Research Article

To Align Multimodal Lumbar Spine Images via Bending Energy Constrained Normalized Mutual Information

Shibin Wu 🆔,[1] **Pin He,**[2,3] **Shaode Yu,**[4] **Shoujun Zhou** 🆔,[1] **Jun Xia,**[2] and **Yaoqin Xie** 🆔[1]

[1]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*
[2]*Department of Radiology, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen 518035, China*
[3]*National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, China*
[4]*Department of Radiation Oncology, University of Texas, Southwestern Medical Center, Dallas, TX 75390, USA*

Correspondence should be addressed to Yaoqin Xie; yq.xie@siat.ac.cn

To align multimodal images is important for information fusion, clinical diagnosis, treatment planning, and delivery, while few methods have been dedicated to matching computerized tomography (CT) and magnetic resonance (MR) images of lumbar spine. This study proposes a coarse-to-fine registration framework to address this issue. Firstly, a pair of CT-MR images are rigidly aligned for global positioning. Then, a bending energy term is penalized into the normalized mutual information for the local deformation of soft tissues. In the end, the framework is validated on 40 pairs of CT-MR images from our in-house collection and 15 image pairs from the SpineWeb database. Experimental results show high overlapping ratio (in-house collection, vertebrae 0.97 ± 0.02, blood vessel 0.88 ± 0.07; SpineWeb, vertebrae 0.95 ± 0.03, blood vessel 0.93 ± 0.10) and low target registration error (in-house collection, $\leq 2.00 \pm 0.62$ mm; SpineWeb, $\leq 2.37 \pm 0.76$ mm) are achieved. The proposed framework concerns both the incompressibility of bone structures and the nonrigid deformation of soft tissues. It enables accurate CT-MR registration of lumbar spine images and facilitates image fusion, spine disease diagnosis, and interventional treatment delivery.

1. Introduction

Spine is the backbone of body trunk. It protects the most significant nerve pathway in the spinal cord and the body. On the other hand, spine injury and disorders affect up to 80% world population and may cause deformity and disability, which become a major health and social problem [1–3]. For instance, the lumbar degenerative disease accompanied by pathological changes might result in lumbocrural pain, neural dysfunction, instability of facet joints, and spino-pelvic sagittal imbalance, and thus, the quality of life decreases dramatically. In addition, due to the aging population, the global burden relating to spinal disease remedy is expected to raise significantly in the next decades.

To align intrapatient multimodal images, such as computerized tomography (CT) and magnetic resonance (MR), benefits clinical diagnosis, treatment planning, and delivery for lumbar spinal diseases [4, 5]. However, few methods were dedicated to matching lumbar spine images. Panigrahy et al. developed a method for CT-MR cervical spine images which needed anatomical landmarks to guide image registration [6]. Palkar and Mishra combined different orthogonal wavelet transforms with various transform sizes for CT-MR spine image fusion, while interactive localization of control points was required [7]. Tomazevic et al. implemented an approach for rigid alignment of volumetric CT or MR to X-ray images [8]. To simplify the registration problem in real-world scenarios, images were acquired from a cadaveric lumbar spine phantom and three-dimensional (3D) images contained only

one of the five vertebrae. Otake et al. proposed a registration method for 3D and planar images which was used for spine intervention and vertebral labeling in the presence of anatomical deformation [9]. Harmouche et al. designed an articulated model for MR and X-ray spine images [10]. Hille et al. presented an interactive framework, and rough annotation of the center regions in different modalities was used to guide the registration [11].

Accurate alignment of intrapatient CT-MR images is challenging. From the anatomy, human spine consists of inflexible vertebrae surrounded by soft tissues, such as nerves, vessels, and muscles. Moreover, the vertebrae of lumbar spine are connected by facet joints in the back, which allows for forward and backward extension and twisting movements. Moreover, spinal deformity imposes difficulties on multimodal image registration. Specifically, during image acquisition, patients can lay flatly for a short time due to pain, and subsequently, motion becomes unavoidable. Last but not the least, there are intrinsic differences between CT and MR imaging.

Figure 1 shows a pair of intrapatient CT-MR images. It is found that in CT images, the lumbar spine region easily highlights itself from the rest of soft tissues (the top row), while in MR images, soft tissues show various intensities and in particular, it might be hard to distinguish rigid bones from soft tissues (the bottom row). In the figure, soft tissues in MR images are with various contrast than those in CT images (red arrows), undesirable artifacts caused by the bias field are observed in MR images (green arrows), and these pairs of images show different imaging field of views. It is obvious that these facets pose difficulties in image registration.

2. Related Works

Image registration is important in medical image analysis [12, 13, 14]. Based on similarity metrics, registration methods could be generally classified into intensity- and feature-based methods. Among the intensity-based methods, mutual information (MI) is well known, and it was primarily presented for MR breast image alignment [15]. Afterwards, the metric is used in multimodal medical image registration [16]. For specific applications, MI has been modified to enhance the performance of image registration. For instance, normalized MI (NMI) was proposed for invariant entropy measure [17], regional MI was implemented to capture volume changes when local tissue contrast varied in serial MR images [18], localized MI was designed for atlas matching and prostate segmentation [19], conditional MI was developed to incorporate joint histogram and intensity distribution for image description [20], self-similarity weighted αMI was presented for handheld ultrasound and MR image alignment [21], and MI was also advanced with spatially encoded information [22].

Feature-based methods aim to quantify detected landmarks with features for image registration. Ou et al. collected multiscale multiorientation Gabor features to weight mutual-saliency points for matching [23]. Zhang et al. used scale-invariant features and corner descriptors for lung image registration [24]. Heinrich et al. designed modality indepen-

dent neighborhood descriptor (MIND) which extracted the distinctive structure in small image patches for multimodal deformation registration [25]. Via principal component analysis of deformation, a low-dimension statistical model was learned [26]. Toews et al. combined invariant features of volumetric geometry and appearance for image alignment [27]. Determined by the moments of image patches, a self-similarity inspired local descriptor was presented [28]. Jiang et al. designed a discriminative local derivative pattern which encoded images of different modalities into similar representation [29]. Woo et al. combined spatial and geometric context of detected landmarks [30], and Carvalho et al. considered intensity and geometrical features [31] into a similarity metric. Weistrand and Svensson constrained image registration with anatomical landmarks for local tissue deformation [32].

Embedding a proper penalty term into a similarity metric is helpful in specific applications. Rueckert et al. used a term to regularize the local deformation to be smooth in breast MR image registration [33]. Rohlfing et al. designed a local volume preservation constraint, assuming the soft tissues incompressible in small deformation [34]. Staring et al. proposed a rigidity penalty and modeled the local transform when thorax images with tumors were aligned [35]. To model fetal brain motion, Chen et al. utilized the total-variation regularization and a penalty was adopted toward piece-wise convergence [12]. Due to local tissue rigidity characteristics, Ruan et al. added a regularization term for aligning inhale-exhale CT lung images [36]. Fiorino et al. designed the Jacobian-volumehistogram of deforming organs to evaluate the parotid shrinkage [37].

This study proposes a coarse-to-fine framework to address the registration of intrapatient CT-MR images of lumbar spine. It develops a similarity metric that penalizes a bending energy term into NMI for local deformation of soft tissues. The most similar work is from the comparison of bending energy penalized global and local MI metrics in aligning positron emission tomography and MR images [38], while this study differs itself from the proposed coarse-to-fine registration framework, the bending energy penalized NMI (BEP-NMI) and the application to CT-MR lumbar spine images.

3. Materials and Methods

3.1. Data Collection. Two data sets were analyzed. One is our in-house collection which contains 40 pairs of lumbar spine images from the Department of Radiology, Shenzhen Second People's Hospital, the First Affiliated Hospital of Shenzhen University. CT images were acquired through SIEMENS SOMATO. The voxel resolution is $0.35 \times 0.35 \times 1.00$ mm3, and the matrix size is 512×512 with 180 ± 25 slices. T_2-weighted MR images were acquired using a 1.5 Tesla scanner (SIEMENS Avanto). The physical resolution is $0.7 \times 0.7 \times 3$ mm3, the matrix size is 256×256, and the slice number ranges between 60 and 75.

The other data set is accessible online, namely SpineWeb (http://spineweb.digitalimaginggroup.ca). It includes 15 image pairs of lumbar spine. The physical resolution of CT images is
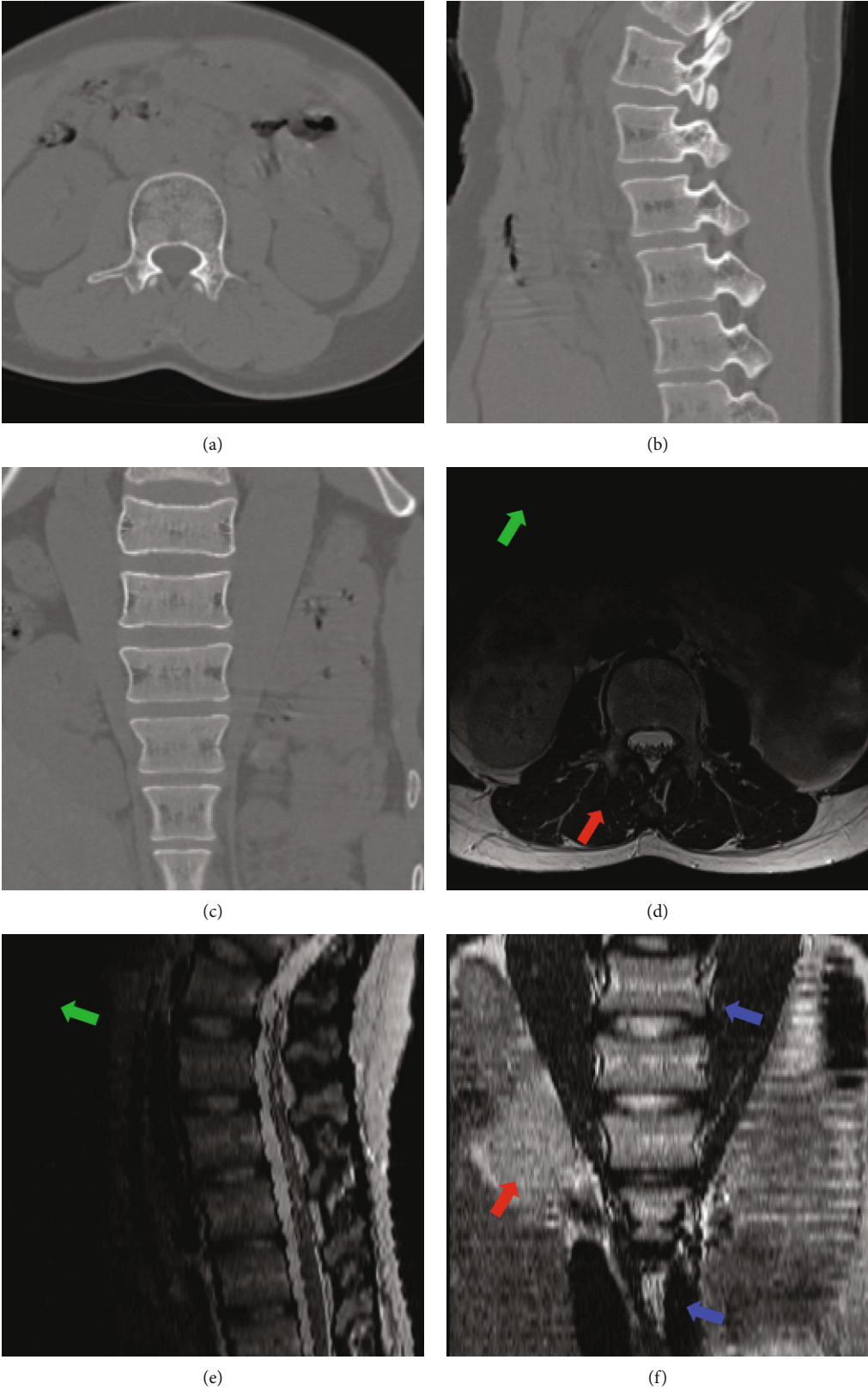
Figure 1: Perceived visual difference between CT and MR images of lumbar spine from three perspective views. The difference of imaging characteristics, fields of view, and unavoidable motion make the registration challenging. Red arrows show different imaging contrast, green arrows direct to the undesirable artifact of bias field in MR images, and blue arrows indicate different field of views. Note that images are cropped and scaled for display purpose.

$0.27 \times 0.27 \times 2.50$ mm3, the image size is 512×512, and the slice number is 77 per volume. The resolution of T_1-weighted MR images is $0.39 \times 0.39 \times 5.00$ mm3, the image size is 512×512, and each volume contains 42 slices.

3.2. The Proposed Framework. The proposed framework consists of two steps both of which use intensity-based image registration methods. An intensity-based registration method can be treated as an optimization problem, and the similarity metric S performs as the cost function. Given a fixed image IF : $\Omega 1 \in R3$ and a moving image $IM : \Omega 2 \in R3$ in 3D space, image registration aims for mapping the moving image IM to the space of the fixed image IF guided by the metric S. When an additional regularization term of P is penalized into S, the registration problem can be formulated as,

$$\widehat{T} = \arg \min_{T} C(T; I_F; I_M) \text{ w.r.t.} C(T\mu; I_F; I_M)$$
$$C(T_\mu; I_F, I_M) = S(T_\mu; I_F, I_M) + \lambda P(T_\mu) \tag{1}$$

where T is a transform model, λ compromises the metric S and the regularity term P, μ is the transform coefficients, and $T\mu$ is the initialized model by μ.

Figure 2 illustrates the proposed framework. It indicates a rigid registration stage and a hierarchical deformation stage, and NMI and BEP-NMI, respectively, perform as the similarity metric. Moreover, adaptive stochastic gradient descent (ASGD) [39] is applied for hyperparameter optimization. Specifically, an affine transformation with 12 degrees of freedom is employed in the first stage, and a B-spline elastic model is used for free-form deformation in the second stage.

3.2.1. Rigid Registration. An affine transform model is used here. The transform $T : \Omega 2 \rightarrow \Omega 1$ can be formulated by

$$T_\mu(x) = R(x - c) + t + c \tag{2}$$

where R is a matrix that contains the rotation, scale, and shear coefficients, c is the center of rotation, t is a translation vector, and μ is a vector of 12 degrees of freedom in volumetric image registration.

Rigid registration attempts for global positioning of the whole body, and thus, an initial alignment of lumbar spine. A 3-level recursive pyramid denotes smoothing that downsamples the source volumes by a factor of 2. Besides, the metric NMI and the affine transform are employed in each scale.

3.2.2. Hierarchical Deformation. Hierarchical deformation is a coarse-to-fine adjustment procedure [40]. This setup utilizes Gaussian pyramid without downsampling to match images from the global structures toward the fine details.

B-spline transform. The B-splines are used to depict the local shape difference between the lumbar vertebrae. To construct the B-spline based free-form deformation model, let $\Omega = \{(x, y, z) \mid 0\,6\,x < X, 0\,6\,y < Y, 0\,6\,z < Z\}$ be a spatial domain of a 3D image. A lattice $(px \times py \times pz)$ of control points is denoted as Ψ, spanning the integer grid in Ω, and Φijk denotes the control point at (i, j, k) on the mesh Ψ. Then, the elastic model can be expressed as a 3D tensor prod-

uct of the uniform B-spline of order 3 as below,

$$T_I(x, y, z) = \sum_{I=0}^{3} \sum_{m-0}^{3} \sum_{n=0}^{3} B_I(u_1) B_m(u_2) B_n(u_3) \Phi_{i+l,j+m,k+n} \tag{3}$$

where $i = \lfloor x/P_x \rfloor - 1, j = \lfloor y/P_Y \rfloor - 1, k = \lfloor z/P_z \rfloor - 1, u_1 = x/P_x - \lfloor x/P_x \rfloor, u_2 = (y/P_y) \lfloor y/P_y \rfloor, u_3 = z/P_Z - \lfloor z/P_z \rfloor$, and Bl repents the lth basis function of the B-spline,

$$\begin{cases} B_0(u) = (1-u)^3/6, \\ B_1(u) = (3u^3 - 6u^2 + 4)/6 \\ B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) = u^3/6 \end{cases} \tag{4}$$

where $0\,6\,u < 1$. The basic functions weigh the contribution of each control point to $Tl(x, y, z)$ based on its distance to the point (x, y, z).

Since the B-splines can be locally controlled, it makes the computation efficient for a large number of control points. In particular, changing a control point affects only the transforms of its local neighborhood.

BEP-NMI. The metric MI is preferred in multimodal image registration. Given IF and IM with intensity bins of f and m, MI is quantified from a joint probability function $p(IF, IM)$ and marginal probability distribution functions. of $p(f) = Pf\{p(f, m)\}$ and $p(m) = Pm\{p(f, m)\}$. The metric MI between a pair of images, IF and IM, can be described as

$$\begin{aligned} MI(I_F; I_M) &= H(I_F) + H(I_M)..H(I_F; I_M) \\ &= \sum_{f \in F} \sum_{m \in M} p(f, m) \log \left\langle \frac{p(f, m)}{p(f)p(m)} \right\rangle \end{aligned} \tag{5}$$

where $H(IF)$ and $H(IM)$ are the marginal entropy and the $H(IF, IM)$ is the joint entropy of IF and IM.

The metric NMI is more robust to the change of overlapped tissue regions. It uses a Parzen-window approach to estimate the probability density function. The entropy of a fixed image IF is defined as $H(IF) = -Pf \in F\,p(f)logp(f)$, where $p(f)$ is a probability distribution estimated by using Parzen-windows. The entropy of a moving image IM can be computed in a similar way. And subsequently, the NMI between IF and IM can be presented as

$$NMI(I_F; I_M) = \frac{H(I_F) + H(I_M)}{H(I_F; I_M)} \tag{6}$$

In order to regularize the B-spline deformation and to prevent the rigid structures from being smoothed, a BEP term $P(u)$ is added to the NMI. The new cost function, BEP-NMI, is formulated as
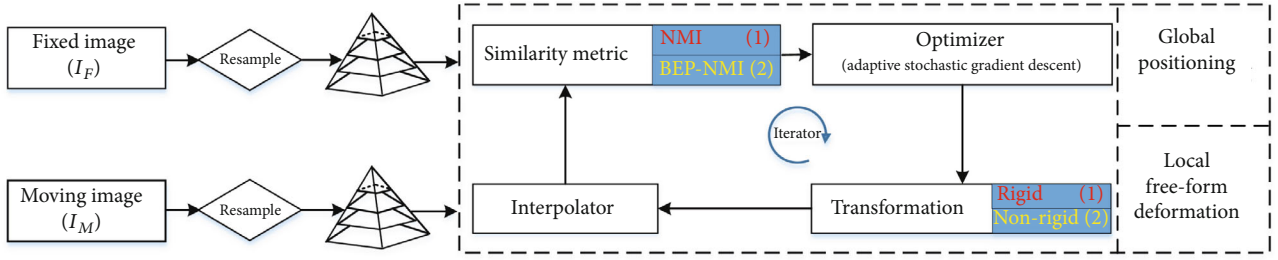
$$C(\mu) = y_1 S(\mu) + y_2 P(\mu) \tag{7}$$

FIGURE 2: The proposed coarse-to-fine framework for aligning CT-MR lumbar spine images. It consists of two stages. The first stage is for global positioning via NMI based rigid registration (highlighted in red), and the second stage is for the local deformation of soft tissues via the bending energy penalized NMI (highlighted in yellow). Both stages utilize the same workflow for iterative optimization.

where $\gamma1$ and $\gamma2$ are predefined constants to weigh between global similarity and local regularity. In this study, off-line experiments indicated that $\gamma1 = \gamma2 = 1$ was a good choice.

The penalty terms are commonly based on the first or second-order spatial derivatives of the transform [35, 36]. In this study, the BEP term is composed of the second-order derivatives [35, 40] in the volumetric space,

$$
P_{BEP}(\mu) = \int_V \left\{ \left(\frac{\partial^2 T}{\partial x^2} \right)^2 + \left(\frac{\partial^2 T}{\partial y^2} \right)^2 + \left(\frac{\partial^2 T}{\partial z^2} \right)^2 \right.
$$
$$
\left. + 2 \left(\frac{\partial^2 T}{\partial x \partial y} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial y \partial z} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial z \partial x} \right)^2 \right\} dxdydz
$$
$$
\text{(8)}
$$

where V is a 3D image. The Equation (8) can be approximated as a discretely sampled sum over the volume V as below,

$$
P_{BEP} = \frac{1}{N_V} \sum_{x \in V} \Phi T(x, y, z) \qquad \text{(9)}
$$

where N is the number of voxels in V, and Φ denotes a sum of the squared second-order derivatives of T inside the integral part in Equation (8) at a voxel location (x, y, z). Specially, the derivative approximation with finite differences can be restricted to the local neighborhood of the control point.

Optimization. Given an initial parameter μ, an optimization algorithm updates an incremental $\Delta\mu$ to reduce the cost function C iteratively. ASGD is used in the study, since it runs faster and less likely to get trapped in the local minima when compared to other gradient-based optimization algorithms [39]. Notably, ASGD implemented in the elastix package (http://elastix.isi.uu.nl) is used for adaptive step size prediction and the initial parameters are set as those in [39, 40].

3.3. A Comparison Method. The MIND is a feature-based method and it has been widely used in multimodal deformable registration [25, 41]. It aims to represent the distinctive image structure in a local neighborhood and explore the similarity of small image patches by using Gaussian-weighted patch distances [25].

MIND can be formulated by a distance Dp, a spatial search region R and a variance estimate V as below,

$$
\text{MIND}(I, x, r) = \frac{1}{n} \exp \left(\frac{D_p(I, x, x+r)}{V(I, x)} \right) r \in R \qquad \text{(10)}
$$

$$
Dp(I, x.x+r) = C * (I - I(r))^2 \qquad \text{(11)}
$$

where n is a normalization constant, r the search region, C a convolution filter of size $(2p+1)d$, $*$ a convolution filter, and $I0(r)$ a dense sampling on r. As such, an image can be represented by a vector of size $|R|$ at each location x. Moreover, $V(I, x)$ can be computed based on a mean of the patch distances within a small neighborhood n ($n \in N$)

$$
V(I, x) = \frac{1}{N} \sum_{n \in N} D_p(I, x, x+n), \qquad \text{(12)}
$$

In Equation (10) to Equation (12), $n = 6$ denotes a six-connected neighborhood and $p = 1$ indicates a $3 \times 3 \times 3$ volume block.

The similarity metric used in MIND comes from the sum of absolute difference. To the fixed image (IF) and the moving image (IM), the local difference at a voxel x is

$$
LD(x) = \frac{1}{|R|} \sum_{r \in R} \left| MIND(I_{F,x,r}) - MIND(I_{M,x,r}) \right| \qquad \text{(13)}
$$

The default value of $|R|$ is 6 and it means 6-connected neighbors are taken into computation.

3.4. Performance Evaluation

3.4.1. Tissue Overlapping. Tissue overlapping quantifies the overlapping ratio of outlined tissue regions in the fixed and its aligned image, which can distinguish the reasonable from the poor registration [42, 43]. This study focuses on the region of lumbar vertebrae and blood vessels. Assuming the outlined tissues in the fixed and aligned image are, respectively, denoted as OF and OA, the voxel-wise Jaccard (J) index and
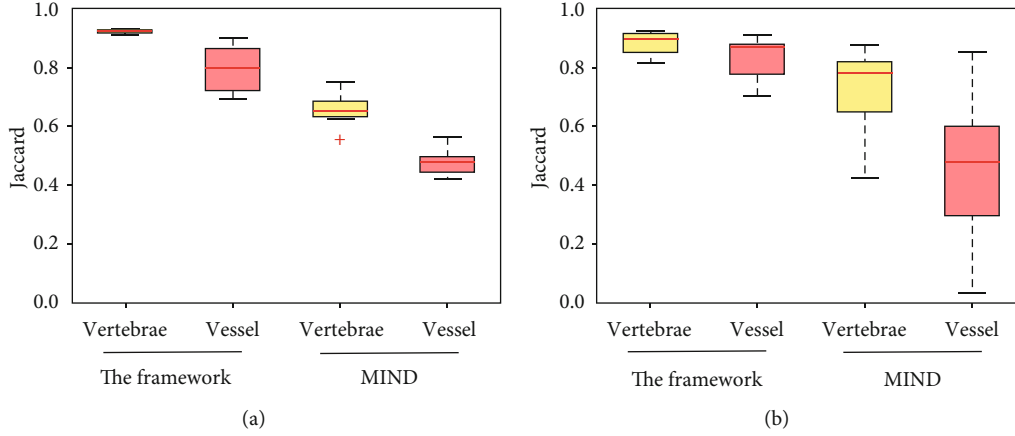
(a)

(b)

FIGURE 3: Jaccard index of the vertebrae and blood vessel overlapping on the in-house dataset (a) and the online dataset (b). Box-and-whisker plots represent the median Jaccard index (horizontal line) and total range (whiskers). The red $^+$ indicates an outlier that causes failure in image registration.
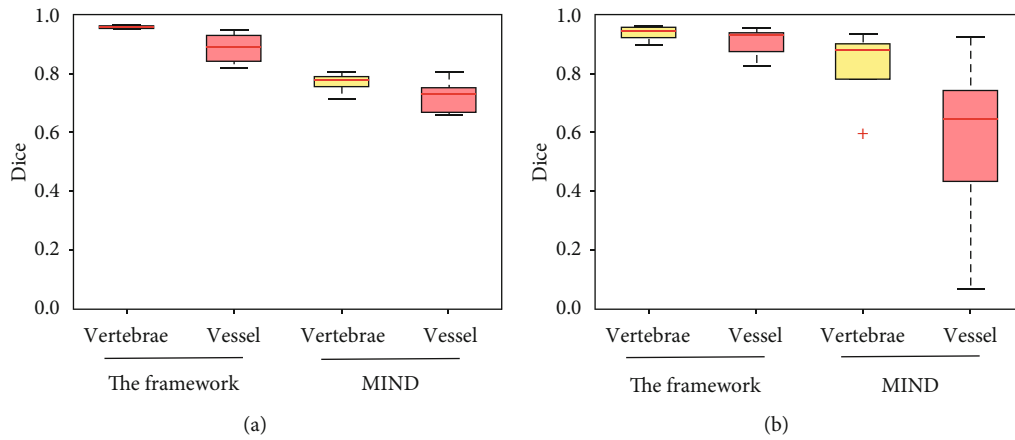

(a)

(b)

FIGURE 4: Tissue overlapping metric of Dice coefficient of the vertebrae and blood vessel on the in-house dataset (a) and the online dataset (b). Box-and-whisker plots show the median coefficient (horizontal line) and total range (whiskers). The red $^+$ indicates a failure case.
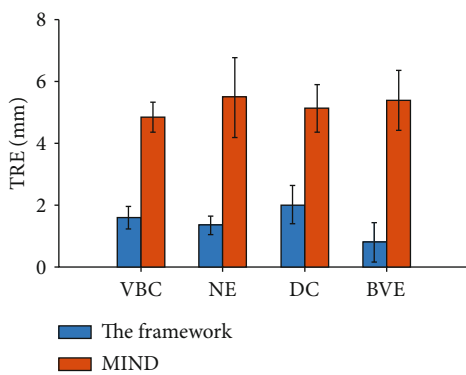

FIGURE 5: TRE values of anatomical landmarks on the in-house collection.

Dice (D) coefficient can be, respectively, described as

$$
\begin{aligned}
J &= \left| \frac{O_F \cap O_A}{O_F \cup O_A} \right|, \\
D &= 2\, \frac{|O_F \cap O_A|}{|O_F| + |O_A|}
\end{aligned}
\tag{14}
$$

where $|\cdot|$ indicates the number of voxels per volume.

3.4.2. Target Registration Errors. As for landmark annotation, ImageJ (http://imagej.nih.gov/ij/) was used. A pair of CT-MR images are displayed side-by-side. Then, landmarks are identified and manually annotated by an imaging radiologist (3+ year experience) and further confirmed by a senior radiologist (10+ year experience). Once landmarks are annotated, their locations in 3D space are recorded. In this study, anatomical landmark points are localized on the vertebral body center (VBC), neural edge (NE), disc center (DC), and blood vessel edge (BVE).

Target registration error (TRE) evaluates the distance between anatomical point pairs in the fixed and moving

Table 1: TRE values (mean ± SD) on the in-house collection images.

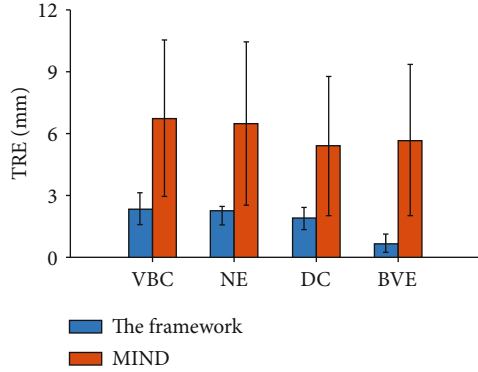|	The framework (mm)	MIND (mm)
VBC	1.52 ± 0.33	5.02 ± 3.76
NE	1.38 ± 0.29	5.07 ± 4.06
DC	2.01 ± 0.62	5.11 ± 3.69
BVE	0.78 ± 0.64	3.77 ± 4.21

Figure 6: TRE values of anatomical landmarks on the SpineWeb dataset.

Table 2: TRE values (mean ± SD) on the SpineWeb images.

	The framework (mm)	MIND (mm)
VBC	2.37 ± 0.76	6.75 ± 3.80
NE	1.91 ± 0.55	5.41 ± 3.38
DC	2.26 ± 0.98	6.49 ± 3.95
BVE	0.66 ± 0.46	5.71 ± 3.65

image. Here, assuming l_i and , respectively, denotes the corresponding landmark point pairs in the fixed and moving image, the mean TRE for a given T is defined as

$$TRE = \frac{1}{n} \sum_{i}^{n} \left\| l_i - T\left(l_i'\right) \right\| \qquad (15)$$

where n is the number of pairs of landmark, and $k \cdot k$ indicates Euclidian distance in 3D space.

3.5. Software and Platform. The whole framework is implemented with Insight Segmentation and Registration Toolkit (http://www.itk.org) and the elastix package [40]. Experiments are performed on a desktop computer equipped with dual-core Intel i7 CPU (3.70 GHz) and 16 GB RAM memory.

4. Results

4.1. Tissue Overlapping. Figure 3 illustrates the tissue overlapping measure J of CT-MR image registration on the in-house collection (left) and the SpineWeb (right). The left shows that the proposed framework outperforms the MIND method on the vertebrae (0.93 ± 0.02 versus 0.69 ± 0.06) and blood vessel (0.81 ± 0.10 versus 0.48 ± 0.07) overlapping. In the right figure, the framework achieves higher values (vertebrae, 0.89 ± 0.05; blood vessel, 0.81 ± 0.12) than the MIND method (vertebrae, 0.75 ± 0.12; blood vessel, 0.52 ± 0.33), and thus, it leads to better performance.

Figure 4 shows the overlapping ratio D of multimodal image registration on the in-house collection (left) and the SpineWeb (right). The left figure indicates that the coarse-to-fine registration framework obtains better results than the MIND method on the vertebrae (0.97 ± 0.02 versus 0.77±0.05) and blood vessel (0.88 ± 0.07 versus 0.74 ± 0.07) overlapping. In the right figure, the MIND method (vertebrae, 0.86 ± 0.12; blood vessel, 0.61 ± 0.33) obtains inferior performance than the proposed framework (vertebrae, 0.95 ± 0.03; blood vessel, 0.93 ± 0.10).

4.2. Target Registration Errors. Figure 5 demonstrates the mean TRE value of anatomical landmark points between the proposed framework and the MIND algorithm on the in-house collection. The error-bar plot indicates that the TRE of the proposed framework is less than 3.00 mm (DC), while that of the MIND algorithm is larger than 4.00 mm (VBC) on average. In addition, statistical analysis indicates that the proposed framework significantly outperforms the MIND algorithm in each of the four sets of landmarks ($p < 0.005$, two-sample t-test).

Table 1 shows the TRE values (mean ± standard deviation, mean ± SD) with respect to different landmark sets. The coarse-to-fine framework achieves TRE between 0.78 ± 0.64 mm (BVE) and 2.01 ± 0.62 mm (DC), while the TRE of the MIND method ranges from 3.77 ± 4.21 mm (BVE) to 5.11 ± 3.69 mm (DC), correspondingly larger than that from the proposed framework.

The mean TRE on the SpineWeb dataset is shown in Figure 6. It is observed that the TRE value of the proposed framework is less than 3.00 mm (VBC and NE), while the MIND algorithm leads to the TRE values larger than 5.00 mm.

Statistical analysis indicates significant difference of the TRE values between the proposed framework and the MIND algorithm on aligning the pairs of VBC and BVE landmarks ($0.01 < p < 0.05$, two-sample t-test).

Table 2 summarizes the mean TRE values on different sets of landmark pairs. The proposed framework achieves the TRE values between 0.66 ± 0.46 mm (BVE) to 2.37 ± 0.76 mm (VBC), and the TRE values of the MIND algorithm ranges from 5.71 ± 3.65 mm (BVE) to 6.75 ± 3.80 mm (VBC).

4.3. Perceived Quality of Image Alignment. Visual assessment of registration quality is perceived from the fusion of CT and MR images and observed from three perspective views in Figure 7, where (A, E, I) are the CT image, (B, F, J) are the MR image, (C, G, K) are the aligned image from the proposed framework, and (D, H, L) are the aligned image from the MIND algorithm. Red arrows directing to the soft tissue regions and green arrows directing to the bone regions are used for comparison. Before registration, both bones and tissues are
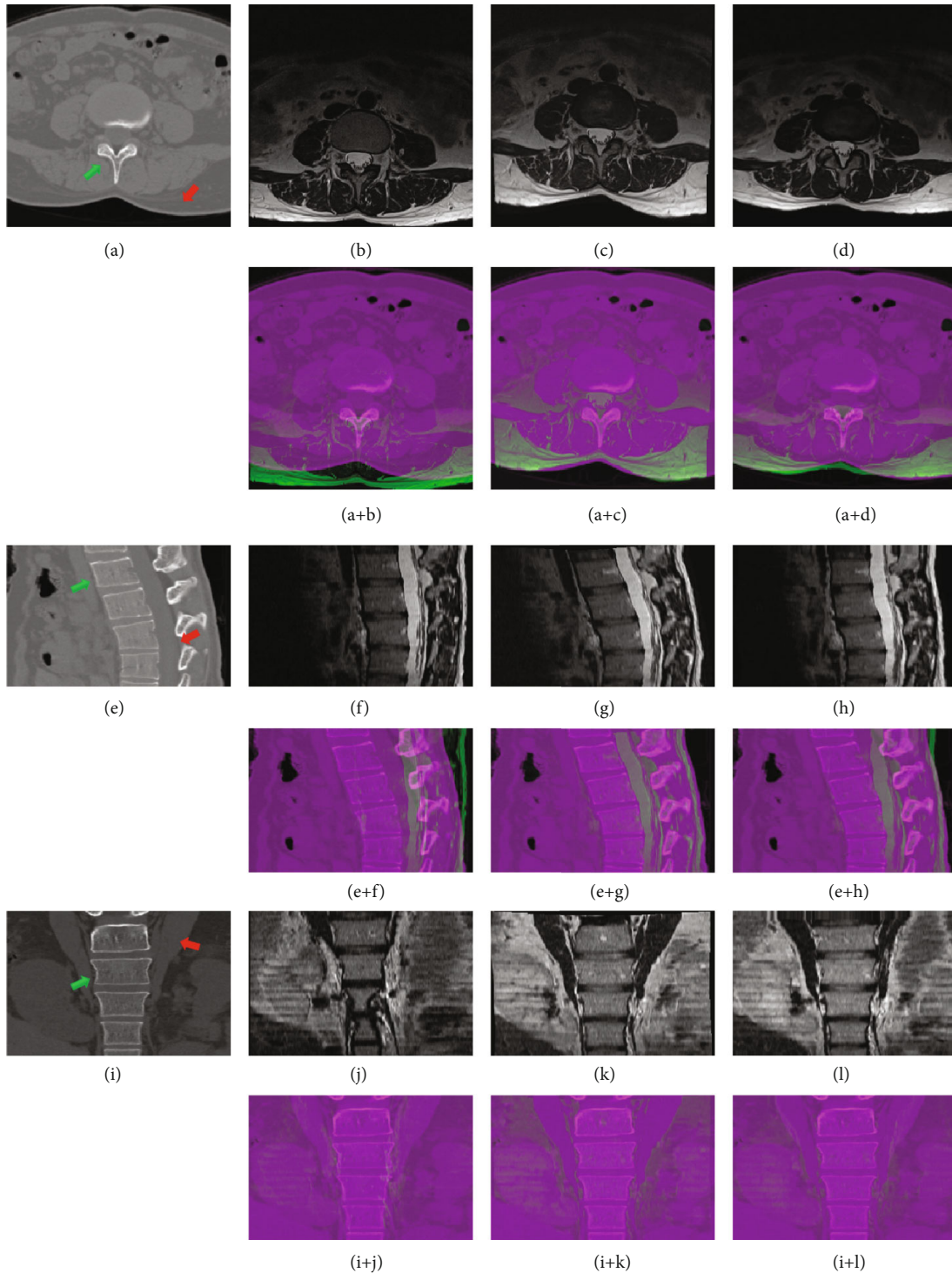
Figure 7: Perceived visual difference of CT-MR images before and after image registration. The regions directed by the arrows are for comparison before and after registration. In addition, images are cropped for display purpose.

misaligned, such as acantha $(A + B)$, bones and nerves $(E + F)$ and muscles $(I + J)$. After image registration, the proposed framework aligns these parts in the MR images with fine deformation to the CT images. Specifically, both rigid bones and soft tissues are well matched, and the anatomical textures shows consistent distributions in the aligned image. On contrary,

the MIND algorithm fails to overlap the acantha $(A + D)$, bones and nerves $(E + H)$ and muscles $(I + L)$ accurately.

4.4. Computation Time. Based on the software and platform, it took about 62 seconds to complete the affine registration and 427 seconds to complete the deformable registration. And

thus, it required a total of 8.15 minutes to fulfill the coarse-to-fine registration for a pair of CT-MR lumbar spine image.

5. Discussion

Intrapatient multimodal image registration can fuse multi-source information that benefits disease diagnosis and treatment delivery. This study develops a coarse-to-fine framework and aligns intrapatient CT-MR lumbar spine images. It first utilizes the similarity metric NMI for global positioning, and then, bending energy penalized NMI for local deformation of soft tissues. The proposed framework achieves high tissue overlapping ratio and low target registration error. It not only preserves the incompressibility of vertebrae but also well matches local soft tissues that provide accurate elastic registration of lumbar spine images for clinical applications.

The proposed framework is a coarse-to-fine approach for multimodal image registration. It aligns anatomical structures and addresses the potential difference on the fields of view and the intrinsic differences between medical imaging. The metric NMI is used, since it is a robust and accurate measure in multimodal image registration [17, 44]. After global positioning, a new similarity metric that integrates a bending energy term into NMI is used for local deformation and registration of soft tissues in medical images. It is worth of note that the term encourages smooth displacements in registration [33]. Ceranka et al. embedded the term to improve multiatlas segmentation of the skeleton from whole-body MR images [45], and de Vos et al. integrated the term into unsupervised affine and deformable image registration by using a convolutional neural network [46]. Both works [45, 46] figured out that the term caused significantly less folding in image registration.

The framework takes the incompressibility of vertebrae into account. Vertebrae are bony structures which are connected to each other by the ligamentum flavum at the neural arch [47]. The proposed framework enables global and local image structures well matched, and inflexible bones and soft tissues properly deformed. Its superior performance has been verified on the in-house collection and the SpineWeb database. Experiential results demonstrate that the overlapping ratio of annotated vertebrae and blood vessels are larger than 0.85, and the target registration error is less than 2.40 mm on average. It outperforms the MIND algorithm partly due to its proper deformation of local soft tissues and incompressible lumbar vertebrae. The registration quality is further perceived in a CT-MR image pair. It is found that the marked tissues keep relative location after image registration by using the proposed framework, since it not only well tackles the local soft tissue deformation but also conserves the rigid lumbar vertebrae.

Even if the proposed framework achieves superior performance on aligning CT-MR lumbar spine images, there is still room for further improvement. One way to enhance registration accuracy is by transferring multimodal image registration into mono-modal image registration. Wachinger and Navab developed structural representations, such as Entropy and Laplacian images, which could represent the images in a third space where the images showed close intensity or gradient distribution [48]. Moreover, deep networks have been explored to estimate CT images from MR images directly and in particular, the mapping between CT and MR images was learned without any patch-level pre- or postprocessing [49]. Another straightforward way is to utilize deep networks to learn the deformation field between different imaging modalities [50]. In addition, interactive image registration is admirable in interventional surgery and a doctor user could localize landmarks to guide and to update the registration procedure [51].

There are several limitations in this study. One limitation comes from no comparison of NMI and BEP-NMI on deformable image deformation, since our off-line experimental results show that the NMI based deformable registration is prone to distortion of lumbar spine and unnatural deformation of soft tissues. Moreover, demons and its variants [52, 53, 54] failed in the registration of lumbar spine images. Thus, this study reports the performance of the proposed framework and the MIND method. In addition, how to properly balance the BEP term and the NMI is always a problem and no existing methods could well tackle this issue, while prior knowledge [35, 37] could be employed for further improvement of the registration accuracy.

6. Conclusions

This paper presents a coarse-to-fine framework for the registration of intrapatient CT-MR lumbar spine images. It integrates the bending energy term into normalized mutual information for fine deformation of soft tissues around the incompressible vertebrae. Its high performance benefits multisource information fusion for accurate spine disease diagnosis, treatment planning, interventional surgery, and radiotherapy delivery.

Data Availability

The in-house collection of MR-CT image pairs used to support the findings of this study are restricted by the Medical Ethics Committee of Shenzhen Second People's Hospital in order to protect patient privacy. The SpineWebdata set of MR-CT images used to support the findings is freely available online (https://spineweb.digitalimaginggroup.ca/spineweb/index.php?n=Main.Datasets). If interested, requests for access to these data can be made to the author Shibin Wu (https://sb.wu@siat.ac.cn). Since the database is freely available, requests for access to these data can also be made to the author Shibin Wu (https://sb.wu@siat.ac.cn).

Conflicts of Interest

The authors declare there is no conflict of interest. The founding sponsors had no role in the design of this study, in the collection, analysis or interpretation of data, in the writing of this manuscript, nor in the decision to publish the experimental results.

Acknowledgments

References

[1] G. Zheng and S. Li, "Medical image computing in diagnosis and intervention of spinal diseases," *Computerized Medical Imaging and Graphics*, vol. 45, pp. 99–101, 2015.

[2] F. Raciborski, R. Gasik, and A. Klak, "Disorders of the spine. A major health and social problem," *Annals of Physics*, vol. 4, no. 4, pp. 196–200, 2016.

[3] G. Zhang, Y. Yang, Y. Hai, J. Li, X. Xie, and S. Feng, "Analysis of Lumbar Sagittal Curvature in Spinal Decompression and Fusion for Lumbar Spinal Stenosis Patients under Roussouly Classification," *BioMed Research International*, vol. 2020, 8 pages, 2020.

[4] A. Toussaint, A. Richter, F. Mantel et al., "Variability in spine radiosurgery treatment planning - results of an international multi-institutional study," *Radiation Oncology*, vol. 11, no. 1, p. 57, 2016.

[5] P. A. Helm, R. Teichman, S. L. Hartmann, and D. Simon, "Spinal navigation and imaging: history, trends, and future," *IEEE Transactions on Medical Imaging*, vol. 34, no. 8, pp. 1738–1746, 2015.

[6] A. Panigrahy, S. Caruthers, J. Krejza et al., "Registration of three-dimensional MR and CT studies of the cervical spine," *American Journal of Neuroradiology*, vol. 21, no. 2, pp. 282–289, 2000.

[7] B. Palkar and D. Mishra, "Fusion of multi-modal lumbar spine images using Kekre's hybrid wavelet transform," *IET Image Processing*, vol. 13, no. 12, pp. 2271–2280, 2019.

[8] D. Tomazevic and F. Pernus, "Robust gradient-based 3-D/2-D registration of C-T and MR to X-ray images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 12, pp. 1704–1714, 2008.

[9] Y. Otake, A. S. Wang, J. W. Stayman et al., "Robust 3D-2D image registration: application to spine interventions and vertebral labeling in the presence of anatomical deformation," *Physics in Medicine and Biology*, vol. 58, no. 23, pp. 8535–8553, 2013.

[10] R. Harmouche, F. Cheriet, H. Labelle, and J. Dansereau, "3D registration of MR and X-ray spine images using an articulated model," *Computerized Medical Imaging and Graphics*, vol. 36, no. 5, pp. 410–418, 2012.

[11] G. Hille, S. Saalfeld, S. Serowy, and K. Tonnies, "Multi-segmental spine image registration supporting image-guided interventions of spinal metastases," *Computers in Biology and Medicine*, vol. 102, pp. 16–20, 2018.

[12] L. Chen, H. Zhang, S. Wu, S. Yu, and Y. Xie, "Estimating fetal brain motion with total-variation-based magnetic resonance image registration," in *World Congress on Intelligent Control and Automation (WCICA)*, pp. 809–813, Guilin, China, June 2016.

[13] A. Khalil, S. Ng, Y. M. Liew, and K. W. Lai, *An Overview on Image Registration Techniques for Cardiac Diagnosis and Treatment, Cardiology Research and Practice*, Hindawi, 2018.

[14] J. Li and Q. Ma, *A Fast Subpixel Registration Algorithm Based on Single-Step DFT Combined with Phase Correlation Constraint in Multimodality Brain Image, Computational and Mathematical Methods in Medicine*, Hindawi, 2020.

[15] P. Viola and W. Wells-III, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[16] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[17] 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.

[18] 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 Transactions on Medical Imaging*, vol. 25, no. 5, pp. 626–639, 2006.

[19] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Medical Physics*, vol. 35, no. 4, pp. 1407–1417, 2008.

[20] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 19–29, 2010.

[21] H. Rivaz, Z. Karimaghaloo, and D. L. Collins, "Self-similarity weighted mutual information: a new nonrigid image registration metric," *Medical Image Analysis*, vol. 18, no. 2, pp. 343–358, 2014.

[22] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, "A non-rigid registration framework using spatially encoded mutual information and free-form deformations," *IEEE Transactions on Medical Imaging*, vol. 30, no. 10, pp. 1819–1828, 2011.

[23] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "Dramms: deformable registration via attribute matching and mutual-saliency weighting," *Medical Image Analysis*, vol. 15, no. 4, pp. 622–639, 2011.

[24] R. Zhang, W. Zhou, Y. Li, S. Yu, and Y. Xie, "Nonrigid registration of lung CT images based on tissue features," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 834192, 2013.

[25] M. P. Heinrich, M. Jenkinson, M. Bhushan et al., "MIND: modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical Image Analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.

[26] J. A. Onofrey, L. H. Staib, and X. Papademetris, "Learning nonrigid deformations for constrained multi-modal image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, pp. 171–178, 2013.

[27] M. Toews, L. Zöllei, and W. M. Wells, "Feature-based alignment of volumetric multimodal images," in *Information*

Processing in Medical Imaging, Lecture Notes in Computer Science, pp. 25–36, 2013.

[28] F. Zhu, M. Ding, and X. Zhang, "Self-similarity inspired local descriptor for non-rigid multi-modal image registration," *Information Sciences,* vol. 372, pp. 16–31, 2016.

[29] D. Jiang, Y. Shi, X. Chen, M. Wang, and Z. Song, "Fast and robust multimodal image registration using a local derivative pattern," *Medical Physics,* vol. 44, no. 2, pp. 497–509, 2017.

[30] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Transactions on Image Processing,* vol. 24, no. 2, pp. 757–769, 2015.

[31] D. D. B. Carvalho, S. Klein, Z. Akkus et al., "Joint intensity-andpoint based registration of free-hand B-mode ultrasound and MRI of the carotid artery," *Medical Physics,* vol. 41, no. 5, pp. 052904–052912, 2014.

[32] O. Weistrand and S. Svensson, "The ANACONDA algorithm for deformable image registration in radiotherapy," *Medical Physics,* vol. 42, no. 1, pp. 40–53, 2015.

[33] 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, 1999.

[34] T. Rohlfing, C. Maurer, D. Bluemke, and M. Jacobs, "Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint," *IEEE Transactions on Medical Imaging,* vol. 22, no. 6, pp. 730–741, 2003.

[35] M. Staring, S. Klein, and J. Pluim, "A rigidity penalty term for nonrigid registration," *Medical Physics,* vol. 34, no. 11, pp. 4098–4108, 2007.

[36] D. Ruan, J. Fessler, M. Roberson, J. Balter, and M. Kessler, "Nonrigid registration using regularization that accommodates local tissue rigidity," *Computers in Biology and Medicine,* vol. 42, no. 1, pp. 123–128, 2012.

[37] C. Fiorino, E. Maggiulli, S. Broggi et al., "Introducing the jacobian-volume-histogram of deforming organs: application to parotid shrinkage evaluation," *Physics in Medicine and Biology,* vol. 56, no. 11, pp. 3301–3312, 2011.

[38] S. Leibfarth, D. Monnich, S. Welz et al., "A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning," *Acta Oncologica,* vol. 52, no. 7, pp. 1353–1359, 2013.

[39] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent Optimisation for image registration," *International Journal of Computer Vision,* vol. 81, no. 3, pp. 227–239, 2009.

[40] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging,* vol. 29, no. 1, pp. 196–205, 2010.

[41] S. Reaungamornrat, T. D. Silva, A. Uneri et al., "MIND demons: symmetric diffeomorphic deformable registration of MR and CT for image-guided spine surgery," *IEEE Transactions on Medical Imaging,* vol. 35, no. 11, pp. 2413–2424, 2016.

[42] S. Yu, S. Wu, L. Zhuang et al., "Efficient segmentation of a breast in B-mode ultrasound tomography using threedimensional GrabCut (GC3D)," *Sensors,* vol. 17, no. 8, p. 1827, 2017.

[43] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable," *IEEE Transactions on Medical Imaging,* vol. 31, no. 2, pp. 153–163, 2012.

[44] T. Veninga, H. Huisman, R. van der Maazen, and H. Huizenga, "Clinical validation of the normalized mutual information method for registration of CT and MR images in radiotherapy of brain tumors," *Journal of Applied Clinical Medical Physics,* vol. 5, no. 3, pp. 66–79, 2004.

[45] J. Ceranka, S. Verga, M. Kvasnytsia et al., "Multi-atlas segmentation of the skeleton from whole-body MRI - impact of iterative background masking," *Magnetic Resonance in Medicine,* vol. 83, no. 5, pp. 1851–1862, 2019.

[46] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis,* vol. 52, pp. 128–143, 2019.

[47] R. Wang and M. M. Ward, "Arthritis of the spine," in *Spinal Imaging and Image Analysis,* pp. 31–66, Springer, Cham, 2015.

[48] C. Wachinger and N. Navab, "Entropy and laplacian images: structural representations for multi-modal registration," *Medical Image Analysis,* vol. 16, no. 1, pp. 1–17, 2012.

[49] W. Li, Y. Li, W. Qin et al., "Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy," *Quantitative Imaging in Medicine and Surgery,* vol. 10, no. 6, pp. 1223–1236, 2020.

[50] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications,* vol. 31, no. 1-2, p. 8, 2020.

[51] A. Herline, J. Herring, J. Stefansic, W. Chapman, R. Galloway, and B. Dawant, "Surface registration for use in interactive, image-guided liver surgery," *Computer Aided Surgery,* vol. 5, no. 1, pp. 11–17, 2000.

[52] J. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis,* vol. 2, no. 3, pp. 243–260, 1998.

[53] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage,* vol. 45, no. 1, pp. S61–S72, 2009.

[54] H. Lombaert, L. Grady, X. Pennec, and N. Ayache, "F. Cheriet-Spectral logdemons: Diffeomorphic image registration with very large deformations," *International Journal of Computer Vision,* vol. 107, no. 3, pp. 254–271, 2014.