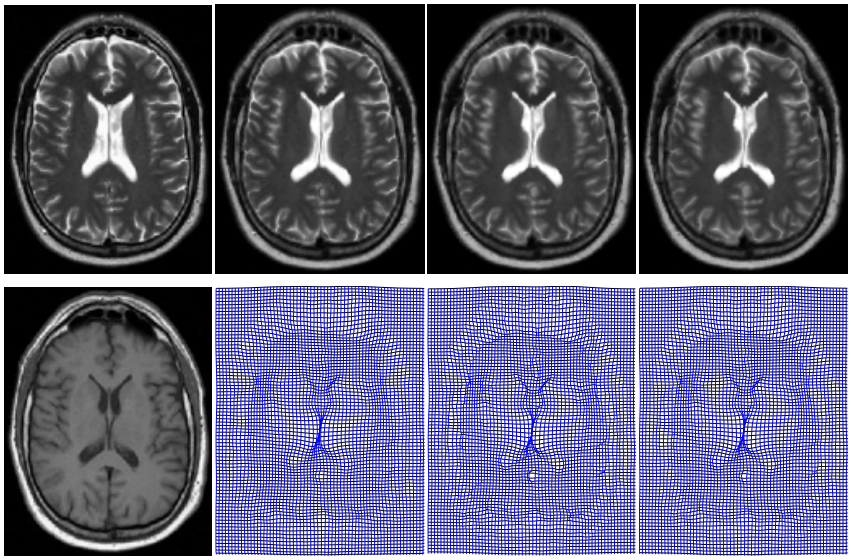


FIG. 6.2. *Comparisons with other models. First column: Source image $S(x)$ and target image $T(x)$. From 2nd -5th columns resulting from CC, CR and MCC respectively: Top: deformed images; Bottom: deformation fields.*

TABLE 6.1
Parameters and results for different models.

–	σ	θ	λ	step size	iteration number	mutual information
CC	0.15	-	0.2	1	300	1.2343
CR	0.15	-	0.2	1	100	1.4834
MCC	0.15	6	0.2	1	100	1.5103

Appendix A. E-L equation of for the global version. By adopting the notations in the paper, we have the following estimates for the means, variances and covariances:

$$S^f_{u+\eta v}(x) = f(S(x+u(x)+\eta v(x))), \qquad T^g(x) = g(T(x)),$$

$$\mu_1^{u+\eta v} = \frac{1}{|\Omega|}\int_\Omega S^f_{u+\eta v}(x)dx, \qquad \mu_1^u = \frac{1}{|\Omega|}\int_\Omega S^f_u(x)dx, \qquad \mu_2 = \frac{1}{|\Omega|}\int_\Omega T^g(x)dx.$$

$$\nu_1^{u+\eta v} = \frac{1}{|\Omega|}\int_\Omega (S^f_{u+\eta v}(x) - \mu_1^{u+\eta v})^2 dx, \quad \nu_1^u = \frac{1}{|\Omega|}\int_\Omega (S^f_u(x) - \mu_1^u)^2 dx,$$

$$\nu_2 = \frac{1}{|\Omega|}\int_\Omega (T^g(x) - \mu_2)^2 dx,$$

$$\nu_{1,2}^{u+\eta v} = \frac{1}{|\Omega|}\int_\Omega (S^f_{u+\eta v}(x) - \mu_1^{u+\eta v})(T^g(x) - \mu_2)dx,$$

$$\nu_{1,2}^{u} = \frac{1}{|\Omega|}\int_\Omega (S^f_{u}(x) - \mu_1^{u})(T^g(x) - \mu_2)dx.$$

It is easy to see that

$$\frac{d}{d\eta}\Big|_{\eta=0} S^f_{u+\eta v}(x) = \nabla S^f_u(x)\cdot v(x), \qquad \frac{1}{|\Omega|}\int_\Omega \mu_1^{u+\eta v}(T^g(x)-\mu_2)dx = 0.$$

Hence,

$$\begin{aligned}\frac{d}{d\eta}\Big|_{\eta=0} \nu_{1,2}^{u+\eta v} &= \frac{d}{d\eta}\Big|_{\eta=0}\left\{\frac{1}{|\Omega|}\int_\Omega S^f_{u+\eta v}(x)(T^g(x)-\mu_2)dx\right\}\\ &= \frac{1}{|\Omega|}\int_\Omega (T^g(x)-\mu_2)\frac{d}{d\eta}\Big|_{\eta=0} S^f_{u+\eta v}(x)dx\\ &= \frac{1}{|\Omega|}\int_\Omega (T^g(x)-\mu_2)\nabla S^f_u(x)\cdot v(x)dx \qquad (A.1)\end{aligned}$$

By noting that

$$\frac{2}{|\Omega|}\int_\Omega (S^f_u(x)-\mu_1^u)\frac{d}{d\eta}\Big|_{\eta=0}\mu_1^{u+\eta v}dx,$$

we have

$$\begin{aligned}\frac{d}{d\eta}\Big|_{\eta=0} \nu_{1}^{u+\eta v} &= \frac{2}{|\Omega|}\int_\Omega (S^f_u(x)-\mu_1^u)\frac{d}{d\eta}\Big|_{\eta=0}(S^f_{u+\eta v}(x)-\mu_1^{u+\eta v})dx\\ &= \frac{2}{|\Omega|}\int_\Omega (S^f_u(x)-\mu_1^u)\frac{d}{d\eta}\Big|_{\eta=0} S^f_{u+\eta v}(x)dx\\ &= \frac{2}{|\Omega|}\int_\Omega (S^f_u(x)-\mu_1^u)\nabla S^f_u(x)\cdot v(x)dx \qquad (A.2)\end{aligned}$$

By using A.1 and A.2, we have

$$\begin{aligned}&\frac{d}{d\eta}\Big|_{\eta=0} CC(S^f_{u+\eta v}(x), T^g(x)) = \frac{d}{d\eta}\Big|_{\eta=0}\nu_{1,2}^{u+\eta v}\cdot(\nu_1^{u+\eta v})^{-1/2}\cdot(\nu_2)^{-1/2}\\ &= \left\{\nu_1^u\cdot\frac{d}{d\eta}\Big|_{\eta=0}\nu_{1,2}^{u+\eta v} - \frac{1}{2}\nu_{1,2}^u\cdot\frac{d}{d\eta}\Big|_{\eta=0}\nu_1^{u+\eta v}\right\}\cdot(\nu_1^u)^{-3/2}\cdot(\nu_2)^{-1/2}\\ &= \frac{1}{|\Omega|}\int_\Omega\left\{\frac{T^g(x)-\mu_2}{(\nu_1^u)^{1/2}\cdot(\nu_2)^{1/2}} - \frac{S^f_u(x)-\mu_1^u}{(\nu_1^u)^{3/2}\cdot(\nu_2)^{1/2}}\cdot\nu_{1,2}^u\right\}\nabla S^f_u(x)\cdot v(x)dx\end{aligned}$$

Hence, the E-L equation for $CC(S^f_{u+\eta v}(x), T^g(x))$ with respect to $u(x)$ is

$$\frac{1}{|\Omega|}\left\{\frac{T^g(x)-\mu_2}{(\nu_1^u)^{1/2}\cdot(\nu_2)^{1/2}} - \frac{S^f_u(x)-\mu_1^u}{(\nu_1^u)^{3/2}\cdot(\nu_2)^{1/2}}\cdot\nu_{1,2}^u\right\}\nabla S^f_u(x)dx = 0.$$

Appendix B. E-L equation of for the local version. By adopting the notations in the paper, we have the following estimates for the means, variances and covariances:

$$S^{f_z}_{u+\eta v}(x) = f_z(S(x+u(x)+\eta v(x))), \qquad T^{g_z}(x) = g_z(T(x)),$$

$$\mu_1^{u+\eta v}(z) = \int_\Omega S^{f_z}_{u+\eta v}(x)G_\theta(x-z)dx, \qquad \mu_1^u(z) = \int_\Omega S^{f_z}_u(x)G_\theta(x-z)dx, \qquad \mu_2(z) = \int_\Omega T^{g_z}(x)G_\theta(x-z)dx.$$

$$\nu_1^{u+\eta v}(z)=\int_\Omega (S_{u+\eta v}^{f_z}(x)-\mu_1^{u+\eta v}(z))^2 G_\theta(x-z)dx,\quad \nu_1^{u}(z)=\int_\Omega (S_{u}^{f_z}(x)-\mu_1^{u}(z))^2 G_\theta(x-z)dx,$$

$$\nu_2(z)=\int_\Omega (T^{g_z}(x)-\mu_2(z))^2 G_\theta(x-z)dx,$$

$$\nu_{1,2}^{u+\eta v}(z)=\int_\Omega (S_{u+\eta v}^{f_z}(x)-\mu_1^{u+\eta v}(z))(T^{g_z}(x)-\mu_2(z)) G_\theta(x-z)dx,$$

$$\nu_{1,2}^{u}(z)=\int_\Omega (S_{u}^{f_z}(x)-\mu_1^{u}(z))(T^{g_z}(x)-\mu_2(z)) G_\theta(x-z)dx.$$

It is easy to see that

$$\left.\frac{d}{d\eta}\right|_{\eta=0} S_{u+\eta v}^{f_z}(x)=\nabla S_u^{f_z}(x)\cdot v(x),\qquad \int_\Omega \mu_1^{u+\eta v}(z)(T^{g_z}(x)-\mu_2(z))G_\theta(x-z)dx=0.$$

Hence,

$$\begin{aligned}&\left.\frac{d}{d\eta}\right|_{\eta=0}\nu_{1,2}^{u+\eta v}(z)=\left.\frac{d}{d\eta}\right|_{\eta=0}\left\{\int_\Omega S_{u+\eta v}^{f_z}(x)(T^{g_z}(x)-\mu_2(z))G_\theta(x-z)dx\right\}\\ &=\int_\Omega (T^{g_z}(x)-\mu_2(z))G_\theta(x-z)\left.\frac{d}{d\eta}\right|_{\eta=0} S_{u+\eta v}^{f_z}(x)dx\\ &=\int_\Omega (T^{g_z}(x)-\mu_2(z))G_\theta(x-z)\nabla S_u^{f}(x)\cdot v(x)dx \qquad \text{(B.1)}\end{aligned}$$

By noting that

$$2\int_\Omega (S_u^{f_z}(x)-\mu_1^u(z))G_\theta(x-z)\left.\frac{d}{d\eta}\right|_{\eta=0}\mu_1^{u+\eta v}(z)dx=0$$

we have

$$\begin{aligned}\left.\frac{d}{d\eta}\right|_{\eta=0}\nu_1^{u+\eta v}(z)&=2\int_\Omega (S_u^{f_z}(x)-\mu_1^u(z))G_\theta(x-z)\left.\frac{d}{d\eta}\right|_{\eta=0}(S_{u+\eta v}^{f_z}(x)-\mu_1^{u+\eta v}(z))dx\\ &=2\int_\Omega (S_u^{f_z}(x)-\mu_1^u(z))G_\theta(x-z)\left.\frac{d}{d\eta}\right|_{\eta=0} S_{u+\eta v}^{f_z}(x)dx\\ &=2\int_\Omega (S_u^{f_z}(x)-\mu_1^u(z))G_\theta(x-z)\nabla S_u^{f_z}(x)\cdot v(x)dx \qquad \text{(B.2)}\end{aligned}$$

By using A.1 and A.2, we have

$$\begin{aligned}&\left.\frac{d}{d\eta}\right|_{\eta=0} CC(S_{u+\eta v}^{f}(x),T^g(x))=\left.\frac{d}{d\eta}\right|_{\eta=0}\int_\Omega \nu_{1,2}^{u+\eta v}(z)\cdot[\nu_1^{u+\eta v}(z)]^{-1/2}\cdot[\nu_2(z)]^{-1/2}dz\\ &=\int_\Omega \left.\frac{d}{d\eta}\right|_{\eta=0}\nu_{1,2}^{u+\eta v}(z)\cdot[\nu_1^{u+\eta v}(z)]^{-1/2}\cdot[\nu_2(z)]^{-1/2}dz\\ &=\int_\Omega\left\{\nu_1^u(z)\cdot\left.\frac{d}{d\eta}\right|_{\eta=0}\nu_{1,2}^{u+\eta v}(z)-\frac{1}{2}\nu_{1,2}^u(z)\cdot\left.\frac{d}{d\eta}\right|_{\eta=0}\nu_1^{u+\eta v}(z)\right\}\cdot[\nu_1^u(z)]^{-3/2}\cdot[\nu_2(z)^{-1/2}dz\\ &=\int_\Omega\int_\Omega G_\theta(x-z)\left\{\frac{T^{g_z}(x)-\mu_2(z)}{[\nu_1^u(z)]^{1/2}\cdot[\nu_2(z)]^{1/2}}-\frac{S_u^{f_z}(x)-\mu_1^u(z)}{[\nu_1^u(z)]^{3/2}\cdot[\nu_2(z)]^{1/2}}\cdot\nu_{1,2}^u(z)\right\}\nabla S_u^{f_z}(x)\cdot v(x)dxdz\end{aligned}$$

Hence, the E-L equation for $CC(S^f_{u+\eta v}(x), T^g(x))$ with respect to $u(x)$ is

$$\int_\Omega G_\theta(x-z)\left\{\frac{T^{g_z}(x)-\mu_2(z)}{[\nu_1^u(z)]^{1/2}\cdot[\nu_2(z)]^{1/2}}-\frac{S_u^{f_z}(x)-\mu_1^u(z)}{[\nu_1^u(z)]^{3/2}\cdot[\nu_2(z)]^{1/2}}\cdot\nu_{1,2}^u(z)\right\}\nabla S_u^{f_z}(x)dz=0.$$

REFERENCES

[1] N Aronszajn, Theory of reproducing kernels, Transactions of the American mathematical society 686, pp.337-404, 1950.

[2] A.C.S.Chung, W.M.W.Wells III, A.Norbash, and W.E.L.Grimson: Multi-modal image registration by minimizing kullback-Leibler distance. International Conference on Medical Image Computing and Computer-Assisted Intervention, vol.2 LNCS pp. 525-532, 2002.

[3] Ho-Ming Chan, Albert C.S. Chung, Simon C.H. Yu, Alexander Norbash, William M. Wells III, "Multi-modal image registration by minimizing Kullback-Leibler distance between expected and observed joint class histograms," cvpr, vol. 2, pp.570, 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03) - Volume 2, 2003

[4] S.Henn and K.Witsch, Multimodal image registration using a variational approach, SIAM J. Sci. Comput., 28(4):1429-1447, 2003.

[5] G.Hermosillo, C.C.Hotel, and O. Faugeras, A variational methods for multimodal image matching, Int. J. Computer Vision, vol. 50(3), pp. 329-343, 2002.

[6] L.R.Jorge, M.S.Juan and V.M.Rafael, Generalized regularization term for non-parametric multimodal image registration, Signal Processing Vol. 87 (11), pp 2837-2842, 2007.

[7] F.Maes, A.Collignon, D.Vandermeulen, G.Marchal, P.Suetens, Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging, Vol.16, pp.187-198, 1997.

[8] J.P.W.Pluim, J.B.A.Maintz and M.A.Viergever, Mutual-information-based registration of medical images: a survey, IEEE Trans. Med. Imaging, vol. 22, pp. 986-1004, 2003.

[9] A. Rényi, On measure of dependence, Acta Mathematica Academiae Scientiarum Hungaria, 10:441-451, 1959.

[10] P.A.Viola and W.M.Wells III, Alignment by maximization of mutual information,Proceedings of International Conference on Computer Vision, 1995, pp. 16-23.

[11] W.M.Wells III, P.Viola, H.Atsumi, S.Nakajima, and R.Kikinis: Multi-modal volume registration by maximizing mutual information, Medical Image Analysis, 1: 35-52, 1996.

[12] L.Zölei, J.Fisher and W.M.Wells III, A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration, Information Processing in Medical Imaging 2003, LNCS 2732, pp. 366-377.

[13] J. Weickert, B. M. H. Romeny and M. A. Viergever: Efficient and Reliable Schemes for Nonlinear Diffusioin Filtering, IEEE Transactions on Image Processing, Vol. 7, No. 3, March 1998

[14] N. Ayache, A. Guimond, A. Roche, and J. Meunier: Three dimensional multimodal brain warping using the demons algorithm and adaptvie intensity correction, IEEE Trans. Med. Imag., vol. 20(1), pp. 58:69, 2001.

[15] R. Bajscy and S. Kovacic, Multiresolution elastic matching, Comput. Vision. Graph. Image Process, vol. 46, pp. 1-12, 1989.

[16] P. Cachier and X. Pennec: 3d non-rigid reigistration by gradient descent on a Gaussian window similarity measure using convolutions, IEEE workshop on mathematical methods in biomedical image analysis, pp. 182-189, 2000.

[17] Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: A variational approach to multimodal image matching. Intl. J. Comp. Vis. 50 (2002) 329-343.

[18] D. Hill, P. Batchelor, M. Holden, and D. Hawkes: Topical review: medical image registration, Physics in Medicine and Biology, vol. 46, pp. 1-45, 2001.

[19] M. Leventon and W.E.L. Grimson: Multi-modal volume registration using joint intensity distributions. In First International Conference on Medical Image Computing and Computer-Assisted Intervention, 1998.

[20] C. Studholme, D.L.G. Hill, and D.J. Hawkes: An overlap invariant entropy measure of 3d medical image alignment. Pattern Recognition, 32(1):71-86, 1999.

[21] Y. Weiss and D. Fleet, Velocity likelihoods in biological and machine vision, Probabilistic Models of the brain, MIT Press, 2002.

[22] Jian, B., Vemuri, B., Marroquin, J.: Robust nonrigid multimodal image registration using local frequency maps. Proc. Inf. Process. Med. Imag. (2005) 504-515
[23] Zhang, Z., Jiang, Y., Tsui, H.: Consistent multi-modal non-rigid registration based on a variational approach. Pattern Recognit. Lett. 27 (2006) 715-725
[24] A.C.S. Chung, W.M.W. Wells III, A. Norbash, and W.E.L. Grimson: Multi-modal image registration by minimizing kullback-leibler distance. In International Conference on Medical Image Computing and Computer-Assisted Intervention, volume 2 of Lecture Notes in Computer Science, pp. 525-532. Springer, 2002.
[25] Viola, P. and Wells, W. 1997. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137-154.
[26] A. Roche, G. Malandain, X. Pennec, and N. Ayache, The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. MICCAI'98,Springer-Verlag Berlin Heidelberg, VOl: 1496, pp. 1115, 1998.
[27] G. E. Christensen and H. J. Johnson, Consistent Image Registration, IEEE Transactions on Medical Imaging. Vol: 20,No.7, pp 721-735,2001
[28] L. Alvarez and R. Deriche and T. Papadopoulo and J. Sanchez, Symmetrical dense optical flow estimation with occlusions detection, European Conference on Computer Vision, pp 721-735, 2002.
[29] N. Akhiezer and I. Glazman, Theory of linear operators in Hilbert space, Dover Publications, 1993
[30] Nachman Aronszajn, Theory of Reproducing Kernels, Transactions of the American Mathematical Society, volume 68, number 3, pages 337404, 1950.
[31] C. Guetter, C. Xu, F. Sauer and J. Hornegger, Learning Based Non-rigid Multi-modal Image Registration Using Kullback-Leibler Divergence, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Volume 3750, 2005
[32] Y. Guo and C. Lu, Multi-modality Image Registration Using Mutual Information Based on Gradient Vector Flow, 18th International Conference on Pattern Recognition (ICPR'06) Volume 3