IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 23, 2017, accepted October 10, 2017, date of publication October 13, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2762901

Joint-Saliency Structure Adaptive Kernel Regression with Adaptive-Scale Kernels for Deformable Registration of Challenging Images

BINJIE QIN[1], (Member, IEEE), ZHUANGMING SHEN[1], ZESHAN FU[1], ZIEN ZHOU[2], YISONG LV[3], AND JINSONG BAO[4]
[1]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China
[3]School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
[4]College of Mechanical Engineering, Donghua University, Shanghai 200051, China

Corresponding authors: Binjie Qin (bjqin@sjtu.edu.cn) and Jinsong Bao (bao@dhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61271320 and Grant 51475301 and in part by the Medical Engineering Cross Fund of Shanghai Jiao Tong University under Grant YG2014MS29.

ABSTRACT This paper proposes a locally adaptive kernel regression with adaptive-scale kernels for deformable image registration with outliers (i.e., missing correspondences and large local deformations). The adaptive kernel regression locally constructs dense deformation elds from the weighted contributions of each pixel's surrounding discrete displacement elds in a moving anisotropic kernel by exploiting the contextual deformations of the corresponding saliency structures in the two images. Speci cally, we rst propose an effective superpixel-based structure scale estimator to estimate the boundary-aware structure scale of each reference structure. We further propose an edge-aware mismatch scale measuring the mismatch degree of the edge structures to be matched in the images. By combining the boundary-aware structure scale with the edge-aware mismatch scale of the underlying saliency structures to be matched, we de ne edge-aware adaptive-scale kernels for the locally adaptive kernel regression to ef ciently construct deformations for deformable registration with outliers. The experiments show that the proposed method achieves not only state-of-the-art matching accuracy for normal corresponding structures but also the best matching ef ciency for outlier structures in deformable image registration.

INDEX TERMS Deformable registration, structure scale, mismatch scale, joint saliency map, outliers, spatially adaptive nonparametric regression, joint-saliency structure.

I. INTRODUCTION

Deformable image registration [1], or optical ow [2] computation (referred to as monomodal deformable image registration), is the task of spatially aligning the points of every corresponding local structure by minimizing the feature-based and/or intensity-based differences between two images. Accurately matching corresponding local structures in two images by deformable image registration has numerous applications in computer vision, image processing and pattern recognition [1] [4]. However, because of the image content changes over a period of time and the different imaging mechanisms of multimodal sensors, some local structures presented in one image appear partially or even disappear completely in another image. These local structures with missing correspondences are closely intermixed with the structures' large local deformations in the deformable image registration. The missing correspondences and local large deformations of local structures are called outliers in this paper, robustly determining the deformation elds or displacement elds representing the correct alignment of the local structures is still a challenging unsolved problem in many elds, such as machine learning [3], signal processing, medical imaging and image guided surgery [4]. Fig. 1 illustrates both missing correspondence and large local deformation problems in the two images (Figs. 1(a)-(b)) to be registered [28]. Compared with traditional registration approach (Fig. 1(c),(e)) that cannot align the normal and outlier structures via realistic and reasonable deformations, our method (Fig. 1(d),(f)) accurately aligns the normal corresponding regions, and maps the outlier regions to the right location but relax the deformations in the outlier-affected regions.

(b) (d) (f)

FIGURE 1. Comparison between using traditional and our image registration methods [28]. (a) and (b) The reference and moving images. (c) and (e) Traditional method introduces eye distortions (red arrow) in registered moving image, producing the conflicts of the deformation fields. (d) and (f) Our method accurately aligns the normal and outlier structures, relaxing the deformations in the outlier regions.

Deformable image registration can be formulated as the problem of globally searching for the optimal transformation T that minimizes the cost function $D(I_R, I_M \circ T) + S(T)$ between the reference image I_R and the moving image I_M. The global cost function consists of two terms: the data term D quantifies the difference and level of alignment between the two images, and the regularization term S regularizes the transformation toward favoring realistic and reasonable deformation solutions and seeks to address the ill-posed problem of deformable image registration. The data term is referred to as a matching criterion and includes intensity- and/or feature-based approaches. Feature-based methods [5], [6] usually establish dense deformation fields by interpolating the sparse correspondences between local features. Locating reliable local invariant features from the outlier structures remains an open problem in feature-based methods. Intensity-based approaches use the information of all image pixels to directly estimate the most flexible dense deformation (or displacement) field for each pixel (or voxel), which can better quantify and represent the matching accuracy of every point in local structure pairs.

Because global regularization introduces excessive flexibility, the intensity-based approach may favor unrealistic and unreasonable local deformations when it diffuses transformations from the structural to non-structural regions. In particular, the local structural regions make stronger contributions in the cost function minimization than do the non-structural regions, and thus, the transformation computation is easily affected by over-smoothing and is limited to the deformations in these structural regions. Furthermore, the multi-resolution strategy for the large local deformation problem has the following inherent disadvantages: the basic sub-sampling procedure in the multi-resolution strategy causes some displacement details of the edge structures to be removed; inaccurate initialization and outlier effects propagated from the coarse level lead to incorrect displacement estimations at finer levels in the refinement procedure; and the multi-resolution strategy cannot correctly predict the relatively large motion of a small-scale structure that exhibits a larger scale deformation than its own scale. Therefore, deformation models with spatially adaptive regularization [7] [9] have been proposed to address the varying deformation properties of local structures. Certain image segmentation-based works [10] [12] use an informative deformation prior for a specific region or tissue type to locally adapt the deformation field at various structures. The image segmentation is also used to address the missing correspondence problem by creating local artificial correspondences [13], [14], discarding the missing correspondences via cost-function masking [15], [16], or developing geometric metamorphosis [17] to separate the normal deformations from the outlier changes. While effective, these methods require explicit structure segmentations or initial outlier localizations. Recently, a low-rank and sparse decomposition technique [18], [19] has been able to separate the outlier structures from the ``healthy'' parts in a collection of images to be registered. Despite the success, this method may be limited in image sequence applications when separating out the sparse components that are not consistent with the low-rank structures [20].

Data-driven approaches include spatially adaptive regularization to distinguish motion differences for different regions. This data-driven strategy exploits a spatially adaptive transformation prior [21], local changes in intensities and deformation fields [22], or measures of local image reliability [23] to affect the local regularization strength. Recently, an optical flow estimation was able to integrate the sparse match propagation or aggregation [24] [26] into a global optimization framework to estimate the large displacements of small structures, while a large deformation diffeomorphic metric mapping [27] was successfully used to address large deformation problems but was highly influenced by the image intensity profile with missing correspondences. Nonetheless, these methods do not fully consider both the outlier structures (i.e., missing correspondences and large local deformations) and motion boundary removal nor do they consider the global-to-local contextual information of the corresponding saliency structures in the two images during the registration procedure.

Actually, the corresponding saliency structures [28] convey the most useful global-to-local contextual information during image registration. Finding the correspondences between the corresponding saliency structures is not only the starting point but also the ultimate goal of image registration. To efficiently match the structures from outliers, the joint saliency structures' contextual consistency is exploited in [28] for the spatially adaptive deformation construction. The idea of the spatially varying treatment [31] of joint-saliency and outlier voxels has also been successfully adopted in the feature-based DRAMMS approach [32] for challenging registration problems involving pathology-induced outliers. However, there exists limitation of utilizing a fixed kernel scale (or kernel width). By determining the sample size of the displacement vectors participating in the deformation construction, the sptially varying kernel scale for nonparametric regression is very important in controlling the balance between the structure matching accuracy and the smoothness of the local deformation fields. This assumption is confirmed by the fact that in density estimation studies in the literature, almost all the adaptive-scale kernels [33], [40] have been shown to be superior to fixed-scale kernels.

To design an appropriate kernel scale for JAKR, we assume that the kernel scale is adaptively selected according to the contextual information about the underlying structures and their displacement vectors. Generally, a large structure has more contextual sample pixels for propagating their deformations to construct the current sample pixel's deformation, whereas a small structure is confined to a small neighborhood to prevent the neighboring structures' distortions from spreading into the current estimate and smearing the motion boundaries. Moreover, large mismatches of local saliency structures require large kernel scales to include more contextual displacement vector samples for the deformation construction, whereas small mismatches of saliency structures need small kernel scales. Therefore, the kernel scale adaptively depends on not only the local size of the underlying structures to be matched but also the degree of mismatch between the local structures. Assuming that the mismatches of the local structures can be appropriately measured by the edge alignment degree of the structures, we propose an edge-aware adaptive-scale kernel for edge-aware deformation construction in JAKR to handle outlier structures and motion boundaries two common and difficult issues facing deformable image registration.

With the above-mentioned thoughts in mind, the proposed method represents three contributions: 1) After presenting a concise review on structure scale estimation for image processing, we propose a simple but effective boundary-aware local structure scale estimator: the estimator first segments the reference image into superpixel-based [34] multi-resolution structural regions; then, it calculates the boundary-aware structure scales of these regions in terms of the local variance of Gaussian smoothing through the Bayesian estimation and minimal description length criterion (MDL) [35], [36]. 2) We present an edge-aware mismatch scale of the overlapping structure pairs of two images, whereby we can judge and control the registration inaccuracy for the underlying structure pairs during the deformable registration procedure. 3) We propose an adaptive edge-aware kernel scale by combining the mismatch scale with the structure scale into the JAKR for deformable image registration. Therefore, the JAKR with the adaptive-scale kernels (JAKRAK) can iteratively guide the local structure deformations to not only achieve the accurate matching of small edge structures but also maintain smooth deformation fields for deformable registration with outliers. The experimental results demonstrate that the proposed JAKRAK method not only achieves state-of-the-art intensity-based registration performance but also achieves the best alignment of all challenging outlier structures. The background and the proposed method are elaborated in Section 2 and Section 3, respectively, followed by the experimental results in Section 4. The whole paper is discussed and concluded in Section 5.

II. BACKGROUND AND MOTIVATION

A. JOINT-SALIENCY STRUCTURE ADAPTIVE NONPARAMETRIC REGRESSION

Inspired by the success of nonparametric-regression-based [37] machine learning for signal reconstruction, we consider the deformable image registration as a nonparametric regression [28] to construct dense deformation fields from discrete deformation fields. This kernel-regression-based strategy is also implemented in deformable image registration [38] and has been recently adopted in optical flow estimation [39]. Suppose that we have some sparse and irregularly distributed deformation vectors $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^{P}$ given in the form

$$\mathbf{y}_i = \mathbf{z}(\mathbf{x}_i) + \mathbf{e}_i, \quad \mathbf{x}_i \in \Omega, \; i = 1, \cdots, P \tag{1}$$

where $\mathbf{y}_i$ is a sparse displacement vector (response variable) at position (explanatory variable) $\mathbf{x}_i$ and $\mathbf{z}(\cdot)$ and

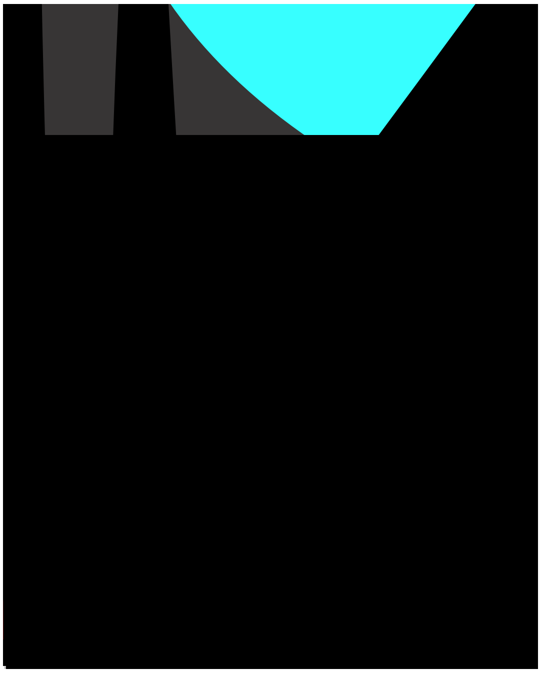

FIGURE 2. Multi-resolution flowchart of the proposed algorithm.

Fig. 2, which shows the three-step multi-resolution structure matching framework, with the different levels having their own resolutions but following the same procedure. First, the moving image I_M is deformed with an initial displacement field obtained via spatial interpolation of the output deformation field obtained on the previous level. The deformed moving image and the reference image on the current level are registered using block matching, with the point-wise mutual information serving as the local similarity measure. In the second step, with the JSM highlighting the overlapping JSSs for the deformation construction, this work estimates the scale of every reference structure and the scale (or degree) of mismatch between every pair of the underlying JSSs. With the anisotropic kernel representing the shape/orientation of the reference structure, we estimate the edge-aware adaptive-scale kernels for JAKRAK by combining the structure scale with the edge-aware mismatch scale; then, we use JAKRAK to construct the current deformation fields from the discrete displacement fields. Finally, the resulting global deformation for the iteration at the next level is composed of the initial deformation and current deformation from sampling the initial deformation fields.

B. SCALE ESTIMATION IN NONPARAMETRIC REGRESSION

The kernel scale of the nonparametric regression [37], [40] is crucial for signal reconstruction when addressing noisy data and outliers. A small scale corresponds to a smaller moving kernel for the nonparametric regression and therefore to noisier estimates, with higher variance and typically decreased estimation bias. A large kernel scale corresponds to smoother estimates, greater bias, and lower variance. Therefore, the kernel scale controls the trade-off between the bias and variance in the local estimation of the nonparametric regression. The are two types of approaches for kernel scale estimation in nonparametric regression. *Plug-in* methods [37], [40] calculate the ideal scale by estimating the bias and the variance in the estimation of the mean squared error (MSE) between the real signal and its approximation. The *quality-of-fit* statistics [37], [40], such as cross-validation and generalized cross-validation, are widely applied for the direct optimization of the estimation accuracy. The second estimation is defined by the accuracy criteria and is always related to data-driven methods disregarding the bias estimates or formulas for the ideal kernel scale selection, with the main goal to achieve an optimal accuracy that balances the bias and the variance of estimation. This work uses this accuracy-based estimation by taking the local structural matching contexts to boost the accuracy of deformable image registration with outliers.

We note that there are very different scale-estimation problems for 2D and/or 3D image analysis in pattern recognition and computer vision, where one goal is to describe the coarseness (or the optimal size for most spatial structures) of an image by any monotonically changing parameter. For example, the gradually changing time parameter t used in the diffusion process [41] [44] of an image in scale space is commonly treated as a scale parameter to globally control the smoothness of the whole image (or gradually remove the object detail within the image). The single global scale is widely used in many applications of multi-scale analysis: A single optimum scale [45], [46] based on Laplacian of Gaussian (LoG) analysis of an image is identified as the smoothing parameter for a normalized LoG filter to delineate blobs with similar sizes in medical images. The optimal scale based on a pre-estimation of the spatial and spectral statistics achieves satisfactory segmentation results with high homogeneity within the segments and high heterogeneity between the segments in multi-scale image segmentation [47]. An optimal scale t determines the stopping of TV-flow [48]-based diffusion to reduce image noise while preserving the maximally stable extremal region features for computer-aided detection. Recently, by decomposing the images into compact and region-boundary-aware superpixels, the structure-guided statistical textural distinctiveness approach [49] illustrates that considering texture at a single scale is sufficient for reliable salient region detection in natural images. However, the global scale estimation has an intrinsic limitation: the single coarseness for image structures in the whole image. The results of this class of algorithms might not be sufficient when the underlying fine and coarse image structures should be discriminatively analyzed in spatially adaptive schemes.

Rather than assuming a single global scale or multiple scales for a whole image using prior knowledge of the scales of the various objects of interest in the image, researchers always define a local scale to measure the size of local structures for each location of the image [50]. In estimating space-variant local scales, linear [41] [43] and nonlinear (such as morphological operations) [51] [53] scale spaces, Laplacian and Gaussian pyramids [54] are widely used to achieve multi-scale image representation. Among the various local scale estimation methods, the method proposed by Lindeberg [50] is widely used for image structures such as blobs, edges and ridges. The detected local scales may not have realistic meaning, as they simply detect local extrema over scales of normalized differential operators for the local image representation of certain sparse locations. The detected points, referred to as scale-space extrema, are sparsely distributed in the image to represent interest points, blobs, corners, edges, ridges and valleys, and they do not consider the actual structural information of the whole image. Alternatively, certain methods have utilized probabilistic approaches to estimate the local scales in an image for edge detection [55] and other low-level tasks [36] such as texture segmentation. An original strategy of local meaningful scale [56] detection relies on the asymptotic properties of perfect shape digitizations to detect what the relevant scales at which each point of the digital contours should be considered. The local adaptive scales for local pattern representation and texture segmentation are also explored in several works by maximizing the changes between the average gradients for different sizes of image blocks [57] or using total variation ows [53], Gabor ltering [58], and energy minimization models [59]. Recently, some segment-based scale selection strategies [60] [63] have been proposed to determine the varying sizes of local segments (or regions) in an image, such that all image pixels within a local segment satisfy a homogeneity or uniformity criterion. However, these local-segment-based scale estimation methods cannot automatically detect locally varying structures for spatially adaptive image processing.

As for image registration, most current image registration and optical ow approaches implicitly assume that the structures in both images are from the same scene and appear at the same scale. Nevertheless, image deformations often occur at different scales. Recently, Pai et al. [64] proposed multi-scale ow-based deformations by exploiting multiple kernels at different scales. A deformable spatial pyramid matching [65] was proposed to match pixels across scale differences coming from a discrete, pre-determined set of scales. The Deep-Matching [25] approach matched patches at several scales to overcome the lack of distinctiveness that affects small patches for optical ow computations. Tau and Hassner [66] established dense correspondences across structures with different scales to estimate the motion of small structures with large displacements and occlusions. However, the spatially varying scales of different geometric structures in the images are not considered in the above-mentioned works.

The locally adaptive scale in nonparametric regression is crucial in searching for an appropriate support of the local estimator for controlling the deformation smoothness and matching accuracy for the underlying saliency structures. On the one hand, a kernel scale is assumed to be spatially adapted to the underlying local structures (and their deformation contexts). For example, the matching of large local structures can use large kernel scales to reduce the deformation variance or increase the deformation smoothness compared with matching local small structures using a small kernel scale to reduce registration (or deformation) bias errors. Moreover, the nonparametric regression of deformation elds may also blur the boundaries and motion details of structures if using kernels crossing the boundaries between different structures. By considering a boundary-aware kernel scale in the kernel regression of deformation elds, we can preserve the intra-structure deformation smoothness and avoid inter-structure deformation smearing.

On the other hand, the gradually re ned deformation of small-scale structures with relatively large deformations will be mistakenly predicted by the deformation of large saliency structures at coarse resolution levels in multi-resolution registration. To achieve a correct prediction, testing the local structures' matching early in the registration procedure and as often as possible is the best way to guide the kernel-regression-based deformation construction toward accurate structure matching. Assuming that the local structure's matching can be validated by the alignment degree of the corresponding edges of the overlapping structures, this work proposes an edge-aware mismatch scale estimation to design an adaptive kernel scale in the registration coarse between the saliency structureg.

theve considemation,ne estimaed the locarly thenethod(underlaw)-372.28-28saliectwe edge-aware mismatch scale achieve accurate structure matching. a multi(resolution)4297(uperpixv)14.999(en)4297[34]npresntmatio the structure boundaries in renmenhing the rft- inoe different saliency structucal re structuro scalso the corresponding saliency structures the multi(resolution)-58[(le)25.003(v)14.995(els)3573.995(uro)-5 local ofan smootding witian the uperpixvei- the (presntede)-250(salienc)14.999(y)2493.995(re)14.956gtiong. at mismatch scale estimatio,e locarly iv ineonstyn-based reaccurce is rdsnattyabjecteevds rhe[67]e(and)-31((rfferectes)-371.003(tferin).s)-371.003(Assum tion,oest registration Destimaedtheanranieformation,s structures iv the registrationinywiveo40710.006jlhseartcn-based372.-28-28stan ranieformatio,nlocal local tmatet- resistance drectevr(tf-f)]TJ 0 -11.955 TD [(als) thisconsiden the localfctsn the local reacccorest in Section 3.2 to quantify the mismatch of underlying local structures.

III. METHODS

A. STRUCTURE SCALE ESTIMATION

The structure scale is considered as the size of every image structure corresponding to each image segment [60]. With the image structure being defined as a group of connected pixels with homogeneous features, structure scale estimation is formulated as a scale labeling assignment for each structure in optimal multi-scale segmentation, which contains the most homogeneous structures and the least edge-smearing mixed heterogeneous structures. To achieve scale invariance, the structure scale is computed in a neighboring region adaptive to the local structures in a multi-resolution image pyramid. Thus, the optimal structure scale refers to the optimal spatial extent or the optimal size of every local structure at every pyramid level.

To estimate the structural scales at every pyramid level, we first segment the image of every pyramid level into a set of superpixel structural units that adhere to the structure boundaries. We denote the whole image region as Φ and the local structures as S_i $(i = 1, \cdots, n)$, with $\Phi = \bigcup_{i=1}^{n} S_i$. The various structure units are then optimally smoothed to be internally homogeneous by the spatially varying Gaussian filters, with some variances in a discrete scale space. The variance σ^2 of the Gaussian filter controls the amount of Gaussian smoothing and thus the homogeneity of each structural region. With the minimal and maximal amounts of smoothness being controlled by the σ_1 and σ_m, respectively, in the discrete scale set σ_k $(k \in \{1, \cdots, m\})$, the optimal structure scale for each smoothed structure unit is obtained by maximizing its posterior probability from Bayes' theorem. Considering the scale coherence between neighboring structure units, we also use a Markov Random Field (MRF) model constraint to create a single large-scale labeling for the neighboring structure units with similar appearances. The final structure scale estimation is an optimal labeling image, with its segments achieving the most homogeneity within structural regions and the least edge-smearing in mixed structural regions.

Specifically, a scale space of the image $I(\mathbf{x})$ is first constructed by a the convolution operation, $I_\sigma(\mathbf{x}) = (I_0 * G_\sigma)(\mathbf{x})$, where $G_\sigma(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-|\mathbf{x}|^2/2\sigma^2}$ denotes the Gaussian kernel and the variance σ^2 is a certain scale parameter from the scale set σ_k. In this work, we assume that the largest scale in the scale set is 15 pixels and that the smallest scale is 1 pixel. Because an image can be decomposed into a smoothed component and a residual component through an anisotropic diffusion filter, the intensity of a local superpixel S_i can be represented by the smoothed component $I_{\sigma_k}(S_i)$ and the residual component

$$I(\mathbf{x}) = I_{\sigma_k}(\mathbf{x}) + \varepsilon_{\sigma_k}(\mathbf{x}), \quad \mathbf{x} \in S_i \tag{4}$$

The residual component ε_{σ_k} can be modeled as a zero-mean Gaussian random variable by the central limit theorem. Thus, the local structure scale estimation assigns a scale $\sigma_k, k \in \{1, \cdots, m\}$ from the scale space generated for each local structure S_i such that the following posterior probability achieves the maximum value

$$\begin{aligned} P(\sigma_k | S_i) &= \frac{P(\sigma_k) P(S_i | \sigma_k)}{P(S_i)} \\ &\propto P(S_i | \sigma_k) = \prod p(\mathbf{x} | \sigma_k), \mathbf{x} \in S_i \end{aligned} \tag{5}$$

where $P(S_i | \sigma_k)$ is the likelihood of the observed structural region S_i at scale σ_k, and $p(\mathbf{x} | \sigma_k) = P(I(\mathbf{x}) | \sigma_k)$ is the likelihood of the observed image at each pixel $\mathbf{x}$ at scale σ_k.

To estimate the likelihood of the observed image at each pixel, we use the well-known MDL criterion [35], [36] to relate the probability of an item with the length of the ideal code used to describe it, namely,

$$P(I|\sigma_k) = 2^{-L(I|\sigma_k)} \tag{6}$$

where $L(I|\sigma_k)$ denotes the description length of I based on its decomposition at scale σ_k. This description length can be expressed as $L(I|\sigma_k) = L(I_{\sigma_k}) + L(\varepsilon_{\sigma_k})$. On the one hand, the sampling theorem states that the number of samples needed for describing a Gaussian smoothed image is proportional to the Gaussian filter bandwidth in frequency space. Due to the uncertainty principle, this bandwidth is inversely proportional to σ_k^2. Therefore, the number of samples needed for describing the Gaussian smoothed image is controlled by the spatial variance of the Gaussian kernel. The description length of the smoothed component $L(I_{\sigma_k})$ is thus assumed [36] to be inversely proportional to σ_k^2 and can be written as $L(I_{\sigma_k}) \sim \frac{1}{\sigma_k^2}$. On the other hand, because the probability distribution of the residual $P(\varepsilon_{\sigma_k})$ is modeled as a zero-mean Gaussian distribution, the description length of the residual component $L(\varepsilon_{\sigma_k}) = -\log_2 P(\varepsilon_{\sigma_k})$ is proportional to $\varepsilon_{\sigma_k}^2$. Therefore, the local description length of I based on its decomposition at the scale σ_k and is estimated as follows:

$$L(I|\sigma_k) = \alpha(\frac{\beta}{\sigma_k^2} + \varepsilon_{\sigma_k}^2(\mathbf{x})) \tag{7}$$

where α and β are positive parameters [36] that depend on the coding precision in bits used to represent the smoothed image and on the assumed noise variance. Using equation (6), we further estimate $p(\mathbf{x} | \sigma_k)$ by the following equation:

$$\hat{p}(\mathbf{x}|\sigma_k) = A\mathbf{e}^{[-\alpha(\frac{\beta}{\sigma_k^2} + \varepsilon_{\sigma_k}^2(\mathbf{x}))]}, \quad \mathbf{x} \in S_i \tag{8}$$

where A is a normalizing constant, and α and β are empirically set to 1 in this work.

The scale field for the neighboring similar pixels is assumed to be inherently smooth due to the intra-structure homogeneity being usually visible in the natural world. Considering the scale coherence between similar neighboring pixels, we implemented the MRF model in the structure scale estimation. As a result, the nal structure scale is estimated as

$$\sigma_S = \arg\max_{\sigma_k} P(\sigma_k|S_i) + \lambda \sum_{\langle i,j \rangle} \delta(\sigma_k, \sigma_l) \exp(-(\mu(S_i) - \mu(S_j))^2)$$

$$\text{with } \delta(\sigma_k, \sigma_l) = \begin{cases} 1, & \text{if } \sigma_k = \sigma_l \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where i, j are the indices of local neighboring structures; $\mu(S_i)$ and $\mu(S_j)$ are the mean intensities of the local structures S_i and S_j, respectively; and σ_k and σ_l are the scales of S_i and S_j from the scale set, respectively. In equation (9), the rst term is the posterior probability of S_i, and the second term is a smoothness function of the local structure S_i and its neighboring local structure S_j. The second term prefers the same scale labeling for neighboring pairs of similar superpixel regions and avoids creating the same scale labeling for neighboring pairs of very dissimilar superpixel regions. The impact of MRF is controlled by the parameter λ, which is usually set to a small value (0.05) to avoid the over-smoothness that increases the structure scales of small local structures.

In Figs. 3(a)-(b), the reference and moving ower images are 384× 288 pixels, having a stamen lament with both missing correspondences and large local deformations in the top-right corner of the images. Figs. 3(f)-(h) show the superpixel-based structure scales for the multi-resolution saliency structures. The process roughly segments the foreground structural regions and background regions at the coarse resolution (Fig. 3(f)). With the increasing image resolution reducing the size of the superpixels and enhancing the image details, a small number of small structure scales are appropriately assigned to the small structures (e.g., the small petals, the petal boundaries and the stamen lament in the upper-right corner of the reference image in Figs. 3(g)-(h)), while a large number of moderate structure scales and the maximal structure scales are displayed for the foreground structural regions and the homogeneous regions, respectively.

B. MISMATCH SCALE CALCULATION USING JSM

Generally, mismatches often coincide with and are driven by intensity changes. Speci cally, homogeneous regions are always least informative and assumed to continuously have large areas of smooth intensity variations at the neighboring pyramid levels; thus, they maintain smooth deformation elds and result in the smallest driving in uences in the multi-resolution deformation construction. Conversely, edge structures, with their narrow areas of high contrast and ne detail, are most informative in driving the deformations and are easily changed or bleared during multi-resolution registration such that their deformation con icts (and the topology changes in the structures) can be widely found in the discrete displacement elds. Under the above-mentioned considerations, the edge structures' mismatches must be evaluated during the registration procedure to guide the kernel scale estimation for the nonparametric regression of the deformation elds.

With the JSM representing the matching degree of the underlying saliency edge structure pairs [28], the mismatch scales are inversely related to the JSM values for the adaptive nonparametric regression during the multi-resolution registration procedure. Figs. 3(c)-(e) show the multi-resolution edge-aware JSM with the color scale representing different joint saliency values. The high joint saliency values (in red) mean that the underlying pixel pairs come from the matched edge structures (or JSSs), whereas the low JSM values (blue and yellow-green) are from either unmatched structural regions (including outlier regions) or homogeneous regions. At every pyramid level, a zero or very small mismatch scale value is thus assigned to the corresponding structural regions with a high JSM value, whereas a large mismatch scale value is given to the unmatched structural regions with a low JSM value. Because they contribute the least to driving the304()14.995(gis)-237(de46malmoothn5 de46mabyde46mah

FIGURE 3. Flower images and their multi-resolution JSM, structure scales, mismatch scales and kernel scales. (a)-(b) The reference and moving images at the 384 x 288 pixels resolution. (c)-(e) multi-resolution JSMs, (f)-(h) multi-resolution structure scales, (i)-(k) multi-resolution mismatch scales, (l)-(n) multi-resolution kernel scales.

where JS represents the JSM at each overlapping pixel pair in the two images. The normalized mismatch scales (Figs. 3(i)-(k)) are thus computed to generally display three types of regions during the image registration: zero-mismatch-scale regions (in blue); low-mismatch-scale regions (in green and yellow-green) being from highly matched edge structure regions; and high-mismatch-scale regions (in red), indicating that the underlying regions are from mismatched structure regions or outlier regions.

C. LOCAL ADAPTIVE KERNEL SCALE

As mentioned above, the mismatch scale of the underlying edge structure pair is estimated to indicate the extent of possible deformation improvement by the kernel regression. Specifically, the moving saliency edge structures with high normalized mismatch scales require large deformation improvements so that the mismatch scales can be used as weights to linearly enlarge the underlying kernel scales gathering more sparse displacement vector samples for the desirable deformation construction. On the other hand, the moving saliency edge structures with low normalized mismatch scales need small deformation adjustments to achieve the desired deformation accuracy. Because the structure scale already indicates the size of the contextual structure, the kernel scale is not only proportional to the structure scale but also weighted by the mismatch scale of the underlying structure pair. Given the structure scale σ_s and the mismatch scale σ_m, we are ready to design the local kernel scale σ_d as

$$\sigma_d = \max\{\sigma_s \times \sigma_m, 1\} \tag{11}$$

where 1 avoids the local kernel scale being less than 1 pixel.

Figs. 3(l)-(n) illustrate the local kernel scales for the reference and moving images (Figs. 3(a)-(b)) for multi-resolution registration, with the color scale representing different normalized scale values. The large corresponding saliency structures with their surrounding homogeneous regions cover a relatively large range of kernel sizes (see the central regions of Fig. 3(l)) that correspond to the large areas of real image contents at the coarse resolution. These areas initialize the smooth deformation construction, while the small saliency structures at the fine resolution refine these deformations to increase the matching accuracy. With the small structures being gradually joined and assigned relatively large kernel scales in the iterative nonparametric regression, the background and homogeneous foreground regions gradually reduce their kernel scales to the smallest values for their expanding overlapping areas (see Fig. 3(n)) so that the deformation construction can be gradually adjusted to achieve the transition from deformation smoothness to deformation (or matching) accuracy.

Meanwhile, the multi-resolution kernel scales of outlier structures and small saliency structures are mostly dependent on their mismatch scales (Figs. 3(i)-(k)). Specifically, the outlier structures and the mismatched saliency structures always have relatively large kernel scales in the multi-resolution scheme. These relatively large kernel scales for the mismatched saliency structures and outlier structures gradually reduce their support regions (see the increasingly converging regions in the upper-right corners in Figs. 3(l)-(n)) in the image space to achieve the transformation from smooth deformation to accurate structure matching during the registration procedure.

According to the aforementioned analysis, the moving image's local saliency structures are gradually matched to the corresponding reference structures by iteratively selecting a locally adaptive scale for the local nonparametric regression. Fig. 4 illustrates why we prefer the edge-aware adaptive kernel scales to the fixed kernel scale in the proposed JAKRAK framework. The black stripes in the local `E' pattern at the top-center region of the hat are small-scale structures with large local deformations (see the reference and moving images in Figs. 4(a)-(b)). Figs. 4(e)-(f) show the two zoomed-in versions of the black stripes for the `E' patterns registered by the JAKR (Fig. 4(c)) and JAKRAK methods (Fig. 4(d)). Compared with the JAKR method introducing local irregular distortion in the stripes, the JAKRAK method can obtain accurate and smooth deformations of these local stripes. Figs. 4(g)-(h) present a performance comparison overview of the mesh deformation process (10-pixel vertex spacing) for the JAKR and JAKRAK methods with fixed-scale and edge-aware adaptive-scale kernels. JAKRAK can ensure a smooth adaptivity of the local mesh deformation to local structures of varying sizes. Specifically, the edge-aware adaptive-scale kernels for JAKRAK obtain smooth mesh deformations that are seamlessly consistent with the boundaries of local structures with varying sizes, while the fixed-scale kernels can produce more or less irregular mesh deformations that are not smoothly adaptive to the local structures (see Fig. 4(g)).

FIGURE 4. Performance comparison of using fixed-scale and adaptive-scale kernels. (a)-(b) The reference and moving images, (c) the

EMPIRE10 [69][2] data sets have been set up specifically for thoracic image registration. However, these data sets do not include outlier structures with both missing correspondences and large local deformations for challenging image registration. The objective and rigorous evaluation of the performance of challenging image registration is demonstrated in the work [70] using an in-house database containing eight patients with recurrent brain tumors. These pathology-bearing images introduce outliers with both missing correspondences and large local deformations and include two independent expert decisions on the corresponding landmark definitions and ROIs. Those landmarks and ROIs served as references for measuring the registration accuracy. However, due to the HIPPA regulation (Health Insurance Portability and Accountability Act), this database was publicly unavailable during the preparation of this manuscript.

Our algorithm has been implemented to support 2D/3D deformable image registration. In this section, we use a set of typical challenging 2D image pairs to validate the performance of the proposed JAKRAK method[3] by comparing it with the JAKR method, SparseFlow method (SF)[4] [26], DeepFlow (DF)[5] [25], Advanced Normalized Tools (ANTs)[6] [71] with greedy symmetric normalization diffeomorphic transformation and mutual information as similarity measure (AGS), and flexible variational non-linear intensity-based (FVNI) method[7] [72]. The JAKR, AGS and FVNI methods have demonstrated [28] state-of-the-art performances for deformable registration on challenging images with outliers. The parameters of the JAKRAK method are the same as those of the JAKR method [28] so that all the algorithms are set with the default parameters for achieving their best performances.

We use both landmark-based [70] registration error measurements and visual valuation to fully evaluate the performances of the six competing methods in the seven challenging image registrations. The landmark-based registration error measurement task measures the matching accuracy for the normal corresponding structures in the two images, while the visual valuation is simply for the outlier structures with both missing correspondences and large local deformations. Specifically, we not only zoom in on some small local structures in the registered moving images, displaying their deviation from the desired locations with several red crosses, but also manually select a large number of densely distributed landmark pairs from two experts in the two images for measuring the registration errors. Considering the uncertainty of manual landmark selection, we use the mean registration error (MRE) and standard deviation (SD) between the landmark pairs as the standard for registration evaluation.

[2]http://empire10.isi.uu.nl
[3]http://www.escience.cn/people/bjqin/research.html
[4]http://www.vision.ee.ethz.ch/~timofter/software/SparseFlow.zip
[5]http://lear.inrialpes.fr/src/deepflow/
[6]http://www.picsl.upenn.edu/ANTs
[7]http://hdl.handle.net/10380/3460

Due to the missing correspondences preventing the corresponding landmark definition, the landmark selection cannot include outlier features with missing correspondences when focusing on the easily identifiable corresponding locations at the JSS pairs. Lower average error distances and lower standard deviations imply a more accurate alignment of normal local structures. In most cases, the visual valuation can be perfectly consistent with the landmark-based registration evaluation. However, due to the inability to define the corresponding landmark pairs within and around the outlier regions with missing correspondences, the landmark-based registration evaluation cannot fully evaluate the real performances of these methods in matching outlier structures. This limitation is compensated by visually evaluating the zoomed-in display of the outlier structures in the following section.

The first experiment involves aligning two grayscale Mickey images (Figs. 5(a)-(b)) with an outlier doctoral cap in the moving image. There are large deformations characterizing Mickey's left thumb, left hand, right thumb (see the red boxes in Fig. 5), and right shoe as well as the right button on Mickey's belly. Therefore, the registration performance evaluations are largely dependent on the deformation results of these structures. Figs. 5(c)-(h) show that the JAKRAK (Fig. 5(c)) and DF methods (Fig. 5(f)) outperform other methods by perfectly deforming the local structures to the desired positions. However, the DF method inadequately diffuses the deformation into the wrist of the left hand. The SF method achieves good structure matching performance in the corresponding regions but deficient deformations in the left thumb, which resulted in a large variance in the following landmark-based registration evaluation. Obviously, the FVNI method has introduced a rounding image artifact around the right button on Mickey's belly. In contrast, the structure of Mickey's left hand is abnormally distorted by the AGS (Fig. 5(g)) method.

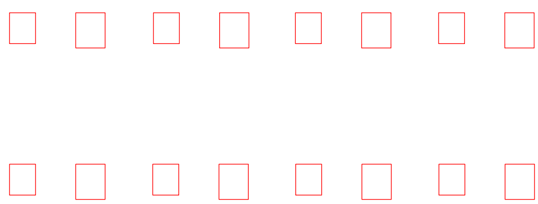

FIGURE 5. Mickey image registration with missing correspondences and large local deformations in the upper-right region. The boxed regions indicate the corresponding small structure regions with large local deformations. (a)-(b) The reference and moving images, (c) JAKRAK, (d) JAKR, (e) SF, (f) DF, (g) AGS, (h) FVNI.

The second experiment, displayed in Fig. 6 for flower image registration, includes both missing correspondences and large local deformations of small structures, where the outlier stamen filament in the right part of the reference image (Fig. 6(a)) has large local deformations driven by

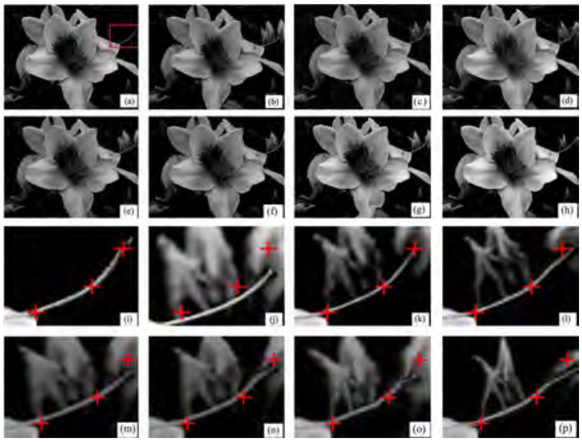

FIGURE 6. Flower image registration with the upper-right outlier regions of the stamen filament. (a)-(b) The reference and moving images, (c) JAKRAK, (d) JAKR, (e) SF, (f) DF, (g) AGS, (h) FVNI, (i)-(p) the corresponding zoomed versions of the red box regions for the stamen filaments (defined at (a)) in images (a)-(h), with the stamen filaments having the desired positions indicated by red crosses.

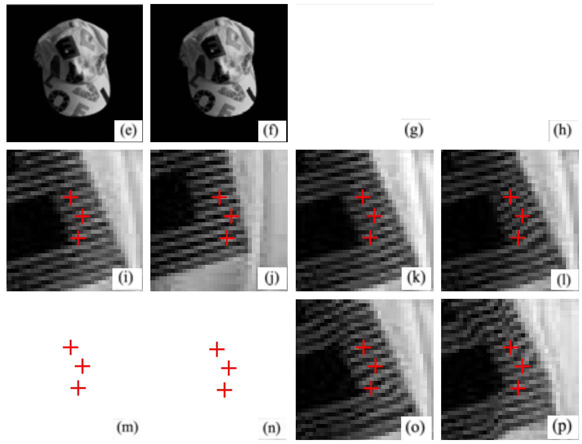

FIGURE 7. Hat image registration with the large local deformations of thin strips. (a)-(b) The reference and moving images, (c) JAKRAK, (d) JAKR, (e) SF, (f) DF, (g) AGS, (h) FVNI, (i)-(p) the zoomed versions of the boxed regions (defined at (a)) in images (a)-(h), with the thin stripes having the desired positions indicated by red crosses.

the movement of the center owers. In addition, some buds behind the stamen lament in the moving image (Fig. 6(b)) disappear in the reference image but appear in the moving image. Except for the AGS method (Fig. 6(g)) introducing excessive deformations in the bottom petal, the JAKRAK method and the other methods in Figs. 6(c)-(h) achieve good registrations, which properly deform the small-scale stamen lament and the large-scale petals simultaneously. However, the zoomed versions of the stamen laments in Figs. 6(i)-(p) can be used to distinguish the best performances of the JAKRAK method (Fig. 6(k)) in matching small-scale structures from outliers compared with the other methods because JAKRAK achieves precise structure matching in the tip of the stamen lament (in the red cross at the top-right corner). Although the JAKR, SF and DF methods (Figs. 6(l)-(n)) can obtain smooth registrations around the stamen lament, they are unable to achieve the desired large deformations in the tip of the stamen lament. Figs. 6(o)-(p) show that the AGS and FVNI methods introduce more or fewer artifacts and unacceptable deformations around the stamen lament. Due to the subsequent landmark-based evaluation having dif culty in de ning suf cient landmarks in the small-scale structures, the zoomed-in display of the visual evaluation performs better than the landmark-based evaluation in evaluating the challenging registration of small-scale structures with missing correspondences and large deformations.

A more challenging experiment is shown in Fig. 7, where the hat distortion deformed all the letters, with the black stripes in `E' in particular having large local deformations. Moreover, the missing `I' in the reference image (Fig. 7(a)) appears in the moving image (Fig. 7(b)). The main challenge in this experiment lies in the reasonable alignment of local small-scale structures such as the stripes in `E'. Because many tiny structures are close to each other, one structure's mismatching will directly affect the deformations of its neighboring local tiny structures and thus lead to poor structure alignment in a certain region. Figs. 7(c)-(h) show the registered moving images obtained by the six methods. The zoomed versions (Figs. 7(i)-(p)) of the stripes of `E' demonstrate that only the JAKRAK, DF and SF methods ((Figs. 7(k), (m), (n))) accurately match every small-scale structure (e.g., in the red crosses) of the stripes, whereas the JAKR, AGS and FVNI methods introduce excessive distortions in the stripes of `E'.

Four experiments involving matching pre- and post-operative brain tumor resection images were performed. In these experiments, surrounding normal brain tissues suppressed by tumor in the preoperative image (Figs. 8(1-a)-(1-b), 8(2-a)-(2-b)) expand after tumor resection, which introduces not only the missing correspondences of the tumor in the post-operative images but also the large local deformations caused by the brain shift. A desirable registration method should smoothly deform the tumor region and surrounding preoperative brain tissues (see the red boxes in Fig. 8) according to the post-operative image structures regardless of tumor resection. Figs. 8(1-c)-(1-h) and (2-c)-(2-h) are the registration results of the JAKRAK, JAKR, SF, DF, AGS and FVNI methods. In general, visual inspection shows that the JAKRAK and JAKR methods apparently perform better than the other four methods. Due to the intensity-based driving force having unexpected effects on the brain tumor resection regions, the AGS and FVNI (Figs. 8(1-g)-(1-h), 8(2-g)-(2-h)) methods produce more or less excessive deformation diffusions within the tumor region and/or some non-smooth distortions across the tumor region boundaries (e.g., in the red boxes or in the red arrows), while the SF and DF methods introduce some artifacts in certain edge structures (red arrows in Figs. 8(1-e)-(1-f)) as well as produce an inappropriate and/or insuf cient contraction of the tumor region and surrounding brain tissues (red boxes in Figs. 8(2-e)-(2-f)). Some diffusion artifacts are clearly displayed in the results of the FVNI method (Figs. 8(1-h) and (2-h)).

FIGURE 8. Two cases of brain tumor image registration. The red arrows indicate unrealistic distortions and/or some artifacts in the results. (1-a)-(1-b) and (2-a)-(2-b) The reference and moving images, (1-c)-(2-c) JAKRAK, (1-d)-(2-d) JAKR, (1-e)-(2-e) SF, (1-f)-(3-f) DF, (1-g)-(2-g) AGS, (1-h)-(2-h) FVNI.

Table 1 demonstrates the landmark-based evaluation of all these methods in the above experiments. Although the landmark-based evaluation is unable to re ect the matching performances of the methods for the outlier structures, Table 1 compares the matching accuracy for the normal corresponding structures in terms of the average registration errors, with standard deviations of approximately 40-50 landmarks.

TABLE 1. Landmark registration errors (Mean+SD) of the six methods for the corresponding structures in the images. The registration errors printed in italic indicate that these methods produced unrealistic artifacts in the deformable registration results according to visual inspection, whereas the registration errors printed in bold indicate the methods that achieved excellent performances in terms of both visual inspection and landmark-based evaluation.

The JAKRAK method achieved satisfying registration performances for all seven experiments, with registration errors of (1.23±0.81, 0.91±0.59, 0.93±0.81, 0.91±0.64, 0.93±0.63, 0.97±0.58, 0.89±0.55), while the registration errors of the DF, SF, JAKR and AGS methods are approximately (1.11±0.51, 0.92±0.83, 0.92±0.66, 0.98±0.71, 0.93±0.52, 0.97±0.62, 0.96±0.62), (1.25±1.23, 0.95±0.68, 0.93±0.82, 1.13±0.87, 0.92±0.68, 1.05±0.57, 0.99±0.59), (1.43±0.87, 0.97±0.81, 1.16±0.85, 0.96±0.63, 0.96±0.61, 1.02±0.74, 0.92±0.58) and (1.86±1.37, 1.08±0.92, 1.05±0.79, 0.91±0.61, 1.01±0.53, 1.15±0.73, 0.94±0.63), respectively. The FVNI method cannot perform well in three cases of these seven challenging image registrations. The landmark-based registration errors printed in italic indicate that these methods produced unrealistic distortions and/or artifacts in the deformable registration results based on visual inspection (Figs. 5-8). Nevertheless, the registration errors printed in bold indicate that these methods achieved excellent performances in terms of both visual inspection and inspection uorprinted in-1uctur(ter3533(mlo6(ein)dTJ3 to achieve a steady transition from smooth deformation to accurate structure matching during the registration procedure.

In general, the JAKRAK method is an effective deformation construction method for accurately matching small structures and outlier structures with smooth deformations compared with state-of-the-art methods. Many other deformable image registration methods for establishing accurate structure correspondences exist and may conceivably be used instead of the block matching method with JAKRAK for challenging image registration with missing correspondences and large local deformations. The challenging 2D/3D deformable image registration problem with missing correspondences and large local deformations is well known to be far from solved in many research elds. At present, there is no doubt that methods and algorithms from intelligent computing and machine learning for addressing this challenging outlier problem in deformable image registration are in high demand. Furthermore, we believe that further experimental studies are required to build ground truth 2D/3D image datasets with outlier structures.

ACKNOWLEDGMENT

The authors thank all the cited authors for providing the source codes used in this work. They are thankful to the anonymous reviewers for their valuable comments, which greatly helped to improve this paper.

REFERENCES

[1] A. Sotiras, C. Davatzikos, and N. Paragios, ``Deformable medical image registration: A survey,'' *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153 1190, Jul. 2013.

[2] D. F. P. Bouthemy and C. Kervrann, ``Optical ow modeling and computation: A survey,'' *Comput. Vis. Image Understand.*, vol. 134, pp. 1 21, May 2015.

[3] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, ``Quicksilver: Fast predictive image registration A deep learning approach,'' *NeuroImage*, vol. 158, pp. 378 396, Sep. 2017.

[4] S. Bayer, A. Maier, M. Ostermeier, and R. Fahrig, ``Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery,'' *Int. J. Biomed. Imag.*, vol. 2017, Jun. 2017, Art. no. 6028645.

[5] G. Wu, Q. Wang, H. Jia, and D. Shen, ``Feature-based groupwise registration by hierarchical anatomical correspondence detection,'' *Hum. Brain Mapping*, vol. 33, no. 2, pp. 253 271, 2012.

[6] O. Lobachev, C. Ulrich, B. S. Steiniger, V. Wilhelmi, V. Stachniss, and M. Guthe, ``Feature-based multi-resolution registration of immunostained serial sections,'' *Med. Imag. Anal.*, vol. 35, pp. 288 302, Jan. 2017.

[7] E. Suárez, C.-F. Westin, E. Rovaris, and J. Ruiz-Alzola, ``Nonrigid registration using regularized matching weighted by local structure,'' in *Proc. MICCAI*, vol. 2489. 2002, pp. 581 589.

[8] R. Stefanescu, X. Pennec, and N. Ayache, ``Grid powered nonlinear image registration with locally adaptive regularization,'' *Med. Imag. Anal.*, vol. 8, no. 3, pp. 325 342, 2004.

[9] I. J. Simpson *et al.*, ``Probabilistic non-linear registration with spatially adaptive regularisation,'' *Med. Imag. Anal.*, vol. 26, no. 1, pp. 203 216, 2015.

[10] M. Staring, S. Klein, and J. P. W. Pluim, ``Nonrigid registration with tissue-dependent ltering of the deformation eld,'' *Phys. Med. Biol.*, vol. 52, no. 23, pp. 6879 6892, 2007.

[11] D. F. Pace, S. R. Aylward, and M. Niethammer, ``A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs,'' *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2114 2126, Nov. 2013.

[12] S. Heldmann *et al.*, ``An image registration framework for sliding motion with piecewise smooth deformation,'' in *Proceedings in Scale Space and Variational Methods in Computer Vision* (Lecture Notes in Computer Science), vol. 9087, J.-F. Aujol, Eds. Berlin, Germany: Springer, 2015, pp. 335 347.

[13] M. Foskey *et al.*, ``Large deformation 3D image registration in image-guided radiation therapy,'' *Phys. Med. Biol.*, vol. 50, no. 24, pp. 5869 5892, 2005.

[14] M. Sdika and D. Pelletier, ``Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping,'' *Hum. Brain Mapping*, vol. 30, no. 4, pp. 1060 1067, 2009.

[15] M. Brett, A. P. Leff, C. Rorden, and J. Ashburner, ``Spatial normalization of brain images with focal lesions using cost function masking,'' *NeuroImage*, vol. 14, no. 2, pp. 486 500, 2005.

[16] P. Ripollés *et al.*, ``Analysis of automated methods for spatial normalization of lesioned brains,'' *NeuroImage*, vol. 60, no. 2, pp. 1296 1306, 2012.

[17] M. Niethammer *et al.*, ``Geometric metamorphosis,'' in *Medical Image Computing and Computer-Assisted Intervention MICCAI*. Berlin, Germany: Springer, 2011, pp. 639 646.

[18] X. Liu, M. Niethammer, R. Kwitt, N. Singh, M. McCormick, and S. Aylward, ``Low-rank atlas image analyses in the presence of pathologies,'' *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2583 2591, Dec. 2015.

[19] Z. Tang, Y. Wu, and Y. Fan, ``Groupwise registration of MR brain images with tumors,'' *Phys. Med. Biol.*, vol. 62, no. 17, pp. 6853 6868, 2017.

[20] M. Jin, R. Li, J. Jiang, and B. Qin, ``Extracting contrast- lled vessels in X-ray angiography by graduated RPCA with motion coherency constraint,'' *Pattern Recognit.*, vol. 63, pp. 653 666, Mar. 2017.

[21] P. Risholm, F. Janoos, I. Norton, A. J. Golby, and W. Wells, III, ``Bayesian characterization of uncertainty in intra-subject non-rigid registration,'' *Med. Image Anal.*, vol. 17, no. 5, pp. 538 555, 2013.

[22] B. W. Papie», M. P. Heinrich, J. Fehrenbach, L. Risser, and J. A. Schnabel, ``An implicit sliding-motion preserving regularisation via bilateral ltering for deformable image registration,'' *Med. Image Anal.*, vol. 18, no. 8, pp. 1299 1311, 2014.

[23] L. Tang, G. Hamarneh, and R. Abugharbieh, ``Reliability-driven, spatially adaptive regularization for deformable registration,'' in *Proc. Int. Workshop Biomed. Image Registration (WBIR)*, vol. 6204. 2010, pp. 173 185.

[24] T. Brox and J. Malik, ``Large displacement optical ow: Descriptor matching in variational motion estimation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500 513, Mar. 2011.

[25] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, ``DeepMatching: Hierarchical deformable dense matching,'' *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 300 323, 2016.

[26] R. Timofte and L. Van Gool, ``Sparse Flow: Sparse matching for small to large displacement optical ow,'' in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 1100 1106.

[27] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, ``Computing large deformation metric mappings via geodesic ows of diffeomorphisms,'' *Int. J. Comput. Vis.*, vol. 61, no. 2, pp. 139 157, 2005.

[28] B. Qin, Z. Shen, Z. Zhou, J. Zhou, and Y. Lv, ``Structure matching driven by joint-saliency-structure adaptive kernel regression,'' *Appl. Soft Comput.*, vol. 46, pp. 851 867, Sep. 2016.

[29] B. Qin, Z. Gu, X. Sun, and Y. Lv, ``Registration of images with outliers using joint saliency map,'' *IEEE Signal Process. Lett.*, vol. 17, no. 1, p. 91 94, Jan. 2010.

[30] Z. Gu and B. Qin, ``Nonrigid registration of brain tumor resection MR images based on joint saliency map and keypoint clustering,'' *Sensors*, vol. 9, no. 12, pp. 10270 10290, 2009.

[31] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, ``A nonrigid registration framework using spatially encoded mutual information and free-form deformations,'' *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1819 1828, Oct. 2011.

[32] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, ``DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting,'' *Med. Image Anal.*, vol. 15, no. 4, pp. 622 639, 2011.

[33] A. Serag *et al.*, ``Construction of a consistent high-de nition spatio-temporal atlas of the developing brain using adaptive kernel regression,'' *NeuroImage*, vol. 59, no. 3, pp. 2255 2265, 2012.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, ``SLIC superpixels compared to state-of-the-art superpixel methods,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274 2281, Nov. 2012.

[35] J. Rissanen, ``Modeling by shortest data description,'' *Automatica*, vol. 14, no. 5, pp. 465 471, 1978.

[36] G. Gomez, J. L. Marroquin, and L. E. Sucar, ``Probabilistic estimation of local scale,'' in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3. Sep. 2000, pp. 790–793.

[37] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, ``From local kernel to nonlocal multiple-model image denoising,'' *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, 2010.

[38] R. Gallea, E. Ardizzone, R. Pirrone, and O. Gambino, ``Three-dimensional fuzzy kernel regression framework for registration of medical volume data,'' *Pattern Recognit.*, vol. 46, no. 11, pp. 3000–3016, 2013.

[39] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, ``EpicFlow: Edge-preserving interpolation of correspondences for optical flow,'' in *Proc. IEEE CVPR*, May 2015, pp. 1164–1172.

[40] M. Köhler, A. Schindler, and S. Sperlich, ``A review and comparison of bandwidth selection methods for kernel regression,'' *Int. Stat. Rev.*, vol. 82, no. 2, pp. 243–274, 2014.

[41] A. Witkin, ``Scale-space filtering: A new approach to multi-scale description,'' in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 9. 1983, pp. 1019–1023.

[42] J. Koenderink, ``The structure of images,'' *Biol. Cybern.*, vol. 50, no. 5, pp. 363–370, 1984.

[43] J. Canny, ``A computational approach to edge detection,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[44] T. Lindeberg, ``Scale-space theory: A basic tool for analyzing structures at different scales,'' *J. Appl. Statist.*, vol. 21, nos. 1–2, pp. 225–270, 1994.

[45] M. Zhang, T. Wu, and K. M. Bennett, ``Small blob identification in medical images using regional features from optimum scale,'' *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1051–1062, Apr. 2015.

[46] A. Basset, J. Boulanger, J. Salamero, P. Bouthemy, and C. Kervrann, ``Adaptive spot detection with optimal scale selection in fluorescence microscopy images,'' *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4512–4527, Nov. 2015.

[47] D. Ming, J. Li, J. Wang, and M. Zhang, ``Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 28–41, Aug. 2015.

[48] J. Liu, S. Wang, E. B. Turkbey, M. G. Linguraru, J. Yao, and R. M. Summers, ``Computer-aided detection of renal calculi from noncontrast CT images using TV-flow and MSER features,'' *Med. Phys.*, vol. 42, no. 1, pp. 144–153, 2015.

[49] C. Scharfenberger, A. Wong, and D. A. Clausi, ``Structure-guided statistical textural distinctiveness for salient region detection in natural images,'' *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 457–470, Jan. 2015.

[50] K. Ikeuchi, Ed., *Computer Vision: A Reference Guide*. New York, NY, USA: Springer, 2014, pp. 701–713.

[51] P. Perona and J. Malik, ``Scale-space and edge detection using anisotropic diffusion,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[52] P. T. Jackway and M. Deriche, ``Scale-space properties of the multiscale morphological dilation-erosion,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 1, pp. 38–51, Jan. 1996.

[53] T. Brox and J. Weickert, ``A TV flow based local scale estimate and its application to texture discrimination,'' *J. Vis. Commun. Image Represent.*, vol. 17, no. 5, pp. 1053–1073, 2006.

[54] P. Burt and E. Adelson, ``The Laplacian pyramid as a compact image code,'' *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

[55] J. H. Elder and S. W. Zucker, ``Local scale control for edge detection and blur estimation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 699–716, Jul. 1998.

[56] B. Kerautret and J.-O. Lachaud, ``Meaningful scales detection along digital contours for unsupervised local noise estimation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2379–2392, Dec. 2012.

[57] D. Liu *et al.*, ``Medical image classification using spatial adjacent histogram based on adaptive local binary patterns,'' *Comput. Biol. Med.*, vol. 72, pp. 185–200, May 2016.

[58] J. Fdez-Valdivia, J. A. Garcia, J. Martinez-Baena, and X. R. Fdez-Vidal, ``The selection of natural scales in 2D images using adaptive Gabor filtering,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 5, pp. 458–469, May 1998.

[59] B.-W. Hong, S. Soatto, K. Ni, and T. Chan, ``The scale of a texture and its application to segmentation,'' in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

[60] X. Zhang and S. Du, ``Learning selfhood scales for urban land cover mapping with very-high-resolution satellite images,'' *Remote Sens. Environ.*, vol. 178, pp. 172–190, Jun. 2016.

[61] A. Souza, J. K. Udupa, and A. Madabhushi, ``Image filtering via generalized scale,'' *Med. Image Anal.*, vol. 12, no. 2, pp. 87–98, 2008.

[62] J. Sun and Z. Xu, ``Scale selection for anisotropic diffusion filter by Markov random field model,'' *Pattern Recognit.*, vol. 43, no. 8, pp. 2630–2645, 2010.

[63] Z. Ji, Y. Xia, Q. Sun, Q. Chen, and D. Feng, ``Adaptive scale fuzzy local Gaussian mixture model for brain MR image segmentation,'' *Neurocomputing*, vol. 134, pp. 60–69, Jan. 2014.

[64] A. Pai, S. Sommer, L. Sørensen, S. Darkner, J. Sporring, and M. Nielsen, ``Kernel bundle diffeomorphic image registration using stationary velocity fields and Wendland basis functions,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1369–1380, Jun. 2016.

[65] J. Kim, C. Liu, F. Sha, and K. Grauman, ``Deformable spatial pyramid matching for fast dense correspondences,'' in *Proc. IEEE Conf. CVPR*, Jun. 2013, pp. 2307–2314.

[66] M. Tau and T. Hassner, ``Dense correspondences across scenes and scales,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 875–888, May 2016.

[67] M. P. Heinrich, I. J. A. Simpson, B. W. Papież, S. M. Brady, and J. A. Schnabel, ``Deformable image registration by combining uncertainty estimates from supervoxel belief propagation,'' *Med. Image Anal.*, vol. 27, pp. 57–71, Jan. 2016.

[68] R. Castillo *et al.*, ``A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets,'' *Phys. Med. Biol.*, vol. 54, no. 7, pp. 1849–1870, 2009.

[69] K. Murphy *et al.*, ``Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge,'' *IEEE Trans. Med. Imag.*, vol. 30, no. 11, pp. 1901–1920, Nov. 2011.

[70] Y. Ou, H. Akbari, M. Bilello, X. Da, and C. Davatzikos, ``Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights,'' *IEEE Trans. Med. Imag.*, vol. 33, no. 10, pp. 2039–2065, Oct. 2014.

[71] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, J. C. Gee, ``A reproducible evaluation of ANTs similarity metric performance in brain image registration,'' *NeuroImage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[72] R. Werner, A. Schmidt-Richberg, H. Handels, and J. Ehrhardt, ``Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: A comparison and evaluation study,'' *Phys. Med. Biol.*, vol. 59, no. 15, pp. 4247–4260, 2014.

BINJIE QIN received the M.Sc. degree in measuring and testing technologies and instruments from the Nanjing University of Science and Technology, Nanjing, and the Ph.D. degree in Biomedical Engineering from the Shanghai Jiao Tong University, Shanghai, China, in 1999 and 2002, respectively. He was a Lecturer and also an Associate Professor with the Department of Biomedical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. From 2012 to 2013, he was a Visiting Professor with the Department of Computer Science, University College London, U.K. He is currently an Associate Professor with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. His group has developed image guided surgery system and small animal optical imaging system that are successfully used in clinical applications. His current research interests include biomedical imaging, image processing, machine learning, computer vision, and biomedical instrumentation.

ZHUANGMING SHEN received the B.S. degree in biomedical engineering from Tianjin Medical University, Tianjin, and the M.S. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2010 and 2013, respectively. His current research interests include medical image processing and visualization, MRI data analysis, and 3-D printing.

ZESHAN FU received the B.S. degree in biomedical engineering from Nanchang University, Nanchang, China, in 2014, and he is currently pursuing the master's degree with the School of Biomedical Engineering, Shanghai Jiao Tong University. His current research interests include medical image registration and computer vision.

ZIEN ZHOU received the B.Eng. degree in biomedical engineering, and the M.D. degree in imaging and nuclear medicine from Shanghai Jiao Tong University, Shanghai, China, in 2007 and 2011, respectively. His current research interests include medical imaging, clinical trials, and systematic reviews of randomize control trials.

YISONG LV received the M.Sc. degree in automatic instruments from the Kunming University of Science and Technology, Kunming, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1999 and 2003, respectively. His current research interests include image analysis, machine learning, and computer vision.

JINSONG BAO received the M.Sc. degree in mechanical engineering from Northeastern University, Shenyang, and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1999 and 2002, respectively. His current research interests include intelligent manufacturing, machine learning, and computer vision.