

RESEARCH

Open Access



# Non-rigid MR-TRUS image registration for image-guided prostate biopsy using correlation ratio-based mutual information

Lun Gong<sup>1,2</sup>, Haifeng Wang<sup>3</sup>, Chengtao Peng<sup>4</sup>, Yakang Dai<sup>1</sup>, Min Ding<sup>5</sup>, Yinghao Sun<sup>3</sup>, Xiaodong Yang<sup>1</sup> and Jian Zheng<sup>1\*</sup>

\*Correspondence:

zhengj@sibet.ac.cn

<sup>1</sup> Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China  
Full list of author information is available at the end of the article

## Abstract

**Background:** To improve the accuracy of ultrasound-guided biopsy of the prostate, the non-rigid registration of magnetic resonance (MR) images onto transrectal ultrasound (TRUS) images has gained increasing attention. Mutual information (MI) is a widely used similarity criterion in MR-TRUS image registration. However, the use of MI has been challenged because of intensity distortion, noise and down-sampling. Hence, we need to improve the MI measure to get better registration effect.

**Methods:** We present a novel two-dimensional non-rigid MR-TRUS registration algorithm that uses correlation ratio-based mutual information (CRMI) as the similarity criterion. CRMI includes a functional mapping of intensity values on the basis of a generalized version of intensity class correspondence. We also analytically acquire the derivative of CRMI with respect to deformation parameters. Furthermore, we propose an improved stochastic gradient descent (ISGD) optimization method based on the Metropolis acceptance criteria to improve the global optimization ability and decrease the registration time.

**Results:** The performance of the proposed method is tested on synthetic images and 12 pairs of clinical prostate TRUS and MR images. By comparing label map registration frame (LMRF) and conditional mutual information (CMI), the proposed algorithm has a significant improvement in the average values of Hausdorff distance and target registration error. Although the average Dice Similarity coefficient is not significantly better than CMI, it still has a crucial increase over LMRF. The average computation time consumed by the proposed method is similar to LMRF, which is 16 times less than CMI.

**Conclusion:** With more accurate matching performance and lower sensitivity to noise and down-sampling, the proposed algorithm of minimizing CRMI by ISGD is more robust and has the potential for use in aligning TRUS and MR images for needle biopsy.

**Keywords:** Prostate, Needle biopsy, Non-rigid registration, CRMI, ISGD

## Background

Prostate cancer is among the most common diseases in males and exhibits a large and increasing morbidity in many countries. As the fifth leading cause of cancer death worldwide, prostate cancer affected approximately 1.1 million men worldwide in 2012 [1]. In China, for example, according to statistics, approximately 49,000 new cases of prostate

cancer were reported, ranking ninth in terms of cancer incidence in men in 2011 [2]. Transrectal ultrasound-guided (TRUS) prostate biopsy is the most common means of diagnosing prostate cancer when an individual exhibits high blood levels of prostate-specific antigen (PSA). Although TRUS has many advantages, including real-time detection, low cost, and easy operation, its poor image quality and lack of clear contrast between malignant and normal tissue lead to false-negative rates of up to 30% for systematic sextant biopsies [3]. In contrast, magnetic resonance (MR) is the most sensitive imaging modality for observing anatomical structures and locating prostate tumor. Therefore, the registration of pre-operative MR images onto inter-operative TRUS images is of important clinical significance for improving biopsy accuracy.

Pre-operative and inter-operative prostate shape may suffer deformation due to extrusion of the ultrasonic probe, inflation of the anorectal coil inside the rectum during the MR scanning, and alterations in patient position [4]. To compensate for these movements, many non-rigid registration algorithms have been proposed over the past 20 years [5–8]. Mutual information (MI) is a widely used similarity criterion in multi-modal image registration and has been independently proposed by Collignon [9] and Viola [10]. Moradi [11] created a label map registration frame (LMRF) that aligned TRUS and MR images by using 3D Slicer, which first used the iterative closest point (ICP) method to rigidly align the outlines of the two images and then combined MI and B-splines to elastically register binary label maps from the two images obtained by manual contouring. Although LMRF aligned contours with high accuracy and could be conveniently implemented in 3D Slicer, it used only label maps to correct the local deformation and ignored pixel intensity, resulting in a high target registration error (TRE) of  $3.6 \pm 1.7$  mm. Mitra [12] utilized directional quadrature filter pairs to convert TRUS and MR images into texture images and then used normalized mutual information (NMI) as a similarity criterion to register the texture images. The average TRE and mean Dice Similarity coefficient (DSC) were  $2.64 \pm 1.47$  mm and  $0.943 \pm 0.039$ , but the computation time exceeded 797 s due to the use of 4 quadrature convolutions and the L-BFGS optimization method. However, several recent studies have shown that MI-based registration can be improved in certain cases. One improvement is to calculate MI over a set of overlapping image blocks to include spatial information. Loeckx [13] proposed the conditional mutual information (CMI), which considered the spatial location in the reference image of each joint intensity pair as the priori condition and calculated conditional entropies between the intensities given the spatial distribution. CMI-based registration resulted in a significant improvement in theoretical, phantom and clinical data compared with MI-based registration, but it was approximately 15 times slower than MI because it regarded the block label as the third channel of the joint histogram. Furthermore, MI assumes that all pixels in the overlapping area affect the calculations equally, but it is clear that different pixels contribute differently to the computation of MI [14]. Another kind of method assigns different weights to pixels using feature detection operators, e.g., the saliency measure [14] and the Harris corner detector [15]. But this method is often unable to extract effective features from TRUS images due to the low signal to noise ratio (SNR).

In this work, we design a mutual information-based method termed correlation ratio-based mutual information (CRMI) that includes the functional dependence of intensity

values. MI only includes the correspondence of intensity classes to correct for the deformation of location but ignores the possible relationship between intensity values. We also deduce the analytical derivative of CRMI in detail. Furthermore, to improve the global searching ability and reduce the registration time, we introduce Metropolis acceptance criteria [16] to an stochastic gradient descent (SGD) optimizer, which increases both the random disturbance and the probability of escaping the local extremum.

**Methods**

Let  $V = \{\mathbf{x} = (x, y) | 0 \leq x < S_x, 0 \leq y < S_y\} \subseteq R^2$  denote the image domain. We use the MR image as a fixed image denoted by  $F(\mathbf{x})$  and the TRUS image as a moving image denoted by  $M(\mathbf{x})$ . To simplify the derivation, we suppose that the intensities of the moving and fixed images have been normalized between zero and the number of histogram bins. The transformation that aligns  $F$  and  $M$  is represented by  $T = (T_x, T_y)$ . The registration can be viewed as the problem of selecting the transformation that best minimizes a cost function

$$C = D(F(\mathbf{x}), M(\mathbf{T}(\mathbf{x}))) + w_R \cdot C_{smooth}(\mathbf{T}) \tag{1}$$

where  $D$  represents the similarity metric,  $C_{smooth}$  is the constraint of the grid that ensures its smoothness, as introduced by Rueckert [17], and  $w_R$  is the weight of the constraint used to balance the metric and the penalty term.  $C_{smooth}$  takes the following form in 2D:

$$C_{smooth}(\mathbf{T}) = \frac{1}{N} \sum_{\mathbf{x} \in V} \left[ \left( \frac{\partial^2 \mathbf{T}}{\partial x^2} \right)^2 + \left( \frac{\partial^2 \mathbf{T}}{\partial y^2} \right)^2 + 2 \times \left( \frac{\partial^2 \mathbf{T}}{\partial x \partial y} \right)^2 \right] \tag{2}$$

where  $\mathbf{x}$  stands for the samples used to calculate the cost function and  $N$  is the total number of samples.

We choose a free-form deformation that is parameterized by the location of cubic B-spline nodes to simulate the transformation of the image. Given an  $n_x \times n_y$  uniform grid of control points on the moving image with spacing  $\delta$  and  $\mathbf{u}$  as the location of the control points in the image plane. The deformation of a pixel at the coordinate  $(x, y)$  can be parameterized by  $\mathbf{u}$  as follows:

$$T(x, y; \mathbf{u}) = \sum_{l=0}^3 \sum_{z=0}^3 B_l(w) B_z(v) u_{p_x + l, p_y + z} \tag{3}$$

where  $p_x = \lfloor x/\delta \rfloor - 1$ ,  $p_y = \lfloor y/\delta \rfloor - 1$ ,  $w = x/\delta - \lfloor x/\delta \rfloor$ ,  $v = y/\delta - \lfloor y/\delta \rfloor$  and  $\lfloor \cdot \rfloor$  is the truncating operation.  $B$  represents the cubic B-spline functions listed in Eq. (4). The multi-resolution Gaussian pyramid mentioned in [18] is applied to improve the searching efficiency.

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

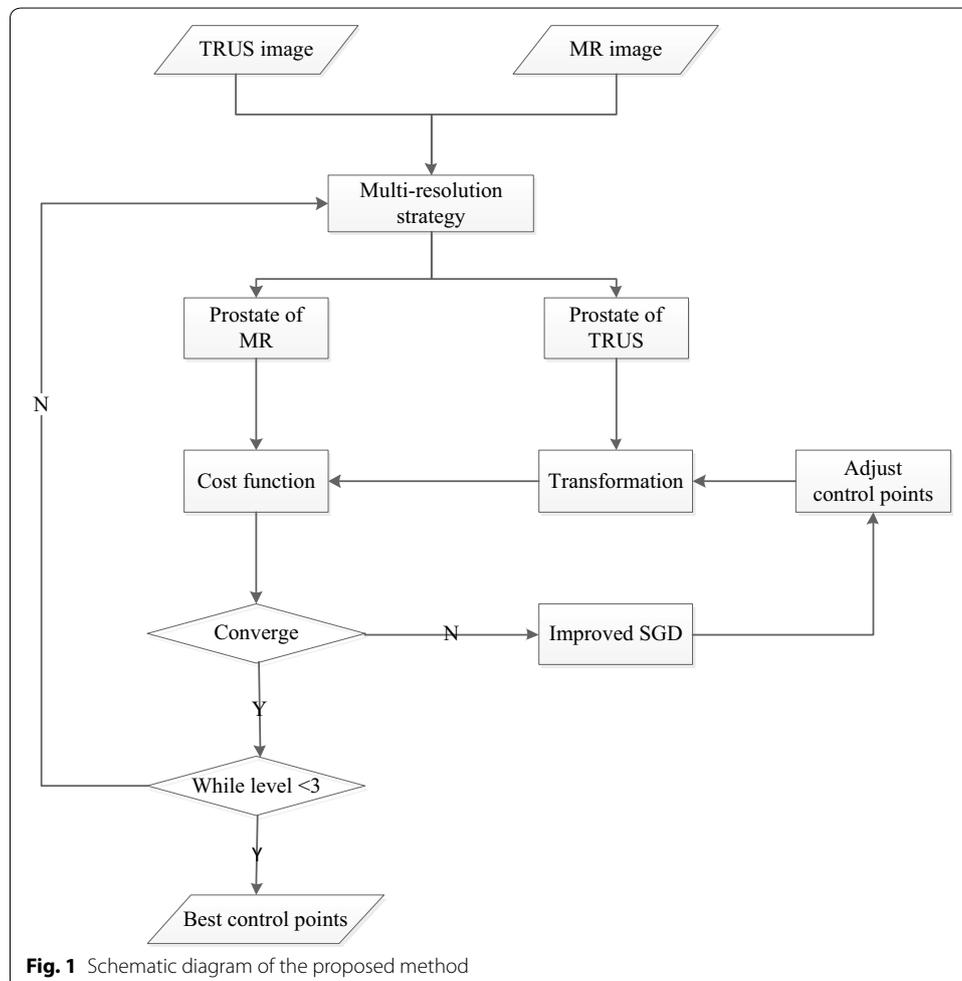
A schematic diagram of our method, which adopts a multi-resolution strategy, is shown in Fig. 1; each procedure will be illustrated in detail below.

**MI**

MI measures the dispersion of the joint density based on the assumption of a correspondence of intensity classes between two images. It finds a balance between the maximization of the marginal entropies and the minimization of the joint entropy. Let  $p(m, f; \mathbf{u})$  be the joint probability density function of  $M$  and  $F$ , and let  $p(m; \mathbf{u})$  and  $p(f)$  be the marginal probability density functions of  $M$  and  $F$ , respectively. MI can be expressed as a function of control points  $\mathbf{u}$  as follows:

$$MI(M, F; \mathbf{u}) = \sum_{m,f} p(m, f; \mathbf{u}) \log_2 \frac{p(m, f; \mathbf{u})}{p(f)p(m; \mathbf{u})}. \tag{5}$$

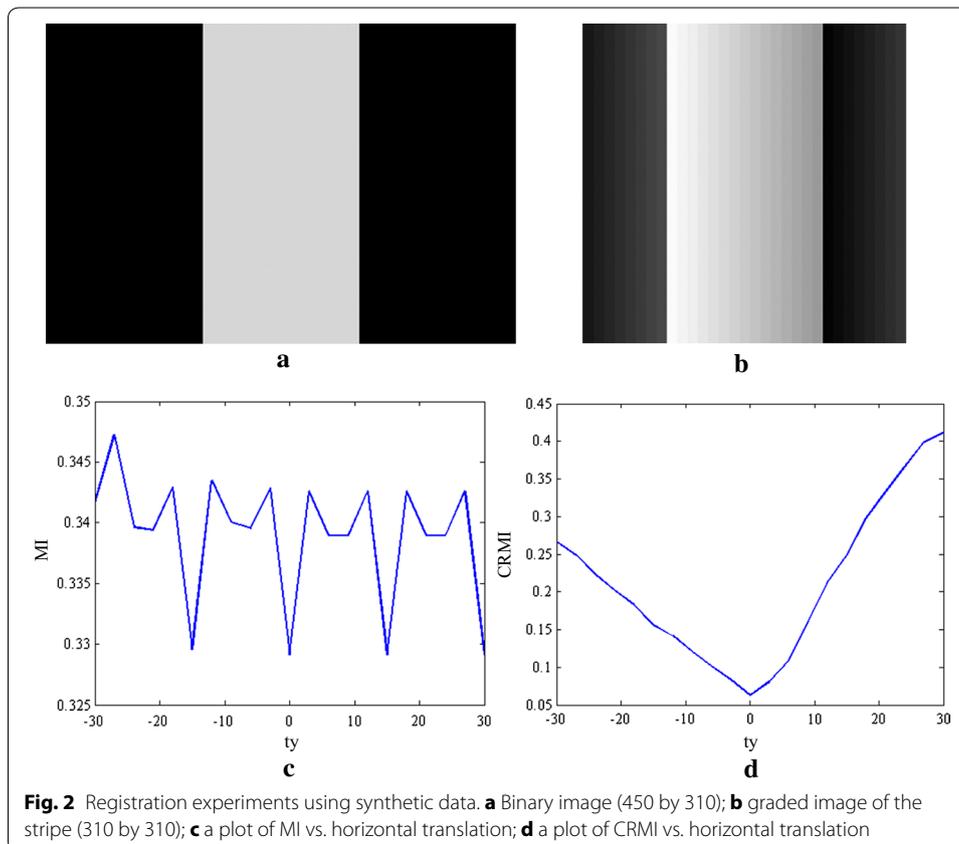
To derive the closed-form solution for the derivative of MI, we employ a second-order polynomial kernel designed by Xu [19] to estimate the probability density functions.



$$h_1(t) = \begin{cases} -1.8|t|^2 - 0.1|t| + 1, & 0 \leq |t| < 0.5 \\ 1.8|t|^2 - 3.7|t| + 1.9, & 0.5 \leq |t| \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

However, MI only corrects the deformation of location and ignores the functional mapping of intensity values. MI-based registration may be inappropriate for the alignment of images with intensity distortion because the intensity bias will disperse the joint density. Figure 2 illustrates an example of registering two synthetic images, where Fig. 2b contains intensity distortion. Figure 2c displays the registration function of MI with respect to a horizontal shift of Fig. 2b, where the origin stands for an accurate alignment of the two images. It is clear that mutual information does not have an absolute minimum when the stripes completely overlap, possibly because the optimal alignment does not correspond to the assumption of intensity class correspondence.

On the other hand, MI requires enough samples to compute the probability density functions. This limits the application of MI in multi-resolution cases because the error between the estimation of the joint density and the real density will increase at low resolution. Furthermore, MI establishes a purely qualitative statistical relationship between the intensity classes, which means that the joint density is sensitive to noise.



**CR**

For two random variables  $X$  and  $Y$ , the correlation ratio (CR) [20] is a quantitative description of the functional dependence between them, and the goal is to find the function  $\phi^*$  that best fits  $Y$  for all possible functions of  $X$ . Therefore, the problem is to find

$$\phi^* = \arg \min_{\phi} \|Y - \phi(X)\|_2. \quad (7)$$

Considering that the conditional expectation is the optimal approximator in the sense of the  $L^2$  norm, the quality of fitting can be measured by CR as follows:

$$CR(Y, X) = 1 - \frac{Var[Y - E(Y|X)]}{Var[Y]} \quad (8)$$

where  $Var[Y]$  is the variance of  $Y$ , and  $Var[Y - E(Y|X)]$  measures the part of  $Y$  that is functionally independent on  $X$  (see [20] for more details). CR varies between 0 (no functional dependence) and 1 (perfect functional relationship). In our method, the intensity values of the MR image are denoted as  $X$ , and the intensity values of the TRUS image are denoted as  $Y$ .

CR has been applied to register multi-modal images in [20–22]. Assuming a functional relationship between the intensities of float images and reference images is not an insurmountable obstacle for rigid or affine registration but is too restrictive to conform to the assumption for non-linear registration, especially when the image pairs have completely different presentations for the same anatomical structure. Therefore, CR is only applicable to rigid registration in most cases.

**CRMI**

Although mapping all intensities between a float image and a reference image using one function may be over-constrained, it is reasonable to assume that the intensities of the aligned pixel pairs can be mapped by the same function. It seems feasible to combine the functional dependence of the intensity values with MI by utilizing the deformation parameters of MI-based registration as prior information to CR-based registration. However, this approach is not advisable because it might degrade the performance into the original CR-based method if the alignment obtained by MI is not reasonably accurate. Furthermore, the registration time is excessive due to the need to separately minimize MI and CR. Therefore, we extend the location alignment to include the functional mapping between the intensity pairs by multiplying MI with CR. The scatter-plot dispersion of intensity values is measured by combining the joint density of the intensity classes with the mapping relationship of the intensity values. We not only correct the location deformation, but also decrease the influence of intensity bias on the joint density.

As described in the literature [23], multiplication is preferred over addition because the addition of these terms requires normalization; therefore, we define a new similarity metric as follows:

$$CRMI(M, F; \mathbf{u}) = (w_{mi} - MI(M, F; \mathbf{u})) \cdot (1 - CR(M, F; \mathbf{u})) \quad (9)$$

where  $w_{mi}$  is an empirical term, which ensures that CRMI is positive. In the new metric,  $w_{mi}$  is set to 1.5 and a linear kernel [24] that simplifies the derivation procedure is used to calculate CR as follow:

$$1 - CR(M, F; \mathbf{u}) = \frac{1}{\sigma^2} \left( \frac{1}{N} \sum_{\mathbf{x} \in V} M^2(\mathbf{T}(\mathbf{x}; \mathbf{u})) - \sum_f N_f \mu_f^2 \right) \quad (10)$$

$$\mu_f = \frac{\sum_{\mathbf{x} \in V} \lambda_{\mathbf{x},f} M(\mathbf{T}(\mathbf{x}; \mathbf{u}))}{N_f}, \quad N_f = \sum_{\mathbf{x} \in V} \lambda_{\mathbf{x},f}$$

where  $\sigma^2$  is the variance of the moving image, as for Eq. (11), and  $\lambda_{\mathbf{x},f}$  is the contribution of sample  $\mathbf{x}$  to bin  $f$  as calculated by Eq. (12).  $h_2(t)$  is the linear kernel as for Eq. (13), which ensures that each sample  $\mathbf{x}$  in the MR assigns the weights to its two closest bins  $f_1 - 1$  and  $f_1$  ( $f_1 = \lfloor F(\mathbf{x}) \rfloor + 1$ ).

$$\sigma^2 = Var[M] = \frac{1}{N} \sum_{\mathbf{x} \in V} M^2(\mathbf{T}(\mathbf{x}; \mathbf{u})) - \bar{m}^2, \quad \bar{m} = \frac{1}{N} \sum_{\mathbf{x} \in V} M(\mathbf{T}(\mathbf{x}; \mathbf{u})) \quad (11)$$

$$\lambda_{\mathbf{x},f} = \frac{1}{N} h_2(f - F(\mathbf{x})) \sum_{m=0} h_2(m - M(\mathbf{T}(\mathbf{x}; \mathbf{u}))) = \frac{1}{N} h_2(f - F(\mathbf{x})) \quad (12)$$

$$h_2(t) = \begin{cases} 1 - |t|, & \text{if } 0 \leq |t| < 1 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Figure 2d shows the translational results of CRMI with respect to the horizontal shift. It can be observed that the registration function of CRMI has the only one global minimum that agrees with the correct transformation parameters. This finding demonstrates that CRMI is more reliable for measuring the scatter-plot dispersion of intensity values.

### Gradient

To take advantage of the gradient-based optimization algorithm, we should obtain the derivative of the cost function. The gradient of the cost function with respect to the control points  $\mathbf{u}$  can be calculated as follows:

$$\frac{\partial C}{\partial \mathbf{u}} = (1 - CR) \frac{\partial (w_{mi} - MI)}{\partial \mathbf{u}} + (w_{mi} - MI) \frac{\partial (1 - CR)}{\partial \mathbf{u}} + w_{reg} \frac{\partial C_{smooth}(\mathbf{T})}{\partial \mathbf{u}}. \quad (14)$$

For an arbitrary control point  $u_{i,j}$ , we derive the derivative of the three terms independently.

The derivative of MI was obtained in [18] as follows:

$$\frac{\partial (w_{mi} - MI)}{\partial u_{i,j}} = - \sum_{m,f} \log_2 \left( \frac{p(m, f; \mathbf{u})}{p(m; \mathbf{u})} \right) \frac{\partial p(m, f; \mathbf{u})}{\partial u_{i,j}}. \quad (15)$$

According to the chain rule, the last term in Eq. (15) can be calculated as follows:

$$\frac{\partial p(m, f; \mathbf{u})}{\partial u_{i,j}} = \frac{1}{N} \sum_{\mathbf{x} \in V} \left[ h_1(f - F(\mathbf{x})) \frac{dh_1(\xi_{\mathbf{x}})}{d\xi_{\mathbf{x}}} \Big|_{\xi_{\mathbf{x}}=m-M(\mathbf{T}(\mathbf{x}; \mathbf{u}))} \times \left( \frac{-\partial M(\mathbf{y})}{\partial \mathbf{y}} \Big|_{\mathbf{y}=\mathbf{T}(\mathbf{x}; \mathbf{u})} \right)^T \times \frac{\partial \mathbf{T}(\mathbf{x}; \mathbf{u})}{\partial u_{i,j}} \right] \quad (16)$$

where  $dh_1(\xi)/d\xi_x$  is the first-order derivative of the kernel,  $\partial M(\mathbf{y})/\partial \mathbf{y}$  is the gradient of the moving image, and  $\partial \mathbf{T}(\mathbf{x}; \mathbf{u})/\partial u_{i,j}$  is the derivative of the deformation field with respect to  $u_{i,j}$ , the definition of which is given as follows:

$$\frac{\partial \mathbf{T}(\mathbf{x}; \mathbf{u})}{\partial u_{i,j}} = \begin{cases} B_r(w) B_q(v), & |\mathbf{x} - u_{i,j}| \leq 2\delta \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

where  $r = i - p_x$ ,  $q = j - p_y$  and  $p_x, p_y, w, v$  are defined in Eq. (3).

Using the chain rule, the derivative of the second term in CR can be expressed as follows:

$$\frac{\partial(1 - CR)}{\partial u_{i,j}} = \sum_{\mathbf{x} \in V} \left[ \frac{\partial(1 - CR)}{\partial \xi_{\mathbf{x}}} \Big|_{\xi_{\mathbf{x}}=M(\mathbf{T}(\mathbf{x}; \mathbf{u}))} \times \left( \frac{\partial M(\mathbf{y})}{\partial \mathbf{y}} \Big|_{\mathbf{y}=\mathbf{T}(\mathbf{x}; \mathbf{u})} \right)^T \times \frac{\partial \mathbf{T}(\mathbf{x}; \mathbf{u})}{\partial u_{i,j}} \right] \quad (18)$$

where  $\partial M(\mathbf{y})/\partial \mathbf{y}$  and  $\partial \mathbf{T}(\mathbf{x}; \mathbf{u})/\partial u_{i,j}$  are as defined in Eq. (16), and  $\partial(1 - CR)/\partial \xi_{\mathbf{x}}$  is derived as follows:

$$\begin{aligned} \frac{\partial(1 - CR)}{\partial \xi_{\mathbf{x}}} \Big|_{\xi_{\mathbf{x}}=M(\mathbf{T}(\mathbf{x}; \mathbf{u}))} &= \frac{-\partial \sigma^2 / \partial \xi_{\mathbf{x}}}{\sigma^4} \left( \frac{1}{N} \sum_{\mathbf{x} \in V} \xi_{\mathbf{x}}^2 - \sum_f N_f \mu_f^2 \right) \\ &+ \frac{1}{\sigma^2} \frac{\partial \left( \frac{1}{N} \sum_{\mathbf{x} \in V} \xi_{\mathbf{x}}^2 - \sum_f N_f \mu_f^2 \right)}{\partial \xi_{\mathbf{x}}} \end{aligned} \quad (19)$$

After calculating the derivatives and simplifying, we obtain

$$\begin{aligned} &\frac{\partial(1 - CR)}{\partial \xi_{\mathbf{x}}} \Big|_{\xi_{\mathbf{x}}=M(\mathbf{T}(\mathbf{x}; \mathbf{u}))} \\ &= \frac{2}{N\sigma^2} \left( \xi_{\mathbf{x}} - N(\lambda_{\mathbf{x}f_1-1} \mu_{f_1-1} + \lambda_{\mathbf{x}f_1} \mu_{f_1}) - \frac{1}{\sigma^2} (\xi_{\mathbf{x}} - \bar{m}) \left( \frac{1}{N} \sum_{\mathbf{x} \in V} \xi_{\mathbf{x}}^2 - \sum_f N_f \mu_f^2 \right) \right) \end{aligned} \quad (20)$$

where  $f_1$  and  $\bar{m}$  are as defined in Eq. (11). The details of the derivation of Eq. (20) from Eq. (19) are given in Appendix 1. Combining Eq. (18) with Eq. (20), we can analytically derive the closed-form derivative of CR with respect to  $u_{i,j}$ .

The derivative of the third formula  $C_{smooth}$  is given by Eq. (21), and the calculation of the right side of the equation is presented in Appendix 2.

$$\frac{\partial C_{smooth}(\mathbf{T})}{\partial u_{i,j}} = \frac{1}{N} \sum_{\mathbf{x} \in V} \left[ 2 \frac{\partial^2 \mathbf{T}}{\partial x^2} \frac{\partial(\partial^2 \mathbf{T} / \partial x^2)}{\partial u_{i,j}} + 2 \frac{\partial^2 \mathbf{T}}{\partial y^2} \frac{\partial(\partial^2 \mathbf{T} / \partial y^2)}{\partial u_{i,j}} + 4 \frac{\partial^2 \mathbf{T}}{\partial x \partial y} \frac{\partial(\partial^2 \mathbf{T} / \partial x \partial y)}{\partial u_{i,j}} \right]. \quad (21)$$

### Optimization

A typical gradient-based optimization method using the set of parameters  $\mathbf{u}$  can be expressed as follows:

$$\mathbf{u}_{k+1} = \mathbf{u}_k - a_k \mathbf{d}_k, \quad k = 1, 2, \dots \quad (22)$$

where  $\mathbf{d}_k$  is the search direction in iteration  $k$ , and  $a_k$  is the step size along the search direction. Klein [25], after comparing the performances of eight optimization methods, showed that SGD based on the Robbins-Monro algorithm was the best choice in most medical image registration applications. The main idea of SGD is to use an approximate gradient that is computed using a small, randomly picked subset of pixels to replace the real gradient calculated using all pixels, a method that is more efficient per iteration while not affecting the final accuracy [26]. In practice, SGD always exploits a choice such as that expressed in Eq. (23) to select the step size; thus, it is very difficult to determine the optimal values of the three parameters  $a$ ,  $A$  and  $\alpha$ .

$$a_k = a(k) = \frac{a}{(k + A)^\alpha}. \quad (23)$$

Klein [27] automatically calculated these parameters based on the distribution of the voxel displacements using the adaptive stochastic gradient descent method (ASGD); however, this method required computation of the Jacobin matrix of the deformation model with respect to the transformation parameters, which would be unacceptable if the number of transformation parameters was large. Qiao [28] proposed fast ASGD on the basis of ASGD to estimate the three parameters. This approach simplified the calculation of the bias and variance of the voxel displacements by using frequency statistics, a method that improved the efficiency at the cost of some precision. In addition, SGD requires a sufficient number of iterations to achieve convergence. The maximum number of iterations is also difficult to select. If the number is large, computation time will be increased; otherwise, convergence might not be achieved. Therefore, it is important to achieve the optimal efficiency and precision by setting up an appropriate stop condition.

In this section, we propose a new way to improve the performance of SGD, termed improved SGD (ISGD). The traditional SGD method can be viewed as a non-feedback system that does not depend on the exact value of the cost function. In our work, for each iteration, this value is fed back to the control loop to select a suitable step size, and its invariance within 25 consecutive iterations is used as the stop condition. Note that we utilize the discrete histogram to calculate the exact value of the cost function to save the computation time, but the registration accuracy is not affected because this value is only used for auxiliary judgment. To increase the robustness of ISGD with respect to the parameters used and to prevent falling into the local extremum, we introduce Metropolis acceptance criteria into the stochastic process as follows: If the current exact value is smaller than the previous value, the step size will be accepted; otherwise, the Boltzmann probability factor computed by Eq. (24) will be compared with a small random number distributed in the interval (0, 1). If the probability is larger than the random number, the current step size will be accepted. Otherwise, the current value is moving away from the optimal value, and the step size is decreased.

$$p = \exp\left(-\frac{C - V_b}{T_0}\right) \quad (24)$$

where  $C$  and  $V_b$  are the exact values of the cost function for the current and last iterations, respectively, and  $T_0$  is a constant.

Using the exact cost value to form a closed-loop system, ISGD can be supervised to find the optimal maximum number of iterations. To avoid immature convergence, we add a random disturbance on the basis of SGD by including Metropolis acceptance criteria. This procedure increases the probability that ISGD will jump out of the local extremum. ISGD will make it easier to choose the three parameters, because we only need to ensure that the initial step is large, and the procedure can always find a suitable step size after several attempts.

---

**Input:** the moving image  $M$ , the fixed image  $F$ , the initial deformation  $\mathbf{u}_0$  and the maximum iteration  $T$ .

- (1) Initialize the parameters  $a, A, \alpha, sk, T_0, err$ ;
- (2) Generate a 1-D array  $V_1$  to store the exact value of the cost function  $C$ ;
- (3) Compute  $C$  using all samples and its derivative  $\mathbf{g}_u$  using random samples based on  $\mathbf{u}_0$ ;
- (4)  $V_b = C, \mathbf{u}_b = \mathbf{u}_0, V_b = C, \mathbf{u}_b = \mathbf{u}_0, \mathbf{g}_b = \mathbf{g}_u$ ;
- (5)  $a_0 = a/(0+A)^\alpha$ ;
- (6) **for**  $k = 0:T-1$  **do**
- (7)  $V[k] = V_b$ ;
- (8) **if** 25 consecutive values in array  $V_1$  range within  $err$  **do**
- (9) **break**;
- (10) **if**  $V_b > C$  **do**
- (11) Replace  $V_b$  and  $\mathbf{u}_b$  with  $C$  and  $\mathbf{u}_k$ , respectively;
- (12)  $\mathbf{u}_{k+1} = \mathbf{u}_b - a_k \cdot \mathbf{g}_b$ ;
- (13) Compute  $C$  using all samples and its derivative  $\mathbf{g}_u$  using random samples based on  $\mathbf{u}_{k+1}$ ;
- (14) **if**  $C > V_b$  **do**
- (15) **if**  $p = \exp(-(C - V_b)/T_0) < rand$  **do**
- (16)  $a_k = a_k * sk$ ; **continue**;
- (17)  $\mathbf{u}_b = \mathbf{u}_{k+1}, V_b = C, \mathbf{g}_b = \mathbf{g}_u$ ;
- (18)  $a_k = a/(k+1+A)^\alpha$ ;
- (19) **end for**

**Output:** the best deformation  $\mathbf{u}_b$  that minimizes the cost function.

---

**Algorithm 1.** The pseudo code for the new optimization algorithm

## Experiments and results

In this section, we validate the effective performance of the proposed method using synthetic data and clinical prostate data. The proposed method is implemented in C++ and run on a PC equipped with an Intel Core i7 3.60 GHz CPU and 8 GB RAM.

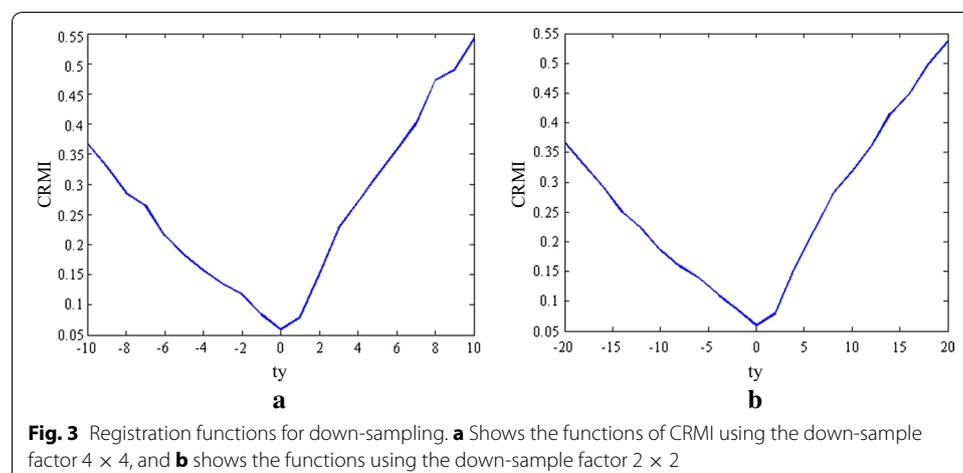
### Synthetic data

Considering the multi-resolution optimization strategy and the low SNR of TRUS images, a set of experiments using synthetic data are presented to demonstrate the robustness of CRMI to low sampling resolutions and its anti-noise performance. Figure 3a, b show the translational results of CRMI when the synthetic images are subsampled by factors of  $4 \times 4$  and  $2 \times 2$ . Figure 4a, c show the images presented in Fig. 2b after the addition of 10- and 20-dB Gaussian noise, respectively. Figure 4b, d are the corresponding translational results of CRMI. It is clear that all registration functions are relatively smooth and have a large capture range around the global minimum. Although the subsampling registration functions are not as smooth as the original, especially when subsampled by the factor of  $4 \times 4$ , the optimal alignment still corresponds to the global minimum.

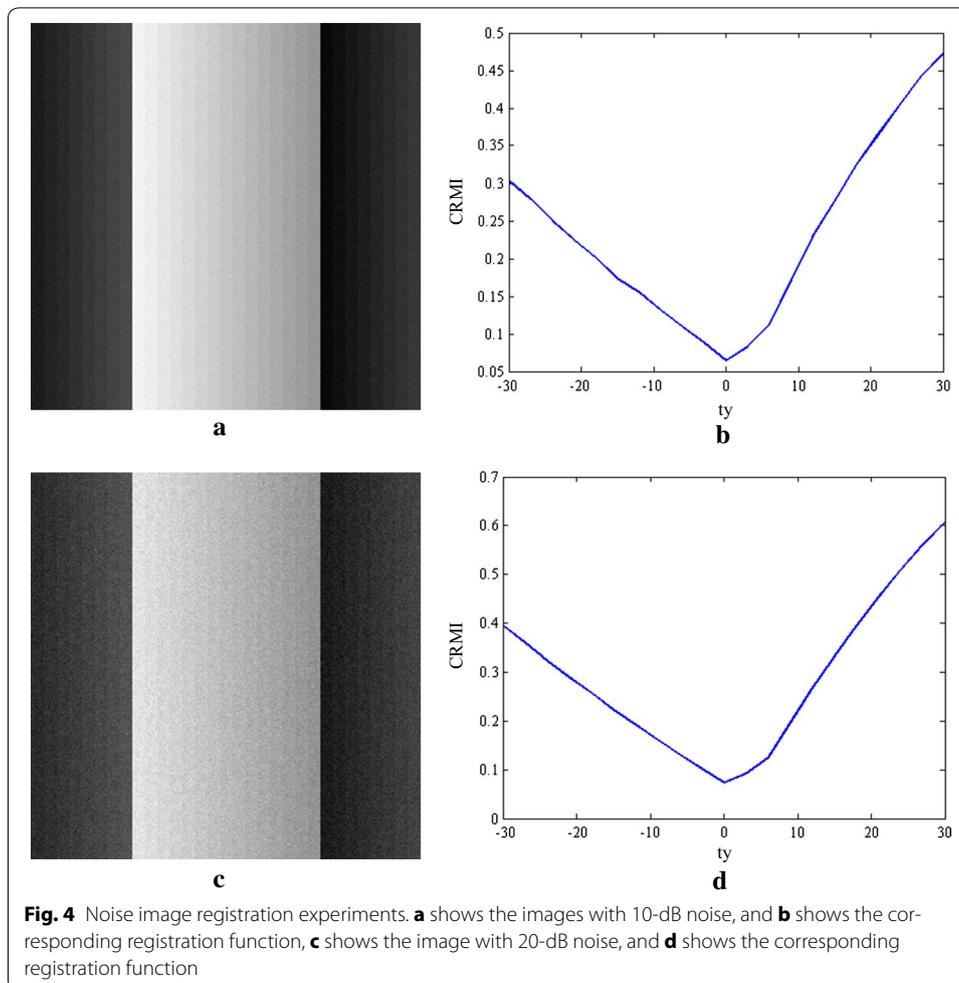
### Clinical prostate data

The proposed method is validated on prostate TRUS and MR images representing 12 patients, which are obtained from Shanghai Changhai Hospital. The size of the MR images is  $384 \times 384$  pixels, and the spatial resolution is  $0.4688 \times 0.4688 \text{ mm}^2$ ; the size of the TRUS images is  $383 \times 691$  pixels, and the spatial resolution is  $0.1667 \times 0.1667 \text{ mm}^2$ . To improve registration accuracy, the TRUS images are resampled to have the same pixel size as the MR images, and a zero-padding strategy is applied to the resampled TRUS images such that the images are the same size with the MR images. The prostate MR and TRUS images are segmented using the Chan-Vese model. For the prostate images of all 12 patients, the corresponding anatomical landmarks are selected by clinical expert.

We employ the common metrics [12] such as DSC, HD, TRE and registration time to evaluate the registration performance. DSC and HD measure the global registration accuracy, and TRE describes how well the local details are aligned. A high value of DSC



**Fig. 3** Registration functions for down-sampling. **a** Shows the functions of CRMI using the down-sample factor  $4 \times 4$ , and **b** shows the functions using the down-sample factor  $2 \times 2$

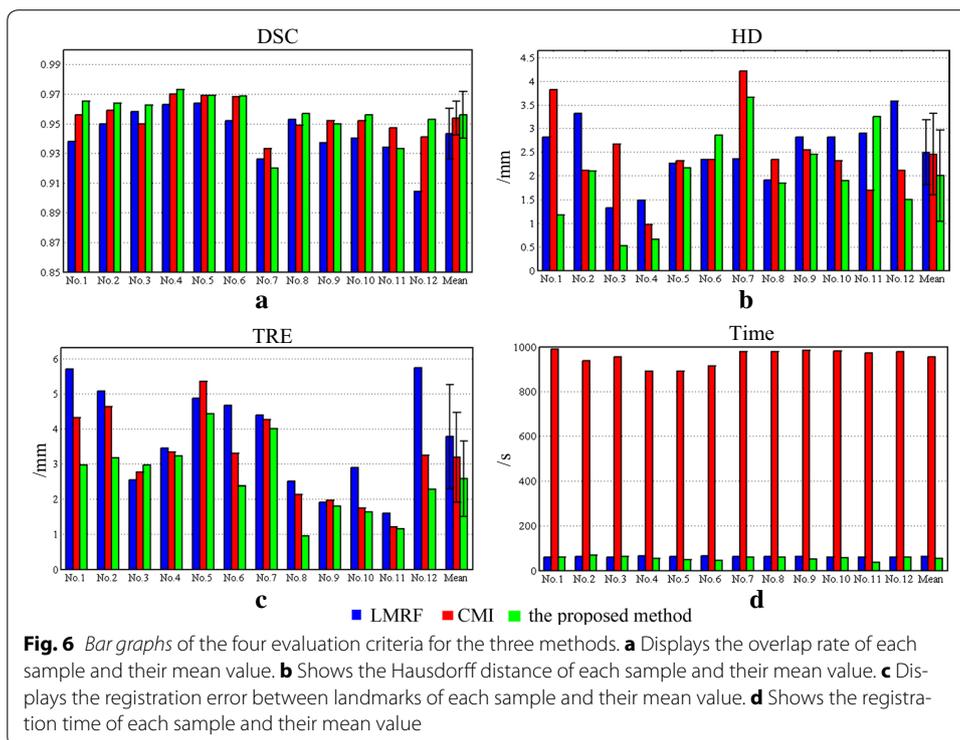
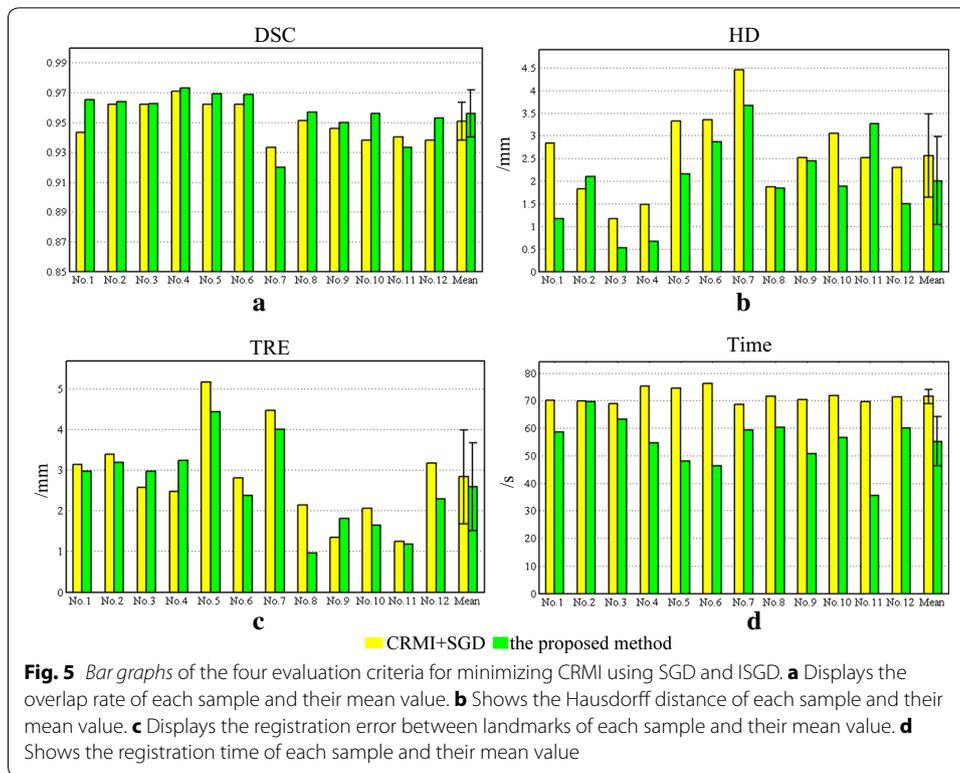


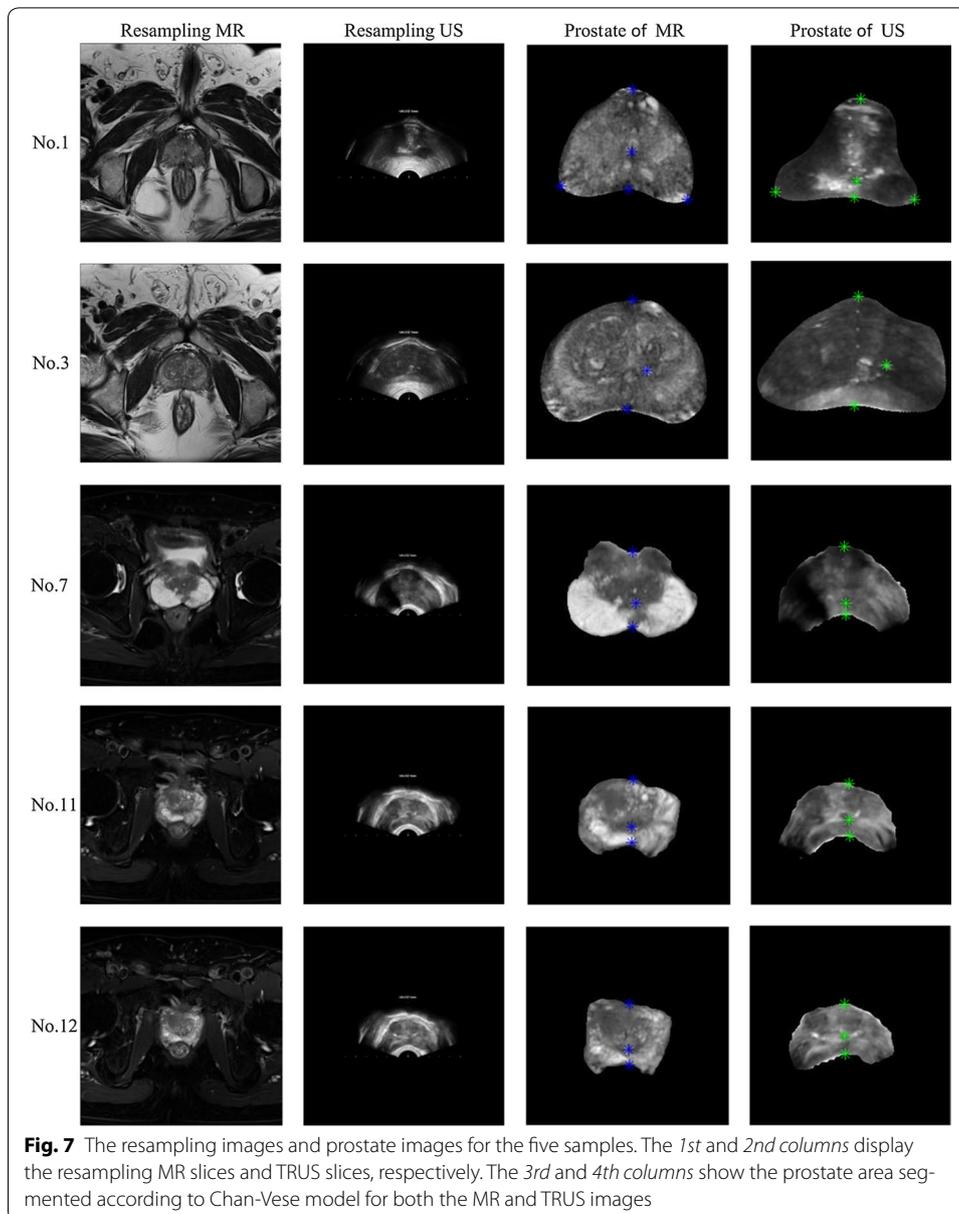
means that the prostate region on a moving image maps well to that on a fixed image. A low value of HD indicates good correspondence between the coordinates of two outlines. The lower the value of TRE is, the better the local registration accuracy has been achieved.

First, we have compared the performance of minimizing CRMI between ISGD and SGD. As shown in Fig. 5, the mean registration time consumed by the proposed method is significantly shorter than SGD. This result demonstrates the effectiveness of the improvement of ISGD. The average values of DSC, HD and TRE also become better, thus verifying that ISGD has a stronger global searching ability and is more likely to jump out of local minima than SGD.

Then we compare the proposed method against two other algorithms: LMRF [11] which is widely used in 3D Slicer and CMI [13] which exhibits higher registration accuracy over MI in many cases. All three algorithms use the same multi-resolution framework to improve search efficiency, and the number of bins is set to 64.

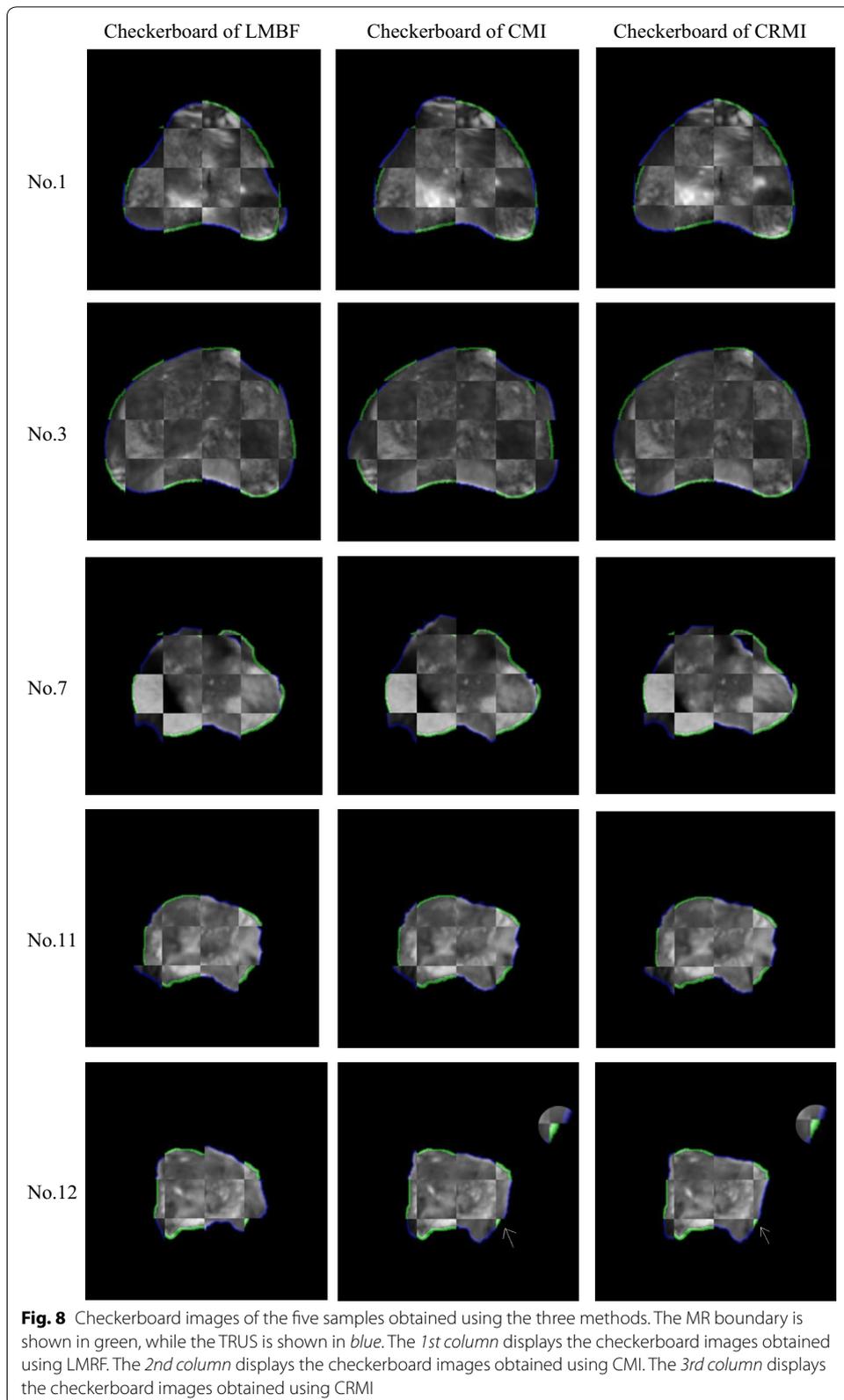
Figure 6 shows the evaluation criteria obtained by LMRF, CMI and the proposed method. Figure 7 presents the five groups (No. 1, No. 3, No. 7, No. 11 and No. 12) of adjusted images and the corresponding prostate images. The samples No. 1 and No. 12

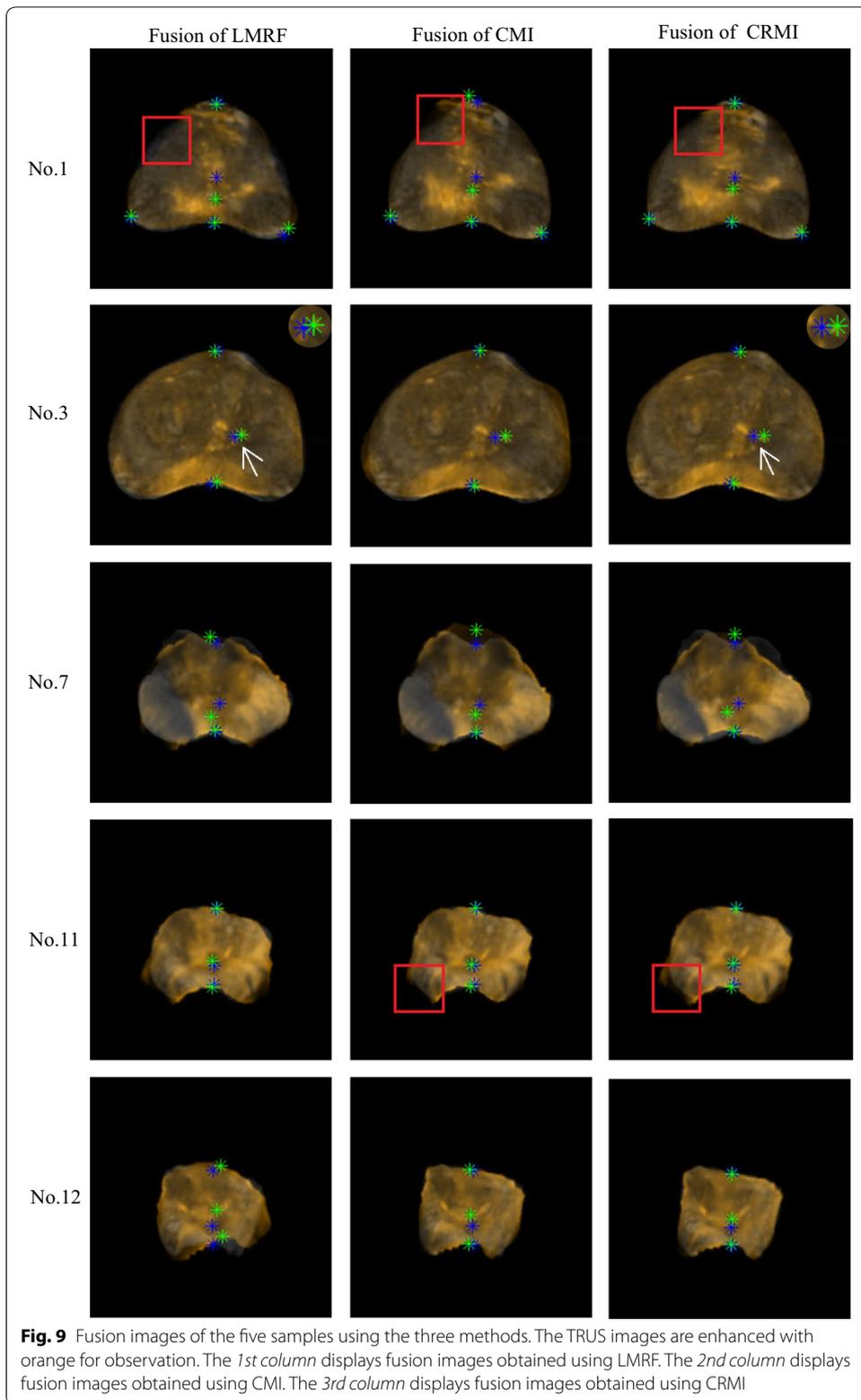




are the representations to show the excellent performance of the proposed method. The samples No. 3 and No. 11 show some special cases that the proposed method does not get the best values for some metrics. And the sample No. 7 is an extreme case that the proposed method fails to correct for the deformation. Figure 8 shows the registration performance of the boundary of the three methods as checkerboard images. Figure 9 presents fusion images that fuse the MR images to the registered TRUS images obtained using the three methods.

As shown in Fig. 6, the average TRE values are  $3.78 \pm 1.47$ ,  $3.19 \pm 1.28$ , and  $2.58 \pm 1.08$  mm for LMRF, CMI and CRMI, respectively, showing that the local homologous structures that are optimized using CRMI have the highest alignment accuracy. The average HD value of CRMI is also significantly reduced. Although the average DSC value





of CRMI is not significantly better than that of CMI, it is increased compared with that of LMRF. The improvement in these global measures validates that the new similarity criterion aligns the global deformation more reliably at low resolution. The registration time consumed by CRMI is approximately 16 times less than that consumed by CMI and is similar to LMRF, although CRMI requires the extra calculation of CR.

For most samples, the three metrics of DSC, HD and TRE of CRMI are the best. For sample No. 1, it is easy to observe that the correspondence between the two outlines aligned using CRMI is the optimal (see No. 1 in Fig. 8), and the distances between the landmarks are the least (see No. 1 in Fig. 9). For sample No. 12, it is clear that the registration accuracy of CRMI is higher than CMI, whereas LMRF fails to correct the displacements, thus further confirming that the new similarity criterion is better at evaluating the scatter-plot dispersion of the intensities in the presence of noise and intensity distortion.

However, some cases remain that should be considered. For sample No. 3, the global measures obtained using CRMI are the best, but the TRE value is not the least. As the partial enlarged drawing of No. 3 shown in Fig. 9, the distances between the landmarks obtained using CRMI are larger than LMRF. The reason might be that CRMI tends to search more accurate global movement parameters at low resolution because of the sample robustness; however, it falls into a local extremum at high resolution due to the inappropriate step factor  $sk$ . For samples No. 11, the checkerboard image obtained using CRMI is more approximate, and the lower left of the fusion image exhibits a smaller overlap rate than CMI, corresponding to the bar graphs shown for DSC and HD. It occurs because the parameter  $T_0$  decides whether ISGD receives the current solution or not is too large; therefore, the current result will be received even if the cost function changes a lot, eventually leading to premature convergence at low resolution. Unexpectedly, as the checkerboard images of sample No. 1 shown in Fig. 8, the checkerboard image obtained using CMI is smoother than LMRF, while the HD values are opposite. This occurs because the maximum distance between the outlines is not observable when the number of checkerboards is inappropriate.

The sample No. 7 is an extreme case that the values of DSC, HD and TRE obtained using CRMI are far worse than their average values. Considering that the other two algorithms also fail to align this group of images, the reason might be that the images are seriously affected by noise and intensity distortion, so that the intensities could not satisfy the assumption of intensity class correspondence or the functional mapping between aligned pixel pairs.

### Discussion and conclusion

In this study, we propose and verify a new similarity metric termed CRMI for 2D MR and TRUS prostate alignment. This method corrects the scatter-plot dispersion of the intensities caused by the deformation of location and intensity bias. Obtaining knowledge about the functional mapping between the intensities of float images and reference images is crucial when searching for the optimal alignment. We incorporate the functional dependence of the intensities quantified by CR into the intensity class correspondence held by MI. Experiments based on simulated data demonstrate that CRMI is more smooth and reliable for aligning images with intensity distortion. The new metric is also

robust to noise due to its calculation of the functional correlation in  $L^2$  space, which indicates that the intensity values of noise will be replaced by those of their neighboring samples because we map the pixel pairs with the smallest distance in intensity space. Furthermore, the computational error between the estimation of the joint density and the real density caused by insufficient sampling will be corrected during the searching of intensity mapping, which ensures that CRMI is not sensitive to down-sampling and is more suitable for multi-resolution optimization strategy.

Another innovation presented here is the improvement of the global searching ability of SGD with less computation time. We use the exact value of the cost function within a closed-loop system to select a suitable step size for each iteration as well as the stop criteria. Therefore, the improved SGD optimizer can adaptively determine the optimal number of iterations for different data, which will avoid hemstitching phenomenon due to the use of redundant iterations. An important parameter to determine the best step is the step factor  $sk$ . At low resolution, this value is set to 0.7 to save time. At medium and high resolution, this value is set to 0.17 to capture a more accurate step size. While searching for the minimum of the cost function, we not only accept smaller values but also consider relatively large values at a certain probability level by using Metropolis acceptance criteria. Therefore, the improved SGD optimizer is more likely to jump out of the local extremum because of the additional random disturbance. The constant  $T_0$  directly affects the global searching ability of ISGD. Considering that the exact cost function value changes within a value of 0.001, this constant is set to 0.0006, 0.00054 and 0.0008 for low, medium and high resolution, as experimentally determined. Comparative experiments using clinical prostate data verify that ISGD yields more accurate optimization performance and requires less registration time.

After comparing two non-rigid registration algorithms of LMRF and CMI, the proposed method of minimizing CRMI using ISGD provides the largest average overlap rate and the least mean HD value. The alignment accuracy of the landmarks also has a significant improvement. The average registration time is similar to that using LMRF, which is approximately 16 times less than that for CMI. These improvements confirm that the proposed method is more suitable for aligning 2D MR and TRUS images. The main limitation is that a few special cases may not fulfil the assumption of the functional mapping between aligned pixel pairs. The proposed algorithm may be extended to 3D MR and TRUS registration which can provide a wide range of information about the prostate, and the parallel computing capability of GPU can be exploited to realize real-time registration for TRUS-guided prostate biopsy.

#### Abbreviations

TRUS: transrectal ultrasound; PSA: prostate specific antigen; MR: magnetic resonance; MI: mutual information; LMRF: label map registration frame; ICP: iterative closest point; TRE: target registration error; NMI: normalized mutual information; DSC: dice similarity coefficient; L-BFGS: Limited-Broyden-Fletcher-Goldfarb-Shanno; CMI: conditional mutual information; SNR: signal to noise ratio; CRMI: correlation ratio-based mutual information; SGD: stochastic gradient descent; CR: correlation ratio; ASGD: adaptive stochastic gradient descent; ISGD: improved stochastic gradient descent; HD: Hausdorff distance.

#### Authors' contributions

Study concept and design LG, CP and MD; drafting of the manuscript LG; critical revision of the manuscript for important intellectual content LG and JZ; obtained funding JZ and HW; administrative, technical, and material support LG, JZ, HW and YS; selecting the anatomical landmarks HW; study supervision YD and XY. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China. <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China. <sup>3</sup> Department of Urology, Shanghai Changhai Hospital, Shanghai 200433, China. <sup>4</sup> Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230061, China. <sup>5</sup> School of Science, Nanjing University of Science and Technology, Nanjing 210094, China.

**Acknowledgements**

The authors would like to thank colleagues Dr. Cheng Zhang and Dr. Ming Li, Shanghai Changhai Hospital for providing the clinical data. The authors are also grateful to anonymous reviewers for their valuable comments and suggestions.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The datasets during the current study are available from the corresponding author on the reasonable request.

**Consent for publication**

The data for experiments has obtained the consent of the patients and Shanghai Changhai hospital, and has been allowed for publication.

**Ethics approval and consent to participate**

All data has been approved by the Medical Ethics Committee of Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, and has been allowed for carrying out experiments.

**Funding**

This work is supported in part by the National Key Research and Development Program of China (Nos. 2016YFC0104500, 2016YFC0104505), the Youth Innovation Promotion Association CAS (No. 2014281), National Natural Science Foundation of China (No. 61201117), the Science and Technology Program of Suzhou (No. ZXY2013001), the Shanghai Health and Family Planning Commission research projects (No. 201540812), the Shanghai Changhai Hospital New Medical Technology Breeding Program (No. NT201507).

**Appendix 1**

Exploiting the independence of  $\lambda_{x,f}$  and  $N_f$  of  $\xi_x$ , we first compute  $\partial\mu_f/\partial\xi_x$  as follows:

$$\frac{\partial \mu_f}{\partial \xi_x} = \partial \frac{\sum_{x \in V} \lambda_{x,f} \xi_x}{N_f} = \frac{\lambda_{x,f}}{N_f}.$$

According to Eq. (19), the key to calculating  $\partial(1 - CR)/\partial\xi_x$  is to calculate  $\partial\sigma^2/\partial\xi_x$  and  $\partial\left(\frac{1}{N} \sum_{x \in V} \xi_x^2 - \sum_f N_f \mu_f^2\right)/\partial \xi_x$ . Therefore, these are derived independently.

$$\frac{\partial \sigma^2}{\partial \xi_x} = \partial \frac{\frac{1}{N} \sum_{x \in V} \xi_x^2 - \bar{m}^2}{\partial \xi_x} = \frac{2}{N} \xi_x - \frac{2}{N} \bar{m}$$

$$\begin{aligned} \partial \frac{\left(\frac{1}{N} \sum_{x \in V} \xi_x^2 - \sum_f N_f \mu_f^2\right)}{\partial \xi_x} &= \frac{2}{N} \xi_x - \sum_f N_f \cdot 2 \mu_f \frac{\partial \mu_f}{\partial \xi_x} \\ &= \frac{2}{N} \xi_x - \sum_f N_f \cdot 2 \mu_f \frac{\lambda_{x,f}}{N_f} = \frac{2}{N} \xi_x - 2 \sum_f \lambda_{x,f} \mu_f \\ &= \frac{2}{N} \xi_x - 2(\lambda_{x,f_1-1} \mu_{f_1-1} + \lambda_{x,f_1} \mu_{f_1}) \end{aligned}$$

Therefore,  $\partial(1 - CR)/\partial\xi_x$  can be calculated as follows:

$$\begin{aligned} \frac{\partial(1 - CR)}{\partial \xi_x} \Big|_{\xi_x = I_m(\mathbf{T}(\mathbf{x}; \mathbf{u}))} &= \frac{-2(\xi_x - \bar{m})}{N\sigma^4} \left( \frac{1}{N} \sum_{x \in V} \xi_x^2 - \sum_f N_f \mu_f^2 \right) + \frac{1}{\sigma^2} \left( \frac{2\xi_x}{N} - 2(\lambda_{x,f_1-1} \mu_{f_1-1} + \lambda_{x,f_1} \mu_{f_1}) \right) \\ &= \frac{2}{N\sigma^2} \left( \xi_x - N(\lambda_{x,f_1-1} \mu_{f_1-1} + \lambda_{x,f_1} \mu_{f_1}) - \frac{1}{\sigma^2} (\xi_x - \bar{m}) \left( \frac{1}{N} \sum_{x \in V} \xi_x^2 - \sum_f N_f \mu_f^2 \right) \right). \end{aligned}$$

**Appendix 2**

Considering the independence of  $B_z(v)$  and  $\mathbf{u}$  of  $x$ , we calculate the first-order derivative of the deformation field with respect to  $x$  using the chain rule as follows:

$$\begin{aligned} \frac{\partial \mathbf{T}}{\partial x} &= \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial B_l(w)}{\partial w} \frac{\partial w}{\partial x} B_z(v) u_{p_x+l, p_y+z} \\ &= \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial B_l(w)}{\partial w} \frac{\partial}{\partial x} \left( \frac{x}{\delta} - \lfloor x/\delta \rfloor \right) B_z(v) u_{p_x+l, p_y+z} \\ &= \frac{1}{\delta} \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial B_l(w)}{\partial w} B_z(v) u_{p_x+l, p_y+z}. \end{aligned}$$

The derivative with respect to  $y$  can be computed in the same way:

$$\frac{\partial \mathbf{T}}{\partial y} = \sum_{l=0}^3 \sum_{z=0}^3 B_l(w) \frac{\partial B_z(v)}{\partial v} \frac{\partial v}{\partial y} u_{p_x+l, p_y+z} = \frac{1}{\delta} \sum_{l=0}^3 \sum_{z=0}^3 B_l(w) \frac{\partial B_z(v)}{\partial v} u_{p_x+l, p_y+z}.$$

Then, we can gain the second-order derivative

$$\begin{aligned} \frac{\partial^2 \mathbf{T}}{\partial x^2} &= \frac{1}{\delta} \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial^2 B_l(w)}{\partial w^2} \frac{\partial w}{\partial x} B_z(v) u_{p_x+l, p_y+z} \\ &= \frac{1}{\delta^2} \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial^2 B_l(w)}{\partial w^2} B_z(v) u_{p_x+l, p_y+z} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{T}}{\partial y^2} &= \frac{1}{\delta} \sum_{l=0}^3 \sum_{z=0}^3 B_l(w) \frac{\partial^2 B_z(v)}{\partial v^2} \frac{\partial v}{\partial y} u_{p_x+l, p_y+z} \\ &= \frac{1}{\delta^2} \sum_{l=0}^3 \sum_{z=0}^3 B_l(w) \frac{\partial^2 B_z(v)}{\partial v^2} u_{p_x+l, p_y+z} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{T}}{\partial x \partial y} &= \frac{1}{\delta} \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial B_l(w)}{\partial w} \frac{\partial B_z(v)}{\partial v} \frac{\partial v}{\partial y} u_{p_x+l, p_y+z} \\ &= \frac{1}{\delta^2} \sum_{l=0}^3 \sum_{z=0}^3 \frac{\partial B_l(w)}{\partial w} \frac{\partial B_z(v)}{\partial v} u_{p_x+l, p_y+z} \end{aligned}$$

An arbitrary control point  $u_{i,j}$  influences the pixels in a local neighbourhood, which is defined as  $\{\mathbf{x} \in \Omega \mid |\mathbf{x} - u_{i,j}| \leq 2\delta\}$ . Therefore, the derivative of the above second-order derivative with respect to  $u_{i,j}$  is

$$\frac{\partial(\partial^2 \mathbf{T} / \partial x^2)}{\partial u_{i,j}} = \begin{cases} \frac{1}{\delta^2} \frac{\partial^2 B_{i-p_x}(w)}{\partial w^2} B_{j-p_y}(v), & |\mathbf{x} - u_{i,j}| \leq 2\delta \\ 0, & otherwise \end{cases}$$

$$\frac{\partial(\partial^2 \mathbf{T} / \partial y^2)}{\partial u_{i,j}} = \begin{cases} \frac{1}{\delta^2} B_{i-p_x}(w) \frac{\partial^2 B_{j-p_y}(v)}{\partial v^2}, & |\mathbf{x} - u_{i,j}| \leq 2\delta \\ 0, & otherwise \end{cases}$$

$$\frac{\partial(\partial^2 \mathbf{T} / \partial x \partial y)}{\partial u_{i,j}} = \begin{cases} \frac{1}{\delta^2} \frac{\partial B_{i-p_x}(w)}{\partial w} \frac{\partial B_{j-p_y}(v)}{\partial v}, & |\mathbf{x} - \mathbf{u}_{i,j}| \leq 2\delta \\ 0, & \text{otherwise} \end{cases}.$$

Received: 10 September 2016 Accepted: 27 December 2016

Published online: 10 January 2017

## References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108.
2. Hao J, Zhao P, Chen WQ. Chinese cancer registration report 2015. Beijing: Press of Military Medical Sciences; 2015.
3. Norberg M, Egevad L, Holmberg L, Sparen P, Norlen B, Busch C. The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. *Urology*. 1997;50(4):562–6.
4. Mitra J, Kato Z, Marti R, Oliver A, Llado X, Sidibe D, Ghose S, Vilanova JC, Comet J, Meriaudeau F. A spline-based non-linear diffeomorphism for multimodal prostate registration. *Med Image Anal*. 2012;16(6):1259–79.
5. Xu S, Kruecker J, Turkbey B, Glossop N, Singh A, Choyke P, Pinto P, Wood B. Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies. *Comput Aided Surg*. 2008;13(5):255–64.
6. Hu Y, Ahmed HU, Taylor Z, Allen C, Emberton M, Hawkes D, Barratt D. MR to ultrasound registration for image-guided prostate interventions. *Med Image Anal*. 2012;16(3):687–703.
7. Hou M, Chen C, Tang D, Luo S, Yang F, Gu N. Magnetic microbubble-mediated ultrasound-MRI registration based on robust optical flow model. *Biomed Eng Online*. 2015;14(Suppl 1):S14.
8. Natarajan S, Marks LS, Margolis DJ, Huang J, Macairan ML, Lieu P, Fenster A. Clinical application of a 3D ultrasound-guided prostate biopsy system. *Urol Oncol*. 2011;29(3):334–42.
9. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. Automated multimodality image registration using information theory. *Inf Process Med Imaging*. 1995;3:263–74.
10. Viola P, Wells WM. Alignment by maximization of mutual information. 5th International Conference on Computer Vision 1995. p. 16–23.
11. Moradi M, Janoos F, Fedorov A, Risholm P, Kapur T. Two solutions for registration of ultrasound to MRI for image-guided prostate interventions. International Conference of the IEEE Engineering in Medicine and Biology Society 2012. p. 1129–32.
12. Mitra J, Marti R, Oliver A, Llado X, Ghose S, Vilanova JC, Meriaudeau F. Prostate multimodality image registration based on B-splines and quadrature local energy. *Int J Comput Assist Radiol Surg*. 2012;7(3):445–54.
13. Loeckx D, Slagmolen P, Maes F, Vandermeulen D, Suetens P. Nonrigid image registration using conditional mutual information. *IEEE Trans Med Imaging*. 2010;29(1):19–29.
14. Luan HX, Qi FH, Xue Z, Chen LY, Shen DG. Multimodality image registration by maximization of quantitative–qualitative measure of mutual information. *Pattern Recogn*. 2008;41:285–98.
15. Woo J, Stone M, Prince JL. Multimodal registration via mutual information incorporating geometric and spatial context. *IEEE Trans Image Process*. 2015;24(2):757–69.
16. Kirkpatrick S, Gelatt C, Vecchi M. Optimization by simulated annealing. *Science*. 1983;220(4598):671–80.
17. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes D. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18(8):712–21.
18. Thevenaz P, Unser M. Optimization of mutual information for multiresolution image registration. *IEEE Trans Med Imaging*. 2000;9(12):2083–99.
19. Xu R, Chen YW, Tang SY, Morikawa S, Kurumi Y. Parzen-window based normalized mutual information for medical image registration. *IEICE Trans Inf Systems*. 2008;91(1):132–44.
20. Alexis R, Grégoire M, Xavier P, Nicholas A. Multimodal image registration by maximization of the correlation ratio. Institut National De Recherche En Informatique Et En Automatique; 1998.
21. Wein W, Brunke S, Khamene A, Callstrom MR, Navab N. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Med Image Anal*. 2008;12(5):577–85.
22. Milko S, Melvaer EL, Samset E, Kadir T. Evaluation of bivariate correlation ratio similarity metric for rigid registration of US/MR images of the liver. *Int J Comput Assist Radiol Surg*. 2009;4(2):147–55.
23. Pluim JPW, Maintz JBA, Viergever MA. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans Image Process*. 2000;19(8):809–14.
24. Rivaz H, Chen S, Collins D. Automatic deformable MR-ultrasound registration for image-guided neurosurgery. *IEEE Trans Med Imaging*. 2015;34(2):365–80.
25. Klein S, Staring M, Pluim JPW. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans Image Process*. 2007;16(12):2879–90.
26. Klein S, Staring M, Pluim JPW. Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration. *Medical Imaging: Image Process*; 2005.
27. Klein S, Pluim JPW, Staring M, Viergever MA. Adaptive stochastic gradient descent optimisation for image registration. *Int J Comput Vision*. 2009;81(3):227–39.
28. Qiao Y, Lew BV, Lelieveldt BPF, Staring M. Fast automatic step size estimation for gradient descent optimization of image registration. *IEEE Trans Med Imaging*. 2016;35(2):391–403.