Applied Soft Computing 46 (2016) 851–867

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Structure matching driven by joint-saliency-structure

density estimation problem with Gaussian mixture model representing the feature points. The recent graph-based CPD work [7] explored graph centralities to bring topological information during the correspondence computation. Furthermore, feature-dependant finite mixture model was proposed in [8]. Though these algorithms are robust to the missing correspondences, due to the smooth transformation computation being very sensitive to the ambiguity in finding local feature correspondences, the joint estimation of correspondence and smooth transformation is still difficult to the outlier issues that have both missing correspondences and local large deformations.

By directly using the complete image data to recover dense correspondences at pixel-level precision, most intensity-based nonrigid registration approaches are regarded as global model based methods that are often formulated as global energy minimization problems with the energy being composed of an regularization term and a similarity term [10–14]. The relative weight of similarity term and regularization term can cause the well-known trade-off between the registration accuracy and the smoothness of the deformation field [12]. In the presence of outliers, the accurate and plausible local structure matching does not exist using whole-intensity driven transformation model. The relative spatial regularization can either cause non-smooth and implausible distortion in these outlier regions or introduce over-smooth and inaccurate mapping artifacts between the whole images. This outlier problem can be partly solved by implementing expectation and maximization algorithm to estimate the missing or partial data [10,11], using a locally varying weight between regularization and image similarity [13–18], creating artificial correspondences [19–24] and cost-function masking [25–27], or developing SIFT flow for large displacement [2]. Most of these approaches are largely dependent on outlier region segmentation without giving full consideration to both the missing correspondences and local large deformations.

At present, there is no doubt that methods and algorithms from intelligent computing and machine learning are very demanded to tackle this challenging outlier problem in structure matching. In these research fields, outliers mean the extreme observations substantially different from all other ones in the real data. In structure matching, the missing correspondences and local large deformations introduce the extreme local geometric and intensity differences between the two images

structure in order to gather more deformation vector samples of the same structure in the kernel regression, whereby the regression of smooth deformations can be locally compliant with the underlying local saliency structures without directing the deformation across the edges and corners of local structures.

This paper is an extended version of a preliminary conference work [38]. The extensions are augmented by more detailed method descriptions, more validations and more thorough discussions for some typical cases of 2D local structure matching, which are general and common to the computer vision and pattern recognition. The rest of this paper is organized as follows. The proposed algorithm is elaborated in Section 2 followed by typical experimental results in Section 3. The whole paper is concluded in Section 4.

2. Methods

2.1. C a e fi e . c . e a chi g ba ed - b c a chi g

The structure matching problem is formulated as an estimation of deformation of a moving image I_M with respect to a reference image I_R using transformation τ parameterized by an unknown displacement vector t (x), where x is the position vector in the reference image. The deformed moving image is defined by $I_M \circ \tau(\mathbf{x}) = I_M(\mathbf{x} + \mathbf{t}(\mathbf{x}))$. A displacement vector field describes deformation at all image pixels to maximize the similarity measure between the deformed moving image and reference image. To obtain the displacement field, our algorithm is built upon a three-step coarse-to-fine iterative block matching scheme proposed in [30]. This block matching framework is well known to deal with large deformation in nonrigid image registration. Fig. 1 displays the three-step coarse-to-fine iterative structure matching framework, where different levels have their own resolutions but the same procedure. First, the moving image I_M is deformed with an initial displacement field obtained by spatial interpolation of the output deformation filed obtained on the previous level. The deformed moving image and the reference image on the current level are registered by block-matching to obtain the discrete displacement filed. Second, after using the JSS adaptive kernel regression (see the green block diagrams in Fig. 1) with the kernel reflecting the edge-structure correspondence and reference structure's orientation, the structure matching scheme is to reconstruct dense current deformation field from the discrete displacement field. At last, the resulted global deformation for the iteration at next level is composed of initial deformation and current deformation by sampling the initial deformation field. The first level's initial global deformation is an affine transformation as the two images are typically already affine-registered before using our structure matching.

In the first step at each level, the moving image is deformed by an initial displacement field from previous level and sampled in the position of the sample of reference image to get deformed moving image. As in red block diagram at Fig. 1, for every pixel in deformed moving image, a block (i.e., a white neighborhood in Fig. 1) around the pixel of moving image I_M is taken and matched against a set of potential correspondent blocks in reference image. The block displacement (the red arrow in Fig. 1) that achieves the best similarity (0.65) between the blocks in moving and reference images, is stored as the discrete displacement for the original pixel. With the nature of realistic structure deformation imposing the spatial transformation to be a diffeomorphism, therefore the Jacobian of the local deformation field must be positive. This condition is preserved while only one pixel displacement is allowed for each pixel at each level of multi-resolution block matching. For every level of the multi-resolution pyramid, the final transformation is obtained by composing the transformation on the previous level with the one on the current level, in order to preserve the Jacobian positiveness condition [30].

In this study, we employ point-wise MI [39] as the local similarity measure for block matching. As a gold-standard registration criterion for two images to be registered, MI of the image intensity values of corresponding pixel pairs is maximal if the two images are geometrically aligned. The MI between the reference and moving images is defined and computed as:

$$\begin{aligned} \mathrm{MI} &= H(R) + H(M) - H(R,M) \\ &= \sum_{i_R, i_M} p(i_R, i_M) \log\left(\frac{p(i_R, i_M)}{p(i_R)\,p(i_M)}\right) \end{aligned} \tag{1}$$

where i_R and i_M are image intensities of reference and moving image. $H(I) = -\sum_i p(i) \log p(i)$ and $H(R,M) = -\sum_{i_R, i_M} p(i_R, i_M) \log p(i_R, i_M)$ are the entropy of the intensities of image I and the entropy of the joint intensities of two images, $p(i)$ is the intensity probabilities with $p(i_R) = \sum_{i_R} p(i_R, i_M)$ and $p(i_M) = \sum_{i_M} p(i_R, i_M)$, $p(i_R, i_M)$ is the joint probability density function (PDF) estimated by the joint intensity histogram $h(i_R, i_M)$.

When the global MI of whole images can be calculated as a sum of local contribution S_{MI}, defined for each individual image pixel pair i_R, i_M, the above global MI can be rewritten as:

$$\begin{aligned} \mathrm{MI} &= \frac{1}{N} \sum_{i_R, i_M} S_{\mathrm{MI}}(i_R, i_M) \\ S_{\mathrm{MI}}(i_R, i_M) &= \log\left(\frac{p(i_R, i_M)}{p(i_R)\,p(i_M)}\right) \end{aligned} \tag{2}$$

where the point-wise S_{MI} is calculated from the joint PDF corresponding to the whole images which have a total of N overlapping image pixels. For a given displacement, this joint PDF will be the global joint intensity histogram of the reference image with the displaced whole moving image. This is important for the multi-resolution block matching, where one pixel displacement can drastically change the PDF estimation. To estimate the optimal displacement for every pixel of moving image at multi-resolution block matching, the moving image is displaced by one pixel displacement to be overlapped with the fixed reference image, and the local evaluation of the MI can be computed just by summing the contributions of pixel pairs from the overlapped region. Therefore, at every pixel the current displacement corresponding to the biggest local MI are stored as the optimal displacement for the current deformation field estimation.

Although block matching has many advantages in obtaining the deformation of an image, implementing only this algorithm is still insufficient to avoid the irregularity of transformations such as tearing, folding and distorting in the challenging structure matching with outliers. Therefore, further reconstruction constraint is indispensable to be integrated into the registration procedure. To this end, local JSS adaptive kernel regression (as in the green block diagrams in Fig. 1) is applied into this discrete displacement fields in the second stage to reconstruct the dense deformation field: First, we compute structure tensor at each pixel position in the reference and the deformed moving images. Based on the structure tensor of reference image, the local structure adaptive anisotropic kernel is designed to spread along the direction of the local edge structure in reference image, whereby we use this anisotropic kernel's orientation information to adaptively steer the local kernel in kernel regression for dense deformation reconstruction. This scheme enables effective edge-aware deformation reconstruction that prevents the unrealistic deformation field to be smeared across object boundaries. Second, the center-surround dissimilarity between neighboring local structure tensors is explored to estimate the saliency map indicating the local saliency edge structure distribution of image. Based on the matching degree of local saliency map, the joint saliency map is constructed to emphasize the JSSs'

Fig. 1. Flowchart of coarse-to-fine structure matching framework. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

discrete displacement fields in the kernel regression for outlier removal. Details of the proposed approach are described in Sections 2.2–2.4. After the above-mentioned processing for each level in this coarse-to-fine framework, a smooth and dense deformation field is iteratively achieved as the global deformation at the last and finest level.

2.2. Ke e eg e i ba ed ca def a i ec . c i

Inspired by the successful applications in modern image deblurring and super-resolution imaging [30,31], we utilize kernel regression to reconstruct the smooth and dense deformation fields from the discrete displacement fields obtained through block matching. Suppose we have irregularly distributed displacement fields $\{y_i, x_i\}_{i=1}^{P}$ given in the form

$$y_i = T(x_i) + \varepsilon_i, \qquad x_i \in \Omega, \quad i = 1, \ldots, P \tag{3}$$

where y_i is a discrete displacement vector (response variable) at position (explanatory variable) x_i, $T(\cdot)$ describes the desired dense deformation fields in the moving windows (kernel) Ω with independent and identically distributed zero mean noise $\varepsilon_i = \varepsilon(x_i)$. In statistics, the function $T(\cdot)$ is treated as a regression of y on x, $T(x) = E\{y|x\}$. In this way, the reconstruction of nonrigid deformation fields is from the field of the regression techniques.

Suppose the point of interest x to be reconstructed is near x_i, then the regression of dense deformation field $T(x_i)$ can be approximated by a local Taylor series expansion

$$\begin{aligned} T(x_i) &\approx T(x) + \{\nabla T(x)\}^T (x_i - x) + \frac{1}{2}(x_i - x)^T \{\text{Hessian}[T(x)]\}^T \\ &\quad \times (x_i - x) + \cdots \approx \beta_0 + \beta_1^T (x_i - x) + \beta_2^T \text{vech}\{(x_i - x)(x_i - x)^T\} + \cdots \end{aligned} \tag{4}$$

where vech(·) is a half-vectorization operator of a symmetric matrix and $\{\beta_0, \beta_1, \beta_2, \ldots, \beta_N\}$ are $N+1$ unknown parameters to be estimated. As in the 2D case, $x = [\,_{\mu}1, \,_{\mu}2]^T$, we can easily estimate the unknown parameters as

$$\begin{aligned} \beta_0 &= T(x) \\ \beta_1 &= \left[\frac{\partial T(x)}{\partial x_1}, \frac{\partial T(x)}{\partial x_2}\right]^T \\ \beta_2 &= \frac{1}{2}\left[\frac{\partial^2 T(x)}{\partial x_1^2}, \frac{\partial^2 T(x)}{\partial x_1 \partial x_2}, \frac{\partial^2 T(x)}{\partial x_2^2}\right]^T \\ &\cdots \end{aligned} \tag{5}$$

Since we have known the discrete displacement vectors $\{y_i\}_{i=1}^{P}$, we can compute $\{\beta_n\}_{n=0}^{N}$ by finding the optimum solution of the following weighted least squares problem:

$$\min_{\{\beta_0, \beta_1, \beta_2, \ldots\}} \sum_{i=1}^{P} [y_i - \beta_0 - \beta_1^T (x_i - x) - \cdots]^2 K_H(x_i - x) \tag{6}$$

where $K_H(\cdot)$ is a kernel function (see the detail in the next section), which not only smoothes the approximation but also penalizes distance away from x.

In addition, if we assume $y=[y_1, y_2, \ldots, y_P]^T$, $b=[\beta_0, \beta_1^T, \ldots, \beta_N^T]^T$, and $K=\mathrm{diag}[K_H(x_1-x), K_H(x_2-x), \ldots, K_H(x-x)]$, then we can rewrite the optimization problem in a matrix form

$$\hat{b}=\underset{b}{\mathrm{argmin}}(y-Xb)^T K(y-Xb) \tag{7}$$

with

$$X=\begin{pmatrix} 1 & (x_1-x) & \mathrm{vech}^T\{(x_1-x)(x_1-x)^T\} & \cdots \\ 1 & (x_2-x) & \mathrm{vech}^T\{(x_2-x)(x_2-x)^T\} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_p-x) & \mathrm{vech}^T\{(x_p-x)(x_p-x)^T\} & \cdots \end{pmatrix}$$

and the least-squares estimation solution can be expressed as

$$\hat{b}=(X^T K X)^{-1} X^T K y \tag{8}$$

Because the zero-order Taylor series expansion known as the *Nada a –Wa* estimator is sufficient to reconstruct the displacement vectors, the estimation of the deformation field at x has the form

$$\hat{T}(x)=\hat{\beta}_0=\frac{\sum_{i=1}^{P} K_H(x_i-x)y_i}{\sum_{i=1}^{P} K_H(x_i-x)} \tag{9}$$

Since images have outliers, it is reasonable to consider uncertainty for each pixel. Therefore, we add a weight (or certainty) function c_i to Eq. (9)

$$\begin{aligned}\hat{T}(x)=\hat{\beta}_0 &= \frac{\sum_{i=1}^{P} K_H(x_i-x)\cdot(y_i\cdot c_i)}{\sum_{i=1}^{P} K_H(x_i-x)\cdot c_i} \\ &= \frac{K\otimes(y\cdot c)}{K\otimes c}\end{aligned} \tag{10}$$

The last line of Eq. (10) can also be expressed in the form of normalized convolution [30], where $\otimes$ denotes convolution operation.

Fig. 2 illustrates the smooth displacement vectors reconstructed for every pixel in brain tumor resection region using local kernel regression. Because block matching results inherently contain incorrect matches, which are exacerbated by the outliers in the tumor resection region. As a result, the conflicts between neighboring displacement vectors (see the several red circles shown in Fig. 2(a)) widely exist in the discrete displacement fields for the tumor region. Those displacement conflicts can easily introduce the topology change of structures, such as tearing and distorting. Fortunately, all the displacement vector conflicts are removed or smoothed by the local kernel regression in Fig. 2(b), where the displacement vectors having opposite directions fully disappear with the displacement magnitudes simultaneously being smoothed. Next, to match local structures in the presence of outliers, we design structural adaptive kernel function and robust weighing scheme for the moving window to further boost the accuracy and robustness of the local deformation reconstruction.

2.3. L ca . c . e ada i_e e- e f.- c i-

As a crucial component of local kernel regression, the shape of the kernel function (or moving window) determines the spatial distribution of samples which are gathered for the quality of the locally reconstructed signal. In principle, isotropic Gaussian kernels are mostly used as kernel functions in nonparametric regression. However, traditional isotropic Gaussian kernels are insufficient to cover more samples of the same modality along some specific orientations in signal reconstruction. Besides, Gaussian kernels' fixed scales and orientations can neither detect nor enhance edge structures. These factors easily cause diffusion across object boundaries. To remedy these restrictions, Pham et al. proposed an anisotropic kernel function to adapt its shape to the density of sampling [40]. Afterwards, Takeda et al. [41] utilized gradient covariance matrices to construct steering kernels, which have been proved to possess the ability to capture the edges of an image and be extremely robust to noise and perturbations of the data.

In dense deformation reconstruction, the desirable local kernel function is assumed to be extended along local structure orientation in the reference image so that it can gather more samples of discrete displacement vectors that correspond to the same saliency structure in the reference image.

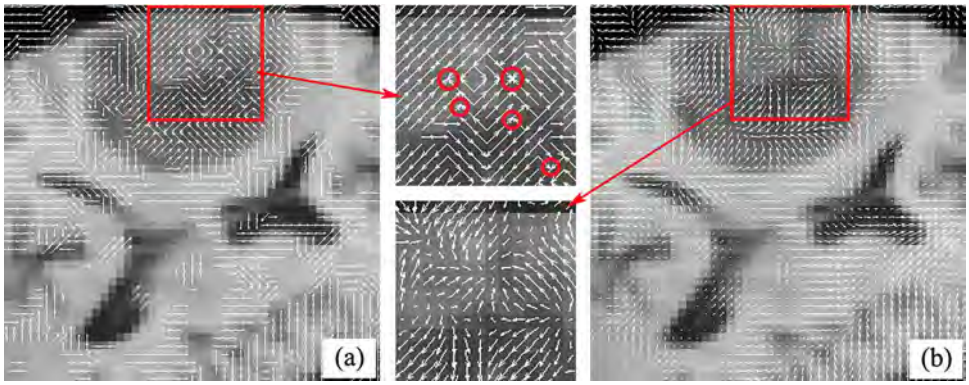

Fig. 2. Application of kernel regression in deformation reconstruction. (a) The conflicts between neighboring displacement vectors (several red circles) widely exist in the discrete displacement field for the tumor region. (b) The displacement vector conflicts are removed or smoothed by the local kernel regression along with the displacement magnitudes simultaneously being smoothed. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

of {u, v}, thus a local structure-adaptive Gaussian kernel in 2D case is designed as

$$a(x,x_0) = \frac{1}{2\pi\sigma_u\sigma_v}\exp\left[-\left(\frac{d_u^2}{2\sigma_u}+\frac{d_v^2}{2\sigma_v}\right)\right] \qquad (13)$$
$$d_u = \langle d, u\rangle, \quad d_v = \langle d, \quad v\rangle, \quad d = x - x_0$$

where d is the vector from x_0 to x, $\{d_u, d_v\}$ are the projections of the vector d on {u, v}, and the directional scales of the Gaussian kernel $\{\sigma_u, \sigma_v\}$ are determined by the anisotropy A as follows:

$$\sigma_u = \frac{\alpha}{\alpha+A}\sigma_c, \quad \sigma_v = \frac{\alpha+A}{\alpha}\sigma_c \qquad (14)$$

where $A = (\lambda_u - \lambda_v)/(\lambda_u + \lambda_v)$. The two directional scales of the Gaussian kernel can be adjusted by the parameter $\alpha > 0$ and the local scale σ_c. Specifically, α determines the eccentricity of the kernel while σ_c affects the number of discrete vectors that contribute to the reconstruction of continuous deformation vectors. For the sake of simplicity, the local scale is set to half the neighborhood window size for each kernel according to our experience. Therefore, we set $\alpha = 0.5$ and $\sigma_c = 1.5$ because we utilize a 3×3 pixel neighborhood window in our experiments.

Two kernel functions for two distinct image structures are displayed at the doll images in Fig. 3, where the crosses indicate two different kinds of typical image structures (blue cross for border and red cross for homogeneous region). In Fig. 3, the two pairs of orthogonal vectors indicate the main axes of their corresponding Gaussian kernels (blue cross at Fig. 3(c) and red cross at Fig. 3(d)). The length of the vector is determined by the scale in the direction of the vector.

Fig. 4 illustratively explains why we prefer local structure-adaptive kernel to isotropic kernel in our structure-adaptive kernel regression. The regions pointed by red arrows are small scale structures which have local large deformations. The directions of the displacement vectors (spaced every 5 pixels) in these small structures are inevitably conflicted with those of the large deformations of the neighboring structures. These deformation conflicts that are introduced by the opposite displacement vectors can easily cause tearing, folding or distorting of the local small scale structures. For example, the eyes in Fig. 4(c) display distortion owing to the conflict of the deformation directions displayed in Fig. 4(e). Comparatively, our local structure-adaptive kernels suppress the contributions of the displacement vectors which are not consistent with the structure orientation, the deformation conflicts are therefore removed in Fig. 4(f) with no distortion existing in the eyes at Fig. 4(d).

Fig. 4(h) demonstrates the overview of mesh deformation (10 pixels vertex spacing) in the local structure-adaptive kernel regression, which can produce the smooth adaptivity of local mesh deformation to the local anisotropic structures. Obviously, this local structure-adaptive kernel regression obtains smooth mesh boundaries which are consistent with the local structures' boundaries. However, isotropic kernels in kernel regression easily produce irregularly deformed local meshes which are not adaptive to the local structures, so that it is very difficult to identify boundaries of local structures from the non-smooth mesh deformation in Fig. 4(g).

2.4. R b . , eigh i. g echa. i . i. g JSM

In kernel regression, the weight function $c(x)$, between 0 and 1, specifies the reliability (or certainty) at x for the local estimation in a moving window and always describes the spatial dependence in the locally varying special context. With locally data-dependent weights, recently popularized and very effective image filters are developed in image and video processing field [28,29]. Since the region with saliency structure information has real influence over locally adaptive image processing based on kernel regression, Suarez et al. [30] proposed a weighting scheme in the kernel regression of registration transformations by using the scalar measure of a local structure in reference image. However, for nonrigid image registration, the salient structural regions in reference image may introduce non-corresponding salient structural regions at same locations in moving image. Therefore, the method proposed by Suarez et al. is most likely to assign wrong high weights to the saliency structures that have missing correspondences.

To avoid the above-mentioned mis-assignment and minimize the outlier effect on the deformation reconstruction, we propose a robust weighting mechanism by simultaneously considering the matching degree of local saliency structures in the overlapping parts of the two images. We deploy the concept of JSM into our robust weighting mechanism and pay more attention to the JSSs in the two images. The JSSs and their incurring deformations should be emphatically treated in

Fig. 3. Gaussian kernels designed for different image local structures. (a) Two labeled positions (red cross and blue one). (b) The scales and orientations of Gaussian Kernels in corresponding positions. (c) Gaussian kernel for the region with blue cross. (d) Gaussian kernel for the region with red cross. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

Fig. 4. Comparison between using isotropic kernels and using local structure-adaptive kernels in kernel regression based deformation reconstruction. (a) and (b) The reference and moving images. (c) and (d) Isotopic kernels in kernel regression introduce eye distortions (red arrow) in registered moving image, while local structure-adaptive kernels can remove these distortions. (e) and (f) Isotopic kernels produce the deformation conflicts in the displacement vector fields, while local structure-adaptive kernels remove these conflicts. (g) and (h) Isotropic kernels easily produce irregularly deformed local meshes, while local structure-adaptive kernels achieve the smooth adaptivity of local mesh deformation to the local anisotropic structures. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

for a given point x_0 and its neighborhood Ω, the saliency value $S(x_0)$ at x_0 in a saliency map can be computed through

$$S(x_0) = \text{avg} \sum_{x \in \Omega} \|\text{LST}(x) - \text{LST}(x_0)\|_D \tag{15}$$

where $\|\cdot\|_D$ defines a distance metric describing the dissimilarity between two LSTs, which is detailed in the following section. The operator avg computes the average of the dissimilarities within the neighborhood Ω of x_0. Traditional tensor similarity measures such as fractional anisotropy (FA) and cosine similarity measure are not appropriate for defining tensor-based saliency operator because they only compare either scales or orientations of two tensors. Fortunately, a few improved tensor similarity measures computing both scales and orientations of two tensors have been reported. In [44], Zhang et al. introduced diffusion tensor metric, which is defined as

$$\|T_1 - T_2\|_D = \sqrt{\frac{8\pi}{15}(\|T_1 - T_2\|_C^2 - \frac{1}{3}\text{Tr}^2(T_1 - T_2))} \tag{16}$$

where $\|T_1 - T_2\|_C = \sqrt{\text{Tr}(T_1 - T_2)^2}$ is the Euclidean distance between two tensors $\{T_1, T_2\}$, Tr is the operator for computing the trace of matrix.

In a tensor-based saliency map, the saliency value is a general representation of the local edge structure distribution in an image. Low saliency values always appear in the homogeneous and background regions while high saliency values are assigned to the edge structures owing to the highlighted contrast among neighboring LSTs in these regions. After the two normalized saliency maps were achieved to indicate the local edge structure distribution, JSM is ready to describe the matching degree between the two saliency maps at every pixel pair in the overlapping regions of the two images. Given a point x_R in the reference image and its corresponding point x_M in the moving image after initial transformation, their joint-saliency value in a JSM is defined as

$$JS(x_R, x_M) = \min\{S_R(x_R), S_M(x_M)\} \frac{A \cdot B}{B + \|\text{LST}(x_R) - \text{LST}(x_M)\|_D} \tag{17}$$

where $\{S_R(\cdot), S_M(\cdot)\}$ denote the saliency values in the saliency maps of the reference and the moving images, respectively. The empirical parameters A and B are used to normalize the JSM values into a final value between 0 and 1 for the definition of weight function $c(x)$. In our experiments, $A = 10$ and $B = \frac{1}{2}\max(\|\text{LST}(x_R) - \text{LST}(x_M)\|_D)$, in which B is half the maximum dissimilarity value between two LSTs in the whole overlapped regions of the two images. It should be noted that it may introduce a situation that both of the corresponding pixels are set to high saliency values in the saliency maps, while their local variations of gradient orientations are in fact totally different. To avoid this situation, we also consider the dissimilarity measure between $\text{LST}(x_R)$ and $\text{LST}(x_M)$ at the denominator in Eq. (17).

Fig. 5 shows some examples of normalized JSM with the color scale representing different joint saliency values. The high joint saliency values represented by red color suggest that the underlying pixel pairs come from the JSSs. On the contrary, the regions with low JSM values are rendered in blue color, which indicates that the

Fig. 5. JSM Examples with color scale representing different joint saliency values. Each column shows one case. The reference images and the moving ones are shown at the top and middle rows. Their JSMs are displayed at the bottom row where the red regions correspond to the higher joint saliency values while the blue regions correspond to the lower joint saliency values. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

Fig. 6. The reference and the moving images and their gradient magnitude, largest eigenvalue profiles for GST and LST, and the JSM magnitude. (a) and (b) The reference and moving images. (c)–(e) Gradient magnitude profiles, largest eigenvalue profiles of GST and largest eigenvalue profiles of LST of the red line in (a). (f) and (h) Gradient magnitude profiles, largest eigenvalue profiles of GST and largest eigenvalue profiles of LST of the red line in (b). (i) and (j) Saliency value profiles of the red lines in (a) and (b). (k) JSM value profiles of the red lines in (a) and (b). (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

underlying pixel pairs originate from either homogeneous regions or outlier regions. The discrete displacement vectors in these red JSS regions are expected to contribute more to the kernel regression than the blue regions having low JSM values, this weighting scheme is therefore called JSS adaptive kernel regression for nonrigid image registration.

The JSM in our study mainly responds to the corresponding high-gradient edge pixels. However, it does not simply highlight the common image gradients in the two images. Fig. 6 vividly presents the differences between the image gradient magnitude, the largest eigenvalue of GSTs and LSTs, the saliency value and the JSM value profiles of the same red line at the two images (Fig. 6(a) and (b)). For easy comparison, the range of ordinates in Fig. 6(c)–(k) are bound to [0, 1]. As shown in Fig. 6, the image gradient features in Fig. 6(c) and Fig. 6(f) for the two images are very sensitive to noise and do not agree with each other at each overlapping location. The noise sensitivity is gradually reduced by using the GSTs (Fig. 6(d) and (g)) and the LSTs (Fig. 6(e) and (h)). The saliency values of the two images in Fig. 6(i) and (j) are robust to noise due to their computing the regional contrast of LSTs through Eq. (15). Moreover, the structural image information in a large region is also comprehensively considered according to Eq. (15). As a result, the JSM values (Fig. 6(k)) computed through the saliency values can accurately preserve the JSSs in larger capture range with smaller variability than the image gradients. Therefore, the effectiveness of JSM is strongly confirmed in Fig. 6.

Because of the outliers introduced by missing correspondences, local large deformations and incorrect block matching, the dense deformation fields cannot be simply interpolated from the discrete displacement vectors in block matching. The JSS adaptive kernel regression is used to reconstruct the dense deformation fields from the discrete displacement vectors, i.e., smooth the outlier effects on the deformation reconstruction. Due to the expected deformations in the outlier region being consistent with its neighboring deformations, the JSM values in the neighboring regions are used to assign different weights to the different displacements of the neighboring structures, only those neighboring deformations with high JSM values indicating the consistency in structure orientations are given high weights in kernel regression based deformation reconstruction.

Fig. 7 illustrates an improvement on the deformation field reconstruction after introducing the JSM-based local JSS adaptive kernel regression. The region pointed by red arrows in Fig. 7(c) and (d) is outlier region with missing correspondences and local large deformations. Without JSM-based robust weighting mechanism, the converged displacement vectors (5 pixels spacing) from conflicting directions (see Fig. 7(e)) in the outlier region spread the distortion effect into the eye region (see Fig. 7(c)). Due to the JSM-based robust weighting mechanism introducing weighted smoothing effect on the magnitudes and directions of displacement vectors (see Fig. 7(f)), moving image's eye distortion is removed (see Fig. 7(d)) with the outlier structure deformations being simultaneously matched to those of the reference image (see Fig. 7(f)). Compared with the deformation meshes (10 pixels vertex spacing) in Fig. 7(g), the deformation meshes at Fig. 7(h) display the overall smoothness improvement for the structural deformations at the outlier structures due to the JSM-based robust weighting mechanism.

3. Experimental results

To validate the proposed algorithm[3] on some challenging images with missing correspondences and local large deformations, we make a comparison among the proposed algorithm and other state-of-the-art intensity-based nonrigid registration methods using some typical 2D image sets, where the moving images with a variety of outlier structures should be matched into the reference images.

We choose ANTs[4] with geodesic symmetric normalization (Syn) diffeomorphic transformation and MI (AGS1) [45], ANTs

[3] http://www.escience.cn/people/bjqin/research.html
[4] http://www.picsl.upenn.edu/ANTs

Fig. 7. Matching performance improvement through JSM-based robust weighting mechanism. (a) and (b) The reference and moving image. (c) and (d) The conflicting deformations in the outlier region (red arrows) spread the distortion

Fig. 8. Chinese character image registration. Our method, AGS2 and the FVNI methods have clearly outperformed the other methods. (a) and (b) The reference and moving images, (c) ours, (d) AGS1, (e) AGS2, (f) AMI, (g) DDD, (h) BMI, (i) AMM, (j) EPPM, (k) LDOF, (l) FVNI. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

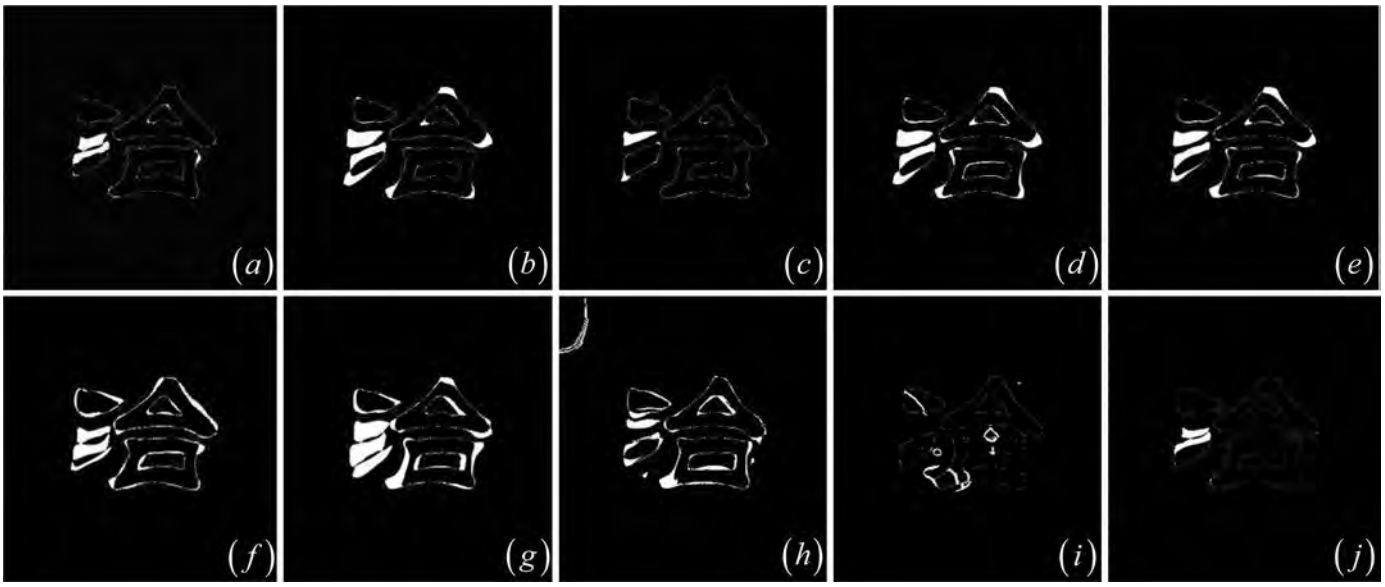

Fig. 9. Difference images of Chinese characters from the 10 matching results. (a) Ours, (b) AGS1, (c) AGS2, (d) AMI, (e) DDD, (f) BMI, (g) AMM, (h) EPPM, (i) LDOF, (j) FVNI.

Fig. 10. Masks and Landmark pairs for the images. (a) and (d) The black mask definitions for Chinese characters, Mickey, brain tumor and flower images. (e)–(g) The corresponding landmark pairs for Mickey, brain tumor and flower images in landmark-based registration error evaluation. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

middle stroke (outlier) in the left part of moving image has missing correspondence (see the red boxes in Fig. 8(a) and (b)). Moveover, the triangular and the rectangular openings at the right part of the character are narrowed. The local large deformations are apparent at lower

Fig. 11. Mickey image registration. The proposed and FVNI methods outperform the other methods. (a) and (b) The reference and moving images, (c) ours, (d) AGS1, (e) AGS2, (f) AMI, (g) DDD, (h) BMI, (i) AMM, (j) EPPM, (k) LDOF, (l) FVNI. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

correspondences into the tumor region of the post-operative image but also incurs local large deformations that are caused by brain shift after tumor resection. A successful registered result of this case should smoothly deform the tumor region and surrounding pre-operative brain tissues (see the red boxes in Fig. 12) according to the post-operative image structures regardless of tumor resection. Visual inspection has revealed that the proposed and

Fig. 12. Brain tumor image registration. The proposed and AMI methods outperform other methods. (a) and (b) The reference and moving images, (c) ours, (d) AGS1, (e) AGS2, (f) AMI, (g) DDD, (h) BMI, (i) AMM, (j) EPPM, (k) LDOF, (l) FVNI. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

gets better deformations for most structures than the DDD and BMI methods (Fig. 13(g) and (h)), but also apparently introduces some artifacts in the petals of the registered moving image. The EPPM method (Fig. 13(j)) still produces bad topology change in the petals.

The MREs and SDs of the manually selected landmark pairs (Fig. 10(e)–(g)) for the 10 methods in the three grayscale image registration are listed in Table 1. Because the EPPM method produced topology changes and complex artifacts in all three experiments such that we cannot find the corresponding landmarks in the deformed moving image, EPPM's registration errors are unavailable in Table 1. The proposed and FVNI methods for the Mickey image matching achieve the smallest and second smallest registration errors of 1.27 ± 3.09 and 1.76 ± 2.96 pixels, respectively, while the registration errors of AGS1, AGS2, AMI, DDD, BMI, AMM, and LDOF methods are greater than or equal to 1.87 ± 3.11 pixels. These landmark-based evaluation results are consistent with the above-mentioned visual evaluation results.

As for the brain images, the missing correspondences resulted from brain tumor resection prevent sufficient

Fig. 13. Flower image registration. The proposed, FVNI and AMM methods perform better than other methods. (a) and (b) The reference and moving images, (c) ours, (d) AGS1, (e) AGS2, (f) AMI, (g) DDD, (h) BMI, (i) AMM, (j) EPPM, (k) LDOF, (l) FVNI. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

structures. We try to analyze several factors that lead to the success of these four methods as follows: first, due to the transformation model with the similarity measure and the optimization strategy being three important components of nonrigid image registration, the high degree of freedom allowed by a deformation mechanism, such as the most flexible deformation model by defining a displacement vector to each pixel as used in our method and FVNI, the large deformation diffeomorphic metric matching (LDDMM) model as used in AGS1 and AGS2 and the B-spline based free from deformation (FFD) model as used in BMI, is assumed to be an important factor contributing to the high registration accuracy in matching challenging structures. However, the high degree of freedom in LDDMM model for AGS1 and AGS2 and FFD model for BMI method also can result in excessive deformations at our experiments because the global intensity-based energy minimization optimization can easily mislead the outlier structure matching. Second, the symmetric diffeomorphic setting in AGS2 and FVNI also contributes to the accuracy and robustness in outlier structure matching. The proposed method adopts one pixel displacement in each level of multi-resolution block matching to guarantee diffeomorphism condition. Generally, these diffeomorphic approaches prevent singularities from arising in the displacement field. A proper transformation model, which is of sufficient degrees of freedom but well regularized by the diffeomorphism or even the symmetric diffeomorphism design with regard to the two images, is the first choice for challenging outlier structure matching.

Third, the symmetric normalization feature as introduced in AGS2 has seemed to at least partly contribute to its accuracy and robustness in matching Chinese character. Specifically, in the binary structure matching with missing correspondences and local large deformations, where two binary images have obvious intensity contrast between the white Chinese character and black background, AGS2 has best accuracies in matching the normal regions (the lower left stroke in Fig. 8(e)) and in the immediately neighboring abnormal regions. This is perhaps due to the symmetric normalization setting, which deforms both images to “hidden middle template” between the two images. That is, a difficult problem of minimizing the full distance (dissimilarity measure) between the two challenging images is decoupled into two relatively simpler subproblems that minimize the half distances between either image and the hidden middle template. Fourth, the same non-parametric variational

Table 2
Computation runtime in seconds for the 10 registration methods (Inter(R) Core(TM) i5-4460 Quad-Core 3.2 GHz CPU, RAM 4.0 GB).

Cases	Ours	AGS1	AGS2	AMI	DDD	BMI	AMM	EPPM	LDOF	FVNI
1	17.67	3563	30.67	13.99	4.42	11.41	13.99	0.79	41.23	4.02
2	18.83	3521.5	54.64	21.52	4.99	16.14	17.29	0.90	55	2.99
3	35.86	3761.5	33.88	15.69	8.97	32	14.94	0.90	84.43	4.53
4	14.67	3271.3	26.16	8.56	4.08	30.57	2.85	0.74	68.48	2.82

Presently, the state-of-art nonrigid registration methods do not explore the consistencies which exist in the normal structures and their local deformations surrounding the outlier regions. For example, simply excluding the outlier structures in minimization of cost function between the images by AMM's cost function masking strategy is not enough to accurately match local saliency structures with missing correspondences and local large deformations. However, the structures and their deformations' consistencies are fully explored by kernel regression of local deformation fields in the proposed method to boost the structure matching performance.

Table 2 lists the computation time needed for the 10 algorithms at the four image registrations, where the image resolution of case 1–3 is 372×392 pixels, and that of case 4 is 384×288 pixels. All the 10 methods are operated on a PC of Intel Core i5-4460 Quad-Core 3.2 GHz CPU with 4 GB memory.

4. Conclusion and discussion

In this paper, we compute saliency map of each image based on the dissimilarity between neighboring local structure tensors to indicate the local saliency structure distribution, and propose new JSM to describe the matching degree between the two saliency maps at every pixel pair in the overlapping regions of the two images. The JSM guides the local adaptive kernel regression for the dense deformation field reconstruction in accurately matching local structures while suppressing outlier effects in the nonrigid image registration. The proposed method also makes the regression kernel not only locally adaptive to the JSSs in the two images but also to reference image's saliency structures. These strategies enhance the proposed algorithm in achieving the accuracy and robustness of structure matching with missing correspondences and local large deformations.

Nevertheless, there are still some issues that need to be further addressed in our future work. Firstly, the local scale of regression kernel in Section 2.3 is set as a constant value (1.5) for the sake of simplicity at each anisotropic Gaussian kernel. As shown in some registered results, however, the deformations of some small scale structures (such as the eyes in the doll image at Fig. 7) are affected somewhat badly by the deformations of the surrounding large scale structures. This situation is caused by the unchangeable local scales. Indeed, the local scale (the width of kernel) controls the number of discrete displacement vectors contributing to the reconstruction of dense deformation vectors. Therefore, the choice of local scale significantly affects the final registration result. A self-adjustable local scale according to the local structure properties is expected to automatically adjust the number of discrete displacement vectors participating in the deformation reconstruction. To the best of our knowledge, almost no attention has been paid on the self-adjustable local scale for nonrigid image registration during the last decade.

Secondly, as for the symmetric diffeomorphism of nonrigid image registration [45,48], the missing correspondences and local large deformations make it unrealistic for nonrigid registration method enforcing the symmetric diffeomorphic local structure matching. Meanwhile, our current work uses the full information from reference image's structure to guide the anisotropic kernel design in kernel regression, which prevents the symmetric diffeomorphic registration strategy in our work. However, symmetry is desired for estimating the distance between reference and moving structures in the two images and makes results independent of arbitrary decisions about which image is "reference" or "moving". However, we can initialize our method with some symmetric diffeomorphic registration algorithms to find the local structure's optimal symmetric diffeomorphic matching for the subsequent JSS adaptive kernel regression.

At last, further experimental researches are required to give thorough tests on 3D structure matching. At present, the objective and rigorous evaluation of 3D structure matching performance was hampered by the absence of public appropriate reference standard 3D image data sets. Though there are two public 3D image data DIR-Lab [50][10] and POPI [51],[11] both of them cannot meet the following two challenges. The desired reference 3D image data to test 3D structure matching should reflect the two major challenges: (1) presenting the outlier structures with both missing correspondences and local large deformations; (2) having expert-defined corresponding structural landmarks in these image data sets, with repeat structure matching performed by multiple observers to estimate the spatial variance in feature identification.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61271320 and 60872102). The authors would like to thank Eduardo Suarez, Rafael Nebot for creating the block matching framework and all authors for opening source codes used in the experimental comparison in this work. The authors thank the open source ANTs project, MIPAV, Diffeomorphic Demons, 3D Slicer and ITK (http://www.itk.org/) project. The authors would like to thank Zeshan Fu for his evaluation work. We are thankful to the anonymous reviewers for their valuable comments that greatly helped to improve this paper.

References

[1] A. Sotiras, C. Davatzikos, N. Paragios, Deformable image registration: a survey, IEEE Trans. Med. Imag. 32 (7) (2013) 1153–1190.
[2] T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, IEEE Trans. Pattern Anal. Mach. Intel. 33 (3) (2011) 500–513.
[3] H. Chui, A. Rangarajan, A new point matching algorithm for non-rigid registration, Comput. Vis. Image Underst. 89 (2-3) (2003) 114–141.
[4] Y. Zheng, E. Daniel, A.A. Hunter III, R. Xiao, J. Gao, H. Li, M.G. Maguire, D.H. Brainard, J.C. Gee, Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix, Med. Image Anal. 18 (2014) 903–913.
[5] X. Yang, J. Pei, J. Shi, Inverse consistent non-rigid image registration based on robust point set matching, Biomed. Eng. Online 13 (Suppl 2) (2014) S2.
[6] A. Myronenko, X. Song, Point set registration: coherent point drift, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 2262–2275.
[7] S. Sousa, W.G. Kropatsch, Graph-based point drift: graph centrality on the registration of point-sets, Pattern Recognit. 48 (2015) 368–379.
[8] Q. Sang, J. Zhang, Z. Yu, Robust non-rigid point registration based on feature-dependant finite mixture model, Pattern Recognit. Lett. 34 (2013) 1557–1565.
[9] Y. Ou, C. Davatzikos, DRAMMS: Deformable Registration Via Attribute Matching and Mutual-Saliency Weighting, IPMI 2009 LNCS, vol. 5636, 2009, pp. 50–62.

[10] http://www.dir-lab.com/
[11] http://www.creatis.insa-lyon.fr/rio/popi-model/

[10] S. Periaswamy, H. Farid, Medical image registration with partial data, Med. Image Anal. 10 (2006) 452–464.
[11] J. Ma, J.C.W. Chan, F. Canters, An operational supperresolution approach for multi-temporal and multi-angle remotely sensed imagery, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (2012) 110–124.
[12] B. Yeo, M. Sabuncu, R. Desikan, B. Fischl, P. Golland, Effects of registration regularization and atlas sharpness on segmentation accuracy, Med. Image Anal. 12 (2008) 603–615.
[13] R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, X. Pennec, Non-Rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy, MICCAI 2004 LNCS, vol. 3216, 2004, pp. 704–711.
[14] F.F. Berendsen, A.N.T.J. Kotte, A.A.C. de Leeuw, I.M. Jurgenliemk-Schulz, M.A. Viergever, J.P.W. Pluim, Registration of structurally dissimilar images in MRI-based brachytherapy, Phys. Med. Biol. 59 (2014) 4033–4045.
[15] P. Risholm, E. Samset, I. Talos, W. Wells, A Non-Rigid Registration Framework that Accommodates Resection and Retraction, IPMI 2009 LNCS, vol. 5636, 2009, pp. 447–458.
[16] L. Tang, G. Hamarneh, R. Abugharbieh, Reliability-Driven, Spatially-Adaptive Regularization for Deformable Registration, WBIR 2010 LNCS, vol. 6204, 2010, pp. 173–185.
[17] N. Chitphakdithai, K.P. Vives, J.S. Duncan, Registration of Brain Resection MRI with Intensity and Location Priors, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011, pp. 1520–1523.
[18] I.J.A. Simpson, J.A. Schnabel, A.R. Groves, J.L.R. Andersson, M.W. Woolrich, Probabilistic inference of regularisation in non-rigid registration, NeuroImage 59 (2012) 2438–2451.
[19] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, S. Joshi, Large deformation 3D image registration in image-guided radiation therapy, Phys. Med. Biol. 50 (24) (2005) 5869–5892.
[20] S. Gao, L. Zhang, H. Wang, R. de Crevoisier, D.D. Kuban, R. Mohan, L. Dong, A deformable image registration method to handle distended rectums in prostate cancer radiotherapy, Med. Phys. 33 (9) (2006) 3304–3312.
[21] M. Bach Cuadra, M. De Craene, V. Duay, B. Macq, C. Pollo, J.-Ph. Thiran, Dense deformation field estimation for atlas-based segmentation of pathological MR brain images, Comput. Methods Programs Biomed. 84 (2006) 66–75.
[22] E.I. Zacharaki, C.S. Hogea, D. Shen, G. Biros, C. Davatzikos, Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth, NeuroImage 46 (2009) 762–774.
[23] A. Mang, S. Becker, A. Toma, T. Buzug, Coupling tumor growth with brain deformation: a constrained parametric non-rigid registration problem, in: Medical Imaging 2010: Image Processing, Proc. SPIE 7623 (2010), 76230C-1-12.
[24] M. Sdika, D. Pelletier, Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping, Hum. Brain Mapp. 30 (4) (2009) 1060–1067.
[25] M. Brett, A.P. Leff, C. Rorden, J. Ashburner, Spatial normalization of brain images with focal lesion using cost function masking, NeuroImage 14 (2) (2001) 486–500.
[26] S.M. Andersen, S.Z. Rapcsak, P.M. Beeson, Cost function masking during normalization of brains with focal lesions: still a necessity, NeuroImage 53 (1) (2010) 78–84.
[27] P. Ripolles, J. Marco-Pallares, R. de Diego-Balaguer, J. Miro, M. Falip, M. Juncadella, F. Rubio, A. Rodriguez-Fornells, Analysis of automated methods for spatial normalization of lesioned brains, NeuroImage 60 (2) (2012) 1296–1306.
[28] V. Katkovnik, A. Foi, K. Egiazarian, J. Astola, From local kernel to nonlocal multiple-model image denoising, Int. J. Comput.