Mixed Metric Random Forest for Dense Correspondence of Cone-Beam Computed Tomography Images

Yuru Pei[1(✉)], Yunai Yi[1], Gengyu Ma[2], Yuke Guo[3], Gui Chen[4], Tianmin Xu[4], and Hongbin Zha[1]

[1] Key Laboratory of Machine Perception (MOE), Department of Machine Intelligence, Peking University, Beijing, China
peiyuru@cis.pku.edu.cn
[2] uSens Inc., San Jose, USA
[3] Luoyang Institute of Science and Technology, Luoyang, China
[4] School of Stomatology, Peking University, Beijing, China

Abstract. Efficient dense correspondence and registration of CBCT images is an essential yet challenging task for inter-treatment evaluations of structural variations. In this paper, we propose an unsupervised mixed metric random forest (MMRF) for dense correspondence of CBCT images. The weak labeling resulted from a clustering forest is utilized to discriminate the badly-clustered supervoxels and related classes, which are favored in the following fine-tuning of the MMRF by penalized weighting in both classification and clustering entropy estimation. An iterative scheme is introduced for the forest reinforcement to minimize the inconsistent supervoxel labeling across CBCT images. In order to screen out the inconsistent matching pairs and to regularize the dense correspondence defined by the forest-based metric, we evaluate consistencies of candidate matching pairs by virtue of isometric constraints. The proposed correspondence method has been tested on 150 clinically captured CBCT images, and outperforms state-of-the-arts in terms of matching accuracy while being computationally efficient.

1 Introduction

Cone-beam computed tomography (CBCT) images have been widely used in clinical orthodontics for treatment evaluations and growth assessments. Efficient dense correspondence and image registration of CBCT images are highly desirable for inter-operative interventions and online attribute transfer, such as landmark location and label propagation. Volumetric image registration has been well-studied in medical image processing for decades. Nevertheless, while the advances made by the large influx of work are dramatic, the efficient online dense correspondence of CBCT images is still a challenging issue. Considering CBCT images with hundreds of millions of voxels, the non-rigid registration of full-sized CBCT images by commonly-used metrics, e.g. the mutual information (MI) [4,8,9] and normalized correlation, by a large-scale non-linear optimization

M. Descoteaux et al. (Eds.): MICCAI 2017, Part I, LNCS 10433, pp. 283–290, 2017.
DOI: 10.1007/978-3-319-66182-7_33

is far from real-time for online applications. Moreover, when given poor initial alignment, the optimization can be trapped into a local minimum, and make things even worse. An efficient and reliable engine for dense correspondence of CBCT images is highly demanded for online inter-operative applications.

The registration based on reduced samples has been used to accelerate the correspondence establishment [2,11]. Although importance sampling speeded up gradient estimation of similarity metrics, the registration based on the iterative optimization still consumed hundreds of seconds [2]. Moreover, the discrete samples were variable and cannot cover the whole volume image [11]. The supervised classification and regression random forests are known for efficient online-testing performance [3,6], and have been applied to correspondence establishment [7,12]. However, the regularization of forest-based correspondence in post-processing was still time-consuming [12]. Moreover, the prior labeling of volumetric medical images is extremely tedious and prone to inter- and intra-observer variations. Without prior labeling, the pseudo labels were defined by supervoxel decomposition of just one volumetric image [7]. It's relatively hard to generalize the classifier with limited training data.

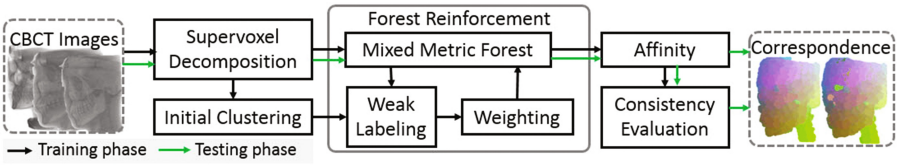

Fig. 1. Flowchart of our system.

In this paper, we propose a mixed metric random forest (MMRF) for correspondence establishment (see Fig. 1). The unsupervised clustering forest [10] is adopted to generate dense correspondence between supervoxels across CBCT images. We propose a novel iterative forest reinforcement method for an optimal forest-based metric to maximize the correspondence consistency in the CBCT image data set. The weak labeling defined by the clustering forest is used to discriminate the badly-clustered supervoxels and related classes. The penalized weights defined according to the confidence scores of weak labels are imposed on the mixed classification and clustering entropy estimation. In each iteration, the updated MMRF favors the previously badly-clustered instances, and in turn improve the forest-based metric for the correspondence establishment. In order to screen out the inconsistent matching pairs and to regularize the final dense correspondence, we evaluate consistencies for candidate matching pairs. The soft consistency label of a matching pair is defined based on supervoxel confidence scores. A conventional regression forest is employed for the consistency evaluation. In the online testing, the decomposed supervoxels of testing images are fed to the MMRF for dense correspondence. The consistency scores of matching pairs are further used to regularize the dense correspondence. The proposed system is totally unsupervised and without prior labeling. The MMRF is reinforced

based on the self-learning of data distribution and matching consistency across images. The dense correspondence by the MMRF is conducted by tree traversals with just a sequence of binary tests, and is computationally efficient.

2 Methods

2.1 Initial Supervoxel-wise Affinity and Weak Labeling

Once given a CBCT image data set $\mathcal{V} = \{V^i | i = 0, \ldots, M\}$ and accompanying decomposed supervoxel set $S = \{s_i | i = 1, \ldots, N\}$, an unsupervised clustering forest [10] is employed to estimate initial pairwise supervoxel affinities. By measuring hyperellipsoid volumes bounding the uncertainty of the data distribution, the criteria function $I_u = \sum_{k=l,r} \alpha_k tr(\sigma(\tilde{S}_k))$. I_u is defined by a trace of covariance matrix σ of supervoxel data sets in left and right children nodes, i.e. $\tilde{S}_l$ and $\tilde{S}_r$. The trace-based criterion is dominant in the information gain estimation [10], which can avoid the ubiquitous rank deficiency of the covariance matrix of high-dimensional data. The coefficient α_k is defined by the node cardinality, and $\alpha_k = |\tilde{S}_k| / \sum_{k=l,r} |\tilde{S}_k|$. Supervoxel s_i and s_j are assumed to be similar if they come to the same leaf node, and $\ell(s_i) = \ell(s_j)$. With respect to the k-th tree, the affinity $a_k(s_i, s_j) = [\ell(s_i) = \ell(s_j)]$, where $[\cdot]$ is an indicator function. The metric is defined as $a(s_i, s_j) = 1/n_T \sum a_k(s_i, s_j)$ for a forest with n_T trees.

Feature Channels. In our system, the supervoxel has three kinds of feature channels, i.e. intensity appearances, spatial contexts, and geodesic coordinates. As in [15], an intensity histogram b of voxels inside a supervoxel and an average histogram $\bar{b}$ of one-ring neighboring supervoxels are used as appearance feature $f_a = (b, \bar{b})$. The contextual features f_c is defined as appearance differences between supervoxel s and a randomly-sampled pattern P in a cube centered at s. $f_c = \left\{ \chi^2 \left(f_a(s), f_a(s + \delta_k) \right) | s + \delta_k \in P \right\}$. The geodesic coordinate f_g is defined as the shortest distance between supervoxel s and boundary background supervoxels s_g. Graph G is built upon each CBCT image with nodes at supervoxel centers and edges weighted by $\exp(-\varrho \| f_a(s_i) - f_a(s_j) \|)$. The geodesic coordinate $f_g(s) = \min d\left(s, s_g | G\right)$. In our system, the bin number of the intensity histogram is set at 20. Pattern P is predefined by sampling 50 voxels in a $150 \times 150 \times 150$ cube. The normalization parameter $\varrho = 1 / \max \| f_a(s_i) - f_a(s_j) \|$.

Weak Labeling. Given the forest-based metric, the supervoxel mapping function between images V^r and V^t is define as $\phi(s_i) = s_j$, where $s_j = \arg\max a(s_i, s_j)$, and $s_i \in V^r, s_j \in V^t$. The supervoxel index label set $Y^r = \{y_i | y_i \in \{1, \ldots, n_s\}\}$ of the reference image V^r with n_s supervoxels can be transferred to other images, and $y(s_i) = y(s_j)$ when $\phi(s_i) = s_j$. In order to avoid labeling bias due to random reference image selection, the image which produces the most consistent label transfer and maximizes $\sum_{m,n=1}^{M} \sum_{y(s_i^m) = y(s_j^n)} a(s_i^m, s_j^n)$, is selected as the reference image V^r.

2.2 Mixed Metric Random Forest

We propose an MMRF to iteratively reinforce the forest-based metric by favoring the previously badly-clustered supervoxels and related classes. The penalized weights are imposed on mixed classification and clustering entropy estimation according to the weak labeling. In order to discriminate the badly-clustered supervoxels, we define a confidence score of label y_i with respect to the k-th volumetric image $V^{(k)}$ as

$$\tau_i^{(k)} = 1 - \frac{1}{|Y_s|} \sum_{j=1}^{|Y_s|} \delta \left(\left\lfloor |Q^{(k)} - Q^r| \right\rfloor_{ij} - \eta \right), \tag{1}$$

where $Q^{(k)}$ and Q^r are n_s by $n_{s'}$ matrices of the normalized Euclidean distance between supervoxels of label Y and Y_s with respect to image $V^{(k)}$ and V^r. Y_s is a subset of Y and has $n_{s'}$ labels. δ is the Heaviside step function. η is a predefined inconsistency constant and set at 0.3 in all our experiments. The i-th row of matrix Q can be viewed as the spatial relationship of the supervoxel of label y_i with the rest supervoxels of label Y_s. When the spatial relationship of the supervoxel with label y_i in image $V^{(k)}$ agrees with that in the reference image, the label y_i is assumed to be confident with respect to image $V^{(k)}$. The confidence score of label y_i is defined by accumulating $\tau_i^{(k)}$ on all images, and $\gamma_i = \frac{1}{M} \sum_{k=1}^{M} \delta(\tau_i^{(k)} - 0.5)$. The weighted information gain with respect to the discrete probability distribution determined by the weak labeling is defined as

$$I_c = - \sum_{k=l,r} \alpha_k \sum_{i=1}^{n_s} \gamma_i \left(p(y_i|\tilde{S}_k) \ln p(y_i|\tilde{S}_k) \right). \tag{2}$$

Moreover, we discriminate the badly-clustered supervoxels and impose penalized weights on the uncertainty evaluation of data distribution in node splitting. The penalized weight ν of a supervoxel is defined as $\nu(s) = K \cdot \delta(0.5 - \tau_{y(s)}^{(k)}) + 1, s \in V^{(k)}$. K is a penalized constant and set at 5 in all our experiments. The clustering-related information gain I_u in Sect. 2.1 is rewritten as

$$I_u = - \sum_{k=l,r} \left(\ln \frac{\sum_i^{|\tilde{S}_k|} \nu_i^2 \left\| f(s_i) - \overline{\tilde{S}_k} \right\|_2^2}{\sum_{i,j}^{|\tilde{S}_k|} \nu_i \nu_j} \right), \tag{3}$$

where $\overline{\tilde{S}_k}$ is the weighted feature mean of supervoxel data set $\tilde{S}_k$. When training MMRF, we integrate penalized weighted information gain I_c of the discrete probability distribution determined by the weak labeling and I_u of the uncertainty evaluation of the data distribution. The criteria function $I = 0.5 \cdot (I_c/I_c^0 + I_u/I_u^0)$, which is normalized by I_c^0 and I_u^0 with respect to the information gains of the classification and the clustering in the root node splitting.

As shown in Fig. 1, given the updated MMRF, the weak labeling together with the penalized weights are updated accordingly. In the further iteration, the MMRF training will favor the previously badly-clustered instances to improve the forest-based metric for the correspondence establishment.

2.3 Soft Consistency Evaluation

When given a volumetric image pair (V^{t_1}, V^{t_2}), the dense matching set $C = \{(s_i, s_j)|s_j = \phi(s_i), s_i \in V^{t_1}, s_j \in V^{t_2}\}$ is obtained by the MMRF-based metric. However, there is no information on the relationship of one supervoxel matching pair with the rest of C. Let's denote the candidate matching pair in set C as $z = (s_i, s_j)$. The feature channels $f_z = \left(\left\|f(s_i) - f(s_j)\right\|, \frac{1}{2}\left(f(s_i) + f(s_j)\right)\right)$. The first term of f_z is the feature difference between supervoxel s_i and s_j. The second term is the location of pair z in the feature space as [14]. Instead of assigning hard labels to z as [14], we introduce the soft label $u(z) = \tau_y^{(t1)}(s_i) \cdot \tau_y^{(t_2)}(s_j)$. y is the supervoxel label of s_i and s_j. The large score u means both supervoxels in matching pair z bear a confident label in image V^{t_1} and V^{t_2}, and in turn the matching pair z is consistent with the rest of C. A conventional regression forest [3] is utilized for the consistency evaluation.

3 Experiments

Data Set. The proposed MMRF is evaluated on 150 clinical CBCT images captured from orthodontic patients for dense correspondence. According to Angle's classification (AC) of malocclusions, the data set includes 54 AC-I, 36 AC-II, 38 AC-III, as well as 22 normal occlusions. The CBCT images are acquired by a NewTom scanner with a 12-in field of view with a resolution of $500 \times 500 \times 476$. The voxel size is $0.4\,\text{mm} \times 0.4\,\text{mm} \times 0.4\,\text{mm}$.

It's not easy to define ground truth supervoxel correspondence considering the independent supervoxel decomposition. Aside from the real data set, we generate a set of toy data viewed as the golden standard. An AC-I CBCT image is supervoxel decomposed with voxels labeled according to the supervoxel indices. Twenty random B-spline based non-rigid deformations are imposed on the CBCT and the label images simultaneously. The resulted volume image data set T_u has the ground truth supervoxel labels.

Implementation Details. The 4-fold cross validation is used. The toy data set T_u is just used for testing. Each volume image is decomposed to 5k supervoxels by the SLIC technique [1]. In the training process, the MMRF is updated n_k times, and $n_k = 5$. All forests, including the clustering forest for the initial affinity and MMRF, have 10 trees and the leaf size set at 5. Given the consistency evaluation (Sect. 2.3), the pairs with scores <0.1 are unlinked, and new counterpart $s_j^{'}$ is located by a straightforward interpolation, and $x(s_{j'}) = \sum_{s_k \in Neib(s_i)} w_k x(\phi(s_k))$, where x denotes 3D coordinates of supervoxel centers. $w_{k'} = \frac{\exp(-\|x(s_i) - x(s_{k'})\|)}{\sum_{s_k \in Neib(s_i)} \exp(-\|x(s_i) - x(s_k)\|)}$. The method with the above regularization is denoted as MMRF-CR.

3.1 Qualitative Assessment

The correspondence accuracy is qualitatively assessed by two metrics: the Dice similarity coefficients (DSC) and the average Hausdorff distance (AHD). As to

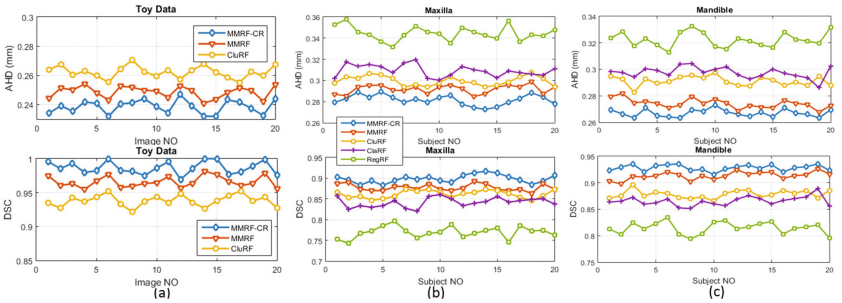

Fig. 2. Comparisons of the label transfer in (a) the toy data set T_u, and of (b) the maxilla and (c) the mandible on clinically-captured twenty images by the proposed MMRF and MMRF-CR, as well as the CluRF [10], RegRF [7], ClaRF [7] methods in terms of the AHD and DSC.

the toy data set, the DSCs are 97% and 99% by MMRF and MMRF-CR respectively, and the AHDs are 0.16 mm and 0.15 mm as shown in Fig. 2(a).

The dense correspondence facilitates the attribute transfer between images. We conduct the qualitative assessments of the proposed MMRF in the label transfer of the mandible and maxilla on captured CBCT images. The upper dentition is assigned to the maxilla, and the lower dentition to the mandible as [13]. The proposed MMRF method is compared with the recent random forest based label propagation methods, including the regression forest (RegRF) and the supervoxel classification forest (ClaRF) [7]. We also compare with the traditional patch-based fusion (PF) [5], and the convex optimization (CO) [13] methods as shown in Table 1. The label transfer accuracies of the maxilla and mandible of twenty images are shown in Fig. 2(b), (c). The consistency scheme is used to find reliable matching pairs for the final correspondence. As shown in Table 1 and Fig. 3(a)–(c), the label transfer with consistency regularization (MMRF-CR) outperforms others. We also compare the label transfer by the MMRFs built on different feature channels, i.e. f_a alone, (f_a, f_c), and (f_a, f_c, f_g) as shown in Fig. 3(d). The forest built upon all feature channels performs best in the mandible label transfer task. The proposed method utilizes the iterative refinement to improve the forest-based metric, where the iteratively updated weak labeling and penalized weights are used to constrain the

Table 1. Comparisons of the proposed MMRF and MMRF-CR, with the CluRF [10], RegRF [7], ClaRF [7], PF [5], and CO [13] methods in terms of the DSC and AHD for the label transfer of the maxilla (Mx) and mandible (Md).

		MMRF	MMRF+CR	CluRF [10]	RegRF [7]	ClaRF [7]	PF [5]	CO [13]
Mx	DSC	0.88 ± 0.02	$\mathbf{0.90 \pm 0.02}$	0.86 ± 0.03	0.76 ± 0.03	0.81 ± 0.03	0.81 ± 0.03	0.87 ± 0.02
	AHD	0.29 ± 0.02	$\mathbf{0.28 \pm 0.02}$	0.30 ± 0.03	0.34 ± 0.03	0.31 ± 0.03	n/a	n/a
Md	DSC	0.91 ± 0.02	$\mathbf{0.93 \pm 0.02}$	0.89 ± 0.02	0.81 ± 0.03	0.88 ± 0.02	0.88 ± 0.02	0.91 ± 0.02
	AHD	0.28 ± 0.02	$\mathbf{0.27 \pm 0.01}$	0.30 ± 0.03	0.32 ± 0.03	0.30 ± 0.03	n/a	n/a

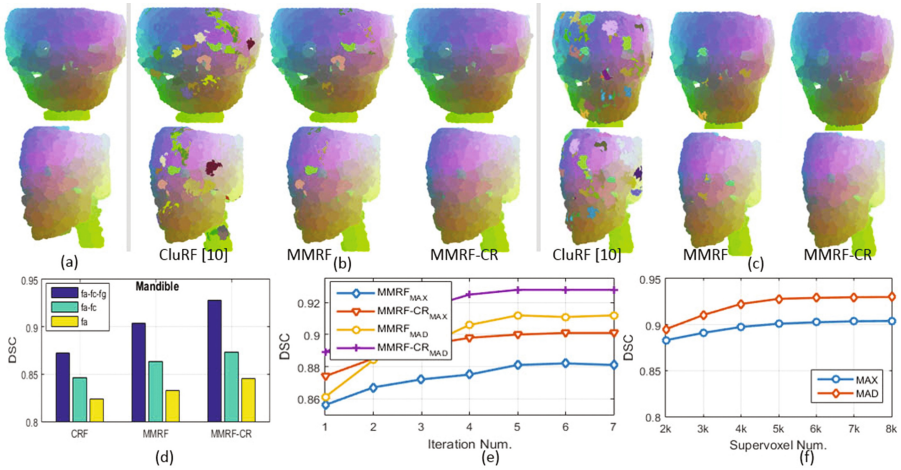

Fig. 3. (a)The reference image and the label transfer of two images, (b) and (c), by the CluRF [10], MMRF, MMRF-CR (from left to right) with pseudo-colors denoting corresponding supervoxel pairs in two viewpoints. (d) Comparisons of label transfer based on different feature channels. (e) Comparisons of label transfer with increasing iteration numbers. (f) Comparisons of label transfer with different supervoxel sizes.

entropy estimation. Figure 3(e) shows the performances after different number of iterations. We observe that the correspondence improves with iterative refinements and reaches a plateau after 5 iterations. We analyze the effects of the supervoxel size on the label transfer. Figure 3(f) shows the label transfer performances based on volumetric decompositions with $2k-8k$ supervoxels for each CBCT image. We observe that the more supervoxels, the more accurate correspondence can be achieved. We think the finer supervoxel decomposition helps to find anatomically-accurate counterparts between images. On the other hand, a large number of supervoxels impose both training and testing burden. In our system, the supervoxel number is set at $5k$ to trade off the accuracy and the computation burden.

4 Discussion and Conclusion

We propose a totally-unsupervised dense correspondence method for CBCT images. The iteratively-updated MMRF-based metric is reinforced to handle the previous poorly-clustered supervoxels. We perform experiments on the toy data generated from CBCT images of an AC-I patient. The label transfer is perfect on the toy data by the MMRF-CR with a DSC of 99%. The DSC and AHD of the maxilla label transfer are 90% and 0.28 mm respectively, and 93% and 0.27 mm for the mandible on the real clinically-captured CBCT images. Also, the computation is efficient by the forest-based metric and consumes average 5 s. The experiments are performed on the label transfer of relatively large structures. However, we observe that the segmentation of some small structures,

e.g. the teeth, are sensitive to the granularity of supervoxel decomposition yet desirable for treatment evaluations. In future work, we will investigate the potential of the MMRF for accurate correspondence for multi-scale structures.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant 61272342, and the Seeding Grant for Medicine and Information Sciences of Peking University under Grant 2014MI24.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. PAMI **34**(11), 2274–2282 (2012)
2. Bhagalia, R., Fessler, J.A., Kim, B.: Accelerated nonrigid intensity-based image registration using importance sampling. IEEE Trans. MI **28**(8), 1208–1216 (2009)
3. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
4. Cevidanes, L.H., Motta, A., Proffit, W.R., Ackerman, J.L., Styner, M.: Cranial base superimposition for 3-dimensional evaluation of soft-tissue changes. Am. J. Orthod. Dentofac. Orthoped. **137**(4), S120–S129 (2010)
5. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**, 940–954 (2011)
6. Criminisi, A., et al.: Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends Comput. Graph. Vis. **7**(23), 81–227 (2012)
7. Kanavati, F., Tong, T., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D., Glocker, B.: Supervoxel classification forests for estimating pairwise image correspondences. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) MLMI 2015. LNCS, vol. 9352, pp. 94–101. Springer, Cham (2015). doi:10.1007/978-3-319-24888-2_12
8. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. MI **16**(2), 187–198 (1997)
9. Park, J.H., et al.: 3-dimensional cone-beam computed tomography superimposition: a review. In: Seminars in Orthodontics, vol. 21, pp. 263–273. Elsevier (2015)
10. Pei, Y., Kim, T., Zha, H.: Unsupervised random forest manifold alignment for lipreading. In: IEEE ICCV (2013)
11. Pei, Y., Ma, G., Chen, G., Zhang, X., Xu, T., Zha, H.: Superimposition of cone-beam computed tomography images by joint embedding. IEEE Trans. BME **64**, 1218–1227 (2016)
12. Rodolà, E., Rota Bulo, S., Windheuser, T., Vestner, M., Cremers, D.: Dense non-rigid shape correspondence using random forests. In: CVPR, pp. 4177–4184 (2014)
13. Wang, L., et al.: Automated segmentation of CBCT image using spiral CT atlases and convex optimization. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8151, pp. 251–258. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40760-4_32
14. Zhu, X., Loy, C.C., Gong, S.: Constrained clustering with imperfect oracles. IEEE Trans. NNLS **27**(6), 1345–1357 (2016)
15. Zikic, D., Glocker, B., Criminisi, A.: Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. Med. Image Anal. **18**(8), 1262–1273 (2014)