

Self-similarity Weighted Mutual Information: A New Nonrigid Image Registration Metric

Hassan Rivaz and D. Louis Collins

Montreal Neurological Institute, McGill University

Abstract. Extending mutual information (MI), which has been widely used as a similarity measure for rigid registration of multi-modal images, to deformable registration is an active field of research. We propose a self-similarity weighted graph-based implementation of α -mutual information (α -MI) for nonrigid image registration. The new Self Similarity α -MI (SeSaMI) metric takes local structures into account and is robust against signal non-stationarity and intensity distortions. We have used SeSaMI as the similarity measure in a regularized cost function with B-spline deformation field. Since the gradient of SeSaMI can be derived analytically, the cost function can be efficiently optimized using stochastic gradient descent. We show that SeSaMI produces a robust and smooth cost function and outperforms the state of the art statistical based similarity metrics in simulation and using data from image-guided neurosurgery.

1 Introduction

The joint intensity histogram of two images, of different or same modalities, is spread (i.e. the joint entropy is high) when they are not aligned, and is compact (i.e. the joint entropy is low) when the two images are aligned. Therefore, mutual information (MI) and its variants such as normalized MI (NMI) have been proposed and widely used for rigid registration of multi-modal images [1,2,3]. MI, being based on global intensity histograms, does not take into account local structures. Therefore, nonrigid registration, which has considerably more degrees of freedom and can distort local structures, is challenging with MI. It is also not robust against spatially varying bias fields. Exploiting the spatial information by conditioning MI calculation to location [4,5,6,7] has been shown to significantly improve nonrigid registration results.

In this work, we propose to incorporate image self-similarity into MI formulation. Self-similarity estimates the similarity of a point in one of the images to other points in the *same* image, and depends on local structures which are ignored by MI. Self-similarity was first proposed for object detection and image retrieval [8], and has since been used in image denoising [9] and registration [10].

Since self-similarity is calculated for pairs of points, it is natural to perceive it in a graph representation where image pixels are vertices and self-similarity is the weight of the edges. α -mutual information (α -MI) similarity

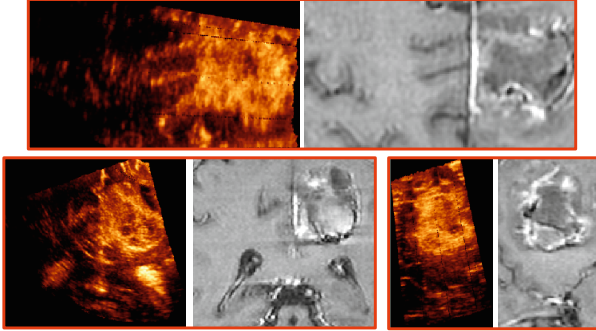


Fig. 1. Corresponding intra-operative US (hot colormap) and pre-operative MR (grayscale colormap) images of neurosurgery. A reconstructed US volume is sliced in the axial (top), coronal (bottom left) and sagittal (bottom right) directions. While local structures correspond, intensities are not related globally.

metric [11,12,13,14,15] is also graph based and has been recently shown to outperform MI in nonrigid registration applications. Therefore, we choose to incorporate self-similarity into this powerful registration framework. We apply the method to register pre-operative magnetic resonance (MR) images to intra-operative ultrasound (US) images in the context of image-guided neurosurgery (IGNS). Other works that nonlinearly register US to other modalities have used local correlation ratio [16] and MI of phase information [17]. Figure 1 shows an example of the registered US and MR images. The US images suffer from strong bias field due to signal attenuation, caused by scattering (from smaller than US wavelength inhomogeneities), specular reflection (from tissue boundaries) and absorption (as heat). In addition, US beam width varies significantly with depth, and therefore the same tissue looks different at different depths. Therefore, it is critical to exploit local structures.

In most image guided applications, one of the images is pre-operative. We therefore perform the self-similarity estimation only on this image offline, resulting in no increase in the on-line computational complexity. The pre-operative image is also usually of higher quality, making it a more attractive choice. We show that SeSaMI outperforms NMI and multi-feature α -MI in terms of producing a smooth dissimilarity function and registration accuracy.

2 Rotationally Invariant Self-similarity Estimation

We first estimate a rotationally invariant 2D histogram descriptor for all pixels; such a sample pixel is marked with an asterisk in Figure 2 (we show 2D images for clarity; the arguments are trivially extended to 3D images). A circular patch with radius r is centered on the pixel, shown in the left image. In this work, we set r to 5 in 2D and to 2 in 3D. The axes of the histogram are d , the normalized Euclidian distance of the pixel from the center point, and i , the pixel's normalized intensity.

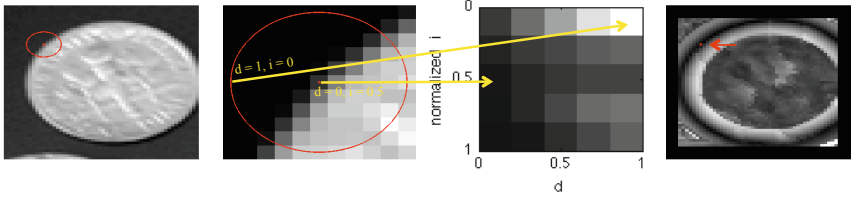


Fig. 2. Construction of the spin image and the resulting self-similarity distance. Left shows the image of a coin, with a circular patch of radius $r = 5$ pixels around a center point. Middle shows the zoomed patch, and the estimated 2D histogram descriptor (i.e. the spin image). Right shows the self-similarity distance to the center point (marked with an asterisk and pointed to by an arrow).

$d = 0$ and $d = 1$ in the histogram respectively correspond to the center pixel and to the pixels on the circle with radius $r = 5$ pixels. The intensity of pixels inside the patch is normalized to $[0, 1]$. Each pixel inside the patch contributes to the 2D histogram: the histogram is constructed using a Gaussian Parzen window, i.e. a pixel with distance d to the center and normalized intensity i contributes to all bins according to a Gaussian centered at (d, i) . The 2D histogram usually has a higher mass at higher d values because the number of pixels at distance d is proportional to d . Since d is the distance to the center (i.e. orientation is ignored), the 2D histogram descriptor is rotation invariant. It is also invariant to affine changes in the intensity because of the intensity normalization step. The histogram descriptor is similar to the spin image used in [18].

After calculating the 2D histogram descriptor for all points, we calculate the similarity between two points by the Earth Mover's Distance (EMD) [19]. The EMD metric avoids quantization and binning problems associated with histograms, and has been shown [19] to outperform other histogram comparison techniques. Figure 2 right shows the EMD distance to the point indicated by the asterisk. Note that small values of the EMD distance (darker pixels) represent more similar regions. It can be seen that the similarity metric is fully rotation invariant. We compute the self-similarity distance between each point and a window, of size The computational complexity of calculating the EMD distance is not an issue since it can be calculated offline on only the pre-operative image.

The histogram descriptor provides stability against small deformations of structures (due to the binning process), while subdividing the distance to the center (d in the histogram) encodes the spatial information. As a result, it is more robust than filter banks and differential invariants, which are also local descriptors [18]. Its disadvantage is its computational complexity. To reduce the running time and memory requirements, the EMD distance of a pixel with respect to pixels within its neighborhood is calculated. The EMD computation for a volume of size 100^3 pixels currently takes about 5 hours on a 3GHz processor.

The EMD distance provides a powerful metric to *condition* or *weight* MI estimation. *Conditioning* MI on EMD distance is motivated by the works which condition MI on spatial location and perform it on regions instead of the entire image [4,5,6,7]. Our preliminary results on conditioning MI on EMD have been

promising, especially since spatial distance can also be incorporated into the EMD distance. This is intuitive since self-similar pixels in one image are more likely to follow the same statistical relationship in the joint histogram. In this work, however, we focus on the second avenue, *weighting* MI with self-similarity. In the next section, we first briefly explain α -MI and then formulate SeSaMI.

3 Self-Similarity α -MI (SeSaMI)

Registration of a moving image I_m to a fixed image I_f can be formulated as

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} C, \quad C = S(I_f(\mathbf{x}), I_m(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}))) + \frac{\omega_R}{2} \|\nabla \boldsymbol{\mu}\|^2 \quad (1)$$

where S is a dissimilarity metric, ω_R is a regularization penalty weight, ∇ is the gradient operator and $\mathbf{T}_{\boldsymbol{\mu}}$ is the transformation modeled by $\boldsymbol{\mu}$. We choose a free-form transformation parameterized by the location of cubic B-spline nodes. $\boldsymbol{\mu}$ is therefore a vector of the location of all the nodes in all directions.

MI similarity metric is usually calculated on the intensities only, and therefore the joint histogram is 2D. α -MI is usually calculated on multiple features like intensities and their gradients. Adopting the notation of [11], let $\mathbf{z}(\mathbf{x}_i) = [z_1(\mathbf{x}_i) \cdots z_q(\mathbf{x}_i)]$ be a q -dimensional vector containing all the features at point \mathbf{x}_i . Similar to [11], we choose image intensity and gradients at two different scales as features, resulting in 5 total features. Let $\mathbf{z}^f(\mathbf{x}_i)$ and $\mathbf{z}^m(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i))$ be respectively the features of the fixed and moving image at \mathbf{x}_i and $\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i)$, and $\mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i))$ be their concatenation $[\mathbf{z}^f(\mathbf{x}_i), \mathbf{z}^m(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i))]$. Minimal spanning tree (MST) [12] and k -nearest neighbor (k NN) [11,15] are among different methods for estimating α -MI from multi-feature samples. With N samples, the complexities of constructing MST and k NN graphs are $O(N^2 \log N)$ and $O(N \log N)$ respectively [14]. Therefore, we choose k NN.

Let $\mathbf{z}^f(\mathbf{x}_{ip})$, $\mathbf{z}^m(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_{ip}))$ and $\mathbf{z}^{fm}(\mathbf{x}_{ip}, \mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_{ip}))$ be respectively the nearest neighbors of $\mathbf{z}^f(\mathbf{x}_i)$, $\mathbf{z}^m(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i))$ and $\mathbf{z}^{fm}(\mathbf{x}_i, \mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i))$. Note that these three nearest neighbors in general do not correspond to the same point. To prevent notation clutter, we show the dependencies on location \mathbf{x}_i or $\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x}_i)$ only through i after this point whenever clear. Let $\mathbf{d}_{ip}^{fm} = \|\mathbf{z}_i^{fm} - \mathbf{z}_{ip}^{fm}\|$, and $\mathbf{d}_{ip}^m = \|\mathbf{z}_i^m - \mathbf{z}_{ip}^m\|$ (\mathbf{d}_{ip}^{fm} and \mathbf{d}_{ip}^m will be used later when we differentiate the cost function) and set

$$\Gamma_i^f = \sum_{p=1}^k \|\mathbf{z}_i^f - \mathbf{z}_{ip}^f\|, \quad \Gamma_i^m(\boldsymbol{\mu}) = \sum_{p=1}^k \|\mathbf{d}_{ip}^m\|, \quad \Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^k \|\mathbf{d}_{ip}^{fm}\|. \quad (2)$$

A k NN estimator for α -MI= $-S$ (the dissimilarity function in Eq. 1) is

$$\widehat{\alpha\text{-MI}}(\boldsymbol{\mu}) = \frac{1}{\alpha - 1} \log \frac{1}{N^\alpha} \sum_{i=1}^N \left(\frac{\Gamma_i^{fm}(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}} \right)^{2\gamma} \quad (3)$$

where $\gamma = (1 - \alpha)q$ and $0 < \alpha < 1$; experimental results of rigid registration in [12] suggest that for MST graphs, α close to 1 gives better registration accuracy, while α close to 0.5 yields a wider capture range. We set $\alpha = 0.99$ in this work.

Weighting α -MI by Self-similarity. In an analogy to MI, small $\Gamma_i^{f^m}$ for majority of locations i means that data in the joint histogram is clustered and compact, and Γ_i^f and Γ_i^m are for normalization. Therefore, accurate estimates of $\Gamma_i^{f^m}$ are essential. Generally, most of the nearest neighbors in the joint feature space are also the most self-similar. However, due to spatially varying bias, small geometrical distortions, lack of enough number of features and misalignment, not all the nearest neighbors are self-similar. Therefore, to penalize points that are close but are not self-similar, we modify $\Gamma_i^{f^m}$ by

$$\Gamma_i^{f^m}(\boldsymbol{\mu}) = \sum_{p=1}^k w_{ip} \|\mathbf{d}_{ip}^{f^m}\|, w_{ip} = \text{EMD}(H(\mathbf{x}_i), H(\mathbf{x}_{ip})) \quad (4)$$

where $\text{EMD}(H(\mathbf{x}_i), H(\mathbf{x}_{ip}))$ is the EMD between the histogram descriptors.

We adopt an iterative stochastic gradient descent optimization method [20] for solving Eq. 1, which is fast and is less likely to get trapped in local minima. Therefore, $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + a_t \nabla_{\boldsymbol{\mu}} C$ where $\nabla_{\boldsymbol{\mu}} C$ is the gradient of C (from Eq. 1) wrt $\boldsymbol{\mu}$. The step size is a decaying function of the iteration number: $a_t = a/(A + t)^\tau$, with $a > 0, A \geq 0$ and $0 < \tau \leq 1$ user-defined constants [20]. From Eq. 1, we have $\nabla_{\boldsymbol{\mu}} C = -\nabla_{\boldsymbol{\mu}} \widehat{\alpha\text{-MI}} + \omega_R \Delta \boldsymbol{\mu}$ where $\Delta = \nabla \cdot \nabla$ is the Laplacian operator. At a specific $\boldsymbol{\mu}$ where the graph topology changes, Eq. 3 can be non-differentiable [12]. However, assuming the topology does not change for small changes in $\boldsymbol{\mu}$, the gradient of $\widehat{\alpha\text{-MI}}$ is calculated analytically in [11,15] using the chain rule; due to space limitations, we refer the reader to them for details. The chain rule finally results in computation of the $\nabla_{\boldsymbol{\mu}} \Gamma_i^{f^m}(\boldsymbol{\mu})$. From Eq. 4, we have

$$\frac{\partial}{\partial \mu_j} \Gamma_i^{f^m}(\boldsymbol{\mu}) = \sum_{p=1}^k \frac{w_{ip}}{\|\mathbf{d}_{ip}^{f^m}\|} \mathbf{d}_{ip}^{f^m T} \cdot \frac{\partial}{\partial \mu_j} \mathbf{d}_{ip}^{f^m} = \sum_{p=1}^k \frac{w_{ip}}{\|\mathbf{d}_{ip}^{f^m}\|} \mathbf{d}_{ip}^{m T} \cdot \frac{\partial}{\partial \mu_j} \mathbf{d}_{ip}^m \quad (5)$$

where T means transpose. w_{ip} is calculated for either I_f or I_m ; in the former case, its derivative wrt $\boldsymbol{\mu}$ is trivially zero, and in the latter case it is zero because the 2D histogram descriptors are invariant to small deformations [18]. Also, even for large global deformations, the histogram patches can be assumed locally rigid. The second equality is true because \mathbf{z}^{f^m} is the concatenation of \mathbf{z}^f and \mathbf{z}^m , and $\partial \mathbf{z}^f / \partial \boldsymbol{\mu} = 0$. Finally,

$$\mathbf{d}_{ip}^{m T} \cdot \frac{\partial}{\partial \mu_j} \mathbf{d}_{ip}^m = \mathbf{d}_{ip}^{m T} \cdot \left(\frac{\partial}{\partial \mathbf{T}(\mathbf{x}_i)} \mathbf{z}_i^m \frac{\partial}{\partial \mu_j} \mathbf{T}(\mathbf{x}_i) - \frac{\partial}{\partial \mathbf{T}(\mathbf{x}_{ip})} \mathbf{z}_{ip}^m \frac{\partial}{\partial \mu_j} \mathbf{T}(\mathbf{x}_{ip}) \right). \quad (6)$$

Note that partial derivative of \mathbf{z}^m wrt \mathbf{T} means calculating derivatives in I_m 's native space, i.e. wrt its own x, y or z coordinate. In our implementation, we pre-compute all the features of the I_f and I_m , and the derivatives of I_m 's features wrt x, y and z directions.

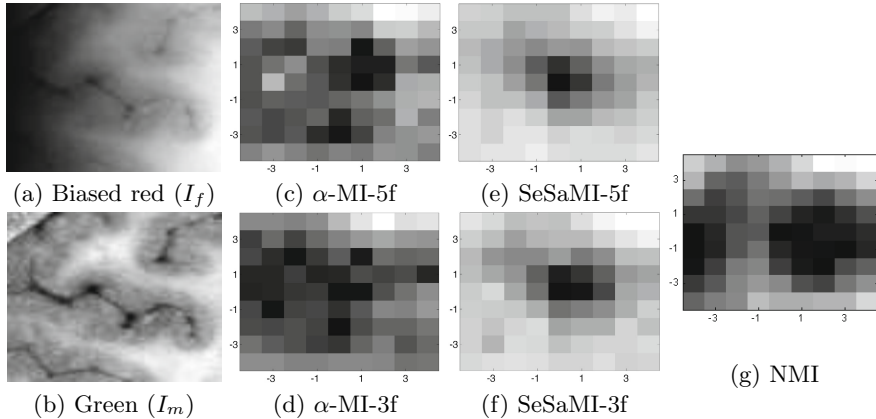


Fig. 3. Effect of the bias on the dissimilarity metrics in the human brain images. (a) The red channel with an additive bias. (b) The green channel. (c)-(f) The α MI and SeSaMI dissimilarity metrics calculated from $N = 400$ points randomly selected throughout I_f with 5 or 3 features (5f or 3f). (g) Our NMI implementation. The x and y axis represent the amount of rigid displacements of I_m in those directions (maximum of ± 4 pixels). Images are registered at 0 displacement. The self-similarity metrics in (e)-(f) are calculated using the biased I_f .

4 Experiments and Results

Visible Human Project. We test the new similarity metric on red and green channels of the visible human project, which are intrinsically registered. The data is publicly available at www.nlm.nih.gov/research/visible/visible_human.html. We set the red image as I_f and add bias to it to show the robustness of our self-similarity measure and SeSaMI (the self-similarity is calculated on the biased I_f image). Figure 3 shows the results; in (c)-(f) a total of $N = 400$ randomly selected points are used (the same random points are used in all the 4 cases) and the number of nearest neighbors k is 100. (g) is our NMI implementation with Parzen window histogram estimation [3]. SeSaMI successfully gives the global optimum at 0, and also produces a relatively smooth dissimilarity metric. In addition, it gives a global minimum even with 3 features (original intensity, smoothed intensity and gradient magnitude), instead of 5 features (calculated at 2 scales). Reduction of the number of features makes the algorithm run faster.

US and MR. We apply our registration algorithm to the clinical data of the IGNS obtained from 13 patients with gliomas in the Montreal Neurological Institute. The pre-operative MR images are gadolinium-enhanced T1 weighted and are acquired approximately 2 weeks before the surgery. The intra-operative US images are obtained using an HDI 5000 (Philips, Bothell, WA) with a P7-4 MHz

Table 1. MR/US registration mTRE (mm) for 3 nonlinear registration methods

Patient	No. of landmarks	Initial	NMI	α -MI	SeSaMI
P1	35	6.30	11.93	2.32	2.05
P2	40	9.38	19.36	3.14	2.76
P3	32	3.93	13.43	1.83	1.92
P4	31	2.62	18.82	2.62	2.71
P5	37	2.30	15.76	1.97	1.89
P6	19	3.04	9.01	2.28	2.05
P7	23	3.75	16.03	3.05	2.89
P8	21	5.09	7.83	2.44	2.93
P9	25	3.00	14.05	2.83	2.75
P10	25	1.52	18.65	1.44	1.28
P11	21	3.70	11.01	2.81	2.67
P12	23	5.15	17.46	3.37	2.82
P13	23	3.78	9.15	2.45	2.34

phased array transducer. The ultrasound probe is tracked with a Polaris camera (NDI, Waterloo, Canada), and 3D US volumes are reconstructed using the tracking information. The tracking information is also used to perform the initial rigid registration of MR to US; a sample of this initial registration is shown in Figure 1. A neurosurgeon and two experts have selected corresponding anatomical landmarks in US and MR images in sub-voxel accuracy, which are used to calculate mTRE for validation. Table 1 shows that multi-feature α -MI and SeSaMI significantly outperform NMI in nonlinear registration of MR to US in all the 13 cases. In 10 out of 13 cases, SeSaMI gives the most accurate results due to its robust self-similarity measure incorporated into the powerful multi-feature α -MI similarity metric.

5 Conclusions

We introduced SeSaMI, a similarity metric that incorporates rotation and bias invariant self-similarity measures into graph-based α -MI. SeSaMI exploits self-similarity in a k NN α -MI registration framework by penalizing clusters (i.e. the nearest neighbors) that are not self-similar. Therefore, it significantly reduces the number of incorrect local minima as shown in Figure 3. We have also, for the first time, shown that multi-feature α -MI and SeSaMI significantly increase the registration accuracy of MR to US registration in our on-going IGNS project. In the future we will investigate GPU implementations of SeSaMI to achieve our goal of near-real time intra-operative US-MRI registration.

Acknowledgements. H. Rivaz is supported by the NSERC PDF. The authors would also like to thank V. Fonov, D. Denigris and L. Mercier.

References

1. Wells, W., Viola, P., Atsumid, H., Nakajimae, S., Kikinis, R.: Multi-modal volume registration maximization of mutual information. *Med. Imag. Anal.* 1, 35–51 (1996)
2. Maes, F., et al.: Multimodality image registration by maximization of mutual information. *IEEE Trans. Medical Imag.* 16, 187–198 (1997)
3. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. *IEEE Trans. Medical Imag.* 22, 986–1004 (2003)
4. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Trans. Medical Imag.* 25, 626–639 (2006)
5. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. *IEEE Trans. Medical Imag.* 29, 19–29 (2010)
6. Klein, S., et al.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med. Phys.* 35, 1407–1417 (2008)
7. Zhuang, S., Arridge, D., Hawkes, D., Ourselin, S.: A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Trans. Medical Imag.* (in press)
8. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8 (2007)
9. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans. Med. Imag.* 27, 425–441 (2008)
10. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, J.M., Schnabel, J.A.: Non-local Shape Descriptor: A New Similarity Metric for Deformable Multi-modal Registration. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part II. LNCS*, vol. 6892, pp. 541–548. Springer, Heidelberg (2011)
11. Staring, M., Heide, U., Klein, S., Pluim, J.: Registration of cervical MRI using multifeature mutual information. *IEEE Trans. Medical Imag.* 28, 1412–1421 (2009)
12. Sabuncu, M., Ramadge, P.: Using spanning graphs for efficient image registration. *Information Processing Medical Imag. (IPMI)* 17, 788–797 (2008)
13. Kybic, J., Vnucko, I.: Approximate all nearest neighbor search for high dimensional entropy estimation for image registration. *Signal Processing* 92, 1302–1316 (2012)
14. Neemuchwala, H., Hero, A.: Entropic graphs for registration. In: Blum, R., Liu, Z. (eds.) *Multi-sensor Image Fusion and its Applications*. CRC Press (2005)
15. Oubel, E., Craene, M., Hero, A., Frangi, A.: Cardiac motion estimation by joint alignment of tagged MRI sequences. *Med. Imag. Anal.* 16, 339–350 (2012)
16. Wein, W., et al.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Imag. Analysis* 12, 577–585 (2008)
17. Zhang, W., Brady, M., Becher, H., Noble, A.: Spatio-temporal (2d+t) non-rigid registration of real-time 3D echocardiography and cardiovascular MR image sequences. *Physics Med. Biol.* 56, 1341–1360 (2011)
18. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. *IEEE Trans. Pattern Anal. Machine Int.* 27, 1265–1278 (2005)
19. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover’s distance as a metric for image retrieval. *IEEE Trans. Pattern Anal. Machine Int.* 40, 99–121 (2000)
20. Klein, S., Staring, M., Pluim, J.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Trans. Imag. Proc.* 16, 2879–2890 (2007)