Hierarchical Multi-Scale Supervoxel Matching using Random Forests for Automatic Semi-Dense Abdominal Image Registration

Pierre-Henri Conze, Florian Tilquin, Vincent Noblet, François Rousseau, Fabrice Heitz, Patrick Pessaux

HIERARCHICAL MULTI-SCALE SUPERVOXEL MATCHING USING RANDOM FORESTS FOR AUTOMATIC SEMI-DENSE ABDOMINAL IMAGE REGISTRATION

P.-H. Conze[⋆,◇], *F. Tilquin*[⋆], *V. Noblet*[⋆], *F. Rousseau*[◇], *F. Heitz*[⋆] *and P. Pessaux*[△]

[⋆] ICube UMR 7357, Université de Strasbourg, CNRS, FMTS, Strasbourg, France
[◇] Institut Mines Télécom Atlantique, LaTIM UMR 1101, INSERM, Brest, France
[△] Institut Hospitalo-Universitaire de Strasbourg, France

ABSTRACT

This paper addresses the estimation of pairwise supervoxel correspondences toward automatic semi-dense medical image registration. Supervoxel matching is performed through random forests (RF) with supervoxel indexes as label entities to predict matching areas in another target image. Ensuring accurate supervoxel boundary adherence requires a fine supervoxel decomposition which highly increases learning complexity. To alleviate this issue, we extend RF based supervoxel matching from single to multi-scale using a recursive hierarchical supervoxel representation. Output RF matching probabilities obtained for the last scale are gathered with ancestor matching probabilities which acts as a coarse-to-fine matching guidance. The effectiveness of our method is highlighted for semi-dense abdominal image registration relying on liver label propagation and consistency assessment.

Index Terms— semi-dense image registration, supervoxel matching, random forests, hierarchical multi-scale

1. INTRODUCTION

Estimating an accurate matching between consistent entities such as voxels, supervoxels or patches is a fundamental task in image processing with applications to medical image analysis. In particular, non-rigid image registration is required for a wide range of medical contexts including organ motion compensation, longitudinal tumor follow-up, pre- and post-operative image matching or cohort analysis [1].

These applications require an initial alignment used as an initialization toward a more complex deformation model estimation. Such complex geometric transformations can be then derived either from physical models (fluid flow or diffusion models, flow of diffeomorphisms...) or interpolation theory (radial basis functions, free-form deformations...) [2]. For accurate initialization matching, spatial prior is usually considered based on anatomy recognition or landmark matching [3]. It requires expert annotations which are time-consuming and prone to misinterpretation and human error.

To avoid this interaction, random forests (RF) [4] have been recently proposed to establish supervoxel correspondences in a unsupervised fashion [5]. Similarly to [6] which encodes a single labeled image as a forest for multi-atlas label propagation, [5] describes a single source image as a collection of supervoxels whose indexes are used *via* RF to predict matching areas in another target image, leading to semi-dense correspondences. For the purpose of supervoxel matching, the selection of the supervoxel resolution is key to ensure accurate boundary adherence of supervoxels. The underlying issue is how to find a good trade-off between large supervoxels which may overlap different tissue types or organs and small supervoxels without enough discriminative power [7], leading to highly complex learning. Merging information from multiple supervoxel layers may achieve a more flexible representation [8] but does not carry a hierarchical scheme allowing coarse-to-fine matching guidance.

In this perspective, we extend automatic semi-dense medical image registration through RF from single [5] to multi-scale supervoxel matching. Relying on a hierarchical multi-scale supervoxel representation successfully applied to multi-class tissue classification [9, 7], our framework implicitly constrains the matching search space in a coarse-to-fine strategy. We demonstrate its effectiveness in the field of CT abdominal image registration which remains an open issue due to wide organ size, shape and appearance heterogeneity.

2. SUPERVOXEL MATCHING USING RF

Automatic semi-dense image registration is considered between two 3D abdominal images I_f and I_s where each image $I_p : \Omega_p \subset \mathbb{N}^3 \rightarrow \mathbb{N}$ associates a greyscale value $I_p(\boldsymbol{x})$ to each voxel $\boldsymbol{v}_p$ located at $\boldsymbol{x}_p \in \Omega_p$ with $p \in \{f, s\}$. Toward this goal, we aim at establishing correspondences between supervoxels over-segmenting I_f and I_s following [5]. Let $\mathcal{F} = \{f_i\}_{i \in \{1,...,|\mathcal{F}|\}}$ and $\mathcal{S} = \{s_j\}_{j \in \{1,...,|\mathcal{S}|\}}$ be respectively the sef of $|\mathcal{F}|$ and $|\mathcal{S}|$ connected supervoxels partitioning I_f and I_s. The supervoxel decomposition is performed using a 3D extension of the Simple Linear Iterative Clustering (SLIC) superpixel algorithm [10] which aggregates neighboring voxels $\boldsymbol{v}_p$ based on spatial and intensity proximity criteria.
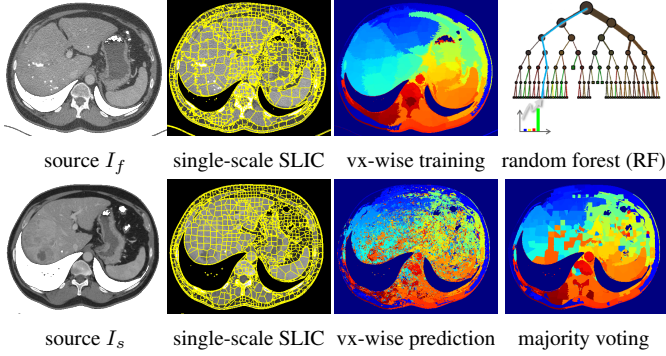
Fig. 1. Single-scale supervoxel matching [5] between I_f and I_s using random forest [4] applied with SLIC [10] supervoxel indexes as label entities followed by majority voting.

Semi-dense image registration translates in automatically learning a mapping function h that maps each supervoxel $s_j \in \mathcal{S}$ of I_s to a given supervoxel $f_i \in \mathcal{F}$ of I_f [5] such that $\forall j \in \{1, \ldots, |\mathcal{S}|\}, \exists i \in \{1, \ldots, |\mathcal{F}|\} \mid h(s_j) = f_i$.

2.1. Single-scale supervoxel matching

Single-scale supervoxel matching (Fig.1) is carried out *via* random forest (RF) [4]. RF builds a voxel/supervoxel index mapping model aiming at assigning a supervoxel $f_i \in \mathcal{F}$ to each voxel $\boldsymbol{v}_s \in \Omega_s$. Thus, a training set is built by considering all voxels $\boldsymbol{v}_f$ of Ω_f with associated supervoxel index, i.e. the index of the supervoxel f_i they belong to. Once trained, RF is applied to I_s to predict for each $\boldsymbol{v}_s$ the index of a supervoxel of $\mathcal{F}$. Majority voting is performed within each supervoxel of $\mathcal{S}$ among all voxel-wise predictions to return for each s_j of I_s the best match $h(s_j)$ among supervoxels of $\mathcal{F}$.

RF consists of T uncorrelated trees made of both internal nodes splitting data according to binary tests Ψ and leaf nodes which reach all together a final data partition. At each internal node, the split sends voxels $\boldsymbol{v}_p$ to left and right children nodes. The associated binary test Ψ focuses on a random subset $\hat{\theta}(\boldsymbol{v}_p)$ ($p \in \{f, s\}$) of visual features $\theta(\boldsymbol{v}_p)$ assigned to $\boldsymbol{v}_p$ (Sect.2.2) and divides the input voxel set based on the following rule:

$$\Psi(\boldsymbol{v}_p, \theta(\boldsymbol{v}_p)) = \begin{cases} \text{true, if } \hat{\theta}(\boldsymbol{v}_p) > \tau \\ \text{false, otherwise} \end{cases} \quad (1)$$

where $\hat{\theta}(\boldsymbol{v}_p)$ is compared to a threshold τ.

Internal node parameters ($\{\tau, \hat{\theta}(\boldsymbol{v}_p)\}$) are optimized *via* information gain maximization with respect to $\mathcal{L} = \{\boldsymbol{v}_f, c(f_i)\}$ combining voxels $\boldsymbol{v}_f$ belonging to f_i with their associated supervoxel index $c(f_i) = i$ with $i \in \{1, \ldots, |\mathcal{F}|\}$. After optimization, each leaf node l_t of the tth tree receives a partition $\mathcal{L}_{l_t}$ of $\mathcal{L}$ and produces an entire class probability distribution: $\mathrm{p}_{l_t}(c(f_i)|\mathcal{L}) \, \forall i \in \{1, \ldots, |\mathcal{F}|\}$.

To predict the corresponding supervoxel index $c(f_i)$ of a given voxel $\boldsymbol{v}_s \in \Omega_s$ with associated visual features $\theta(\boldsymbol{v}_s)$ during testing, $\boldsymbol{v}_s$ is injected into each optimized tree and finally

reaches a leaf node l_t per tree following successive split rules (Eq.1). Let us define $\mathrm{p}(h(\boldsymbol{v}_s) = f_n)$ as the probability that f_n is assigned to voxel $\boldsymbol{v}_s$. For each supervoxel $f_n \in \mathcal{F}$, we get:

$$\mathrm{p}(h(\boldsymbol{v}_s) = f_n) = \frac{1}{T} \sum_{t=1}^{T} \mathrm{p}_{l_t}(c(f_n)|\mathcal{L})$$

$$= \frac{1}{T} \sum_{t=1}^{T} \frac{|\{\boldsymbol{v}_f, c(f_l)\} \in \mathcal{L}_{l_t} \mid l = n|}{|\mathcal{L}_{l_t}|} \quad (2)$$

The final voxel ($\boldsymbol{v}_s$) to supervoxel (f_i) mapping is obtained by maximizing $\mathrm{p}(h(\boldsymbol{v}_s) = f_n)$.

$$h(\boldsymbol{v}_s) = f_i = \arg \max_{f_n \in \mathcal{F}} \mathrm{p}(h(\boldsymbol{v}_s) = f_n) \quad (3)$$

The last step deals with majority voting among all the voxels of a given supervoxel $s_j \in \mathcal{S}$. The most represented supervoxel index is thus selected as follows:

$$h(s_j) = f_i \mid c(f_i) = \arg \max \mathtt{hist}(\{c(h(\boldsymbol{v}_s)) \mid \boldsymbol{v}_s \in s_j\}) \quad (4)$$

2.2. Voxel-wise context appearance features

Denoting $\bar{I}_p^w(\boldsymbol{v}_p)$ the mean intensity on a local box of size w centered on v_p, context appearance features [6, 3, 5] $\theta(\boldsymbol{v}_p) = \{\theta_m(\boldsymbol{v}_p)\}_{m \in \{0, \ldots, K_a - 1\}}$ assigned to voxels $\boldsymbol{v}_p$ following are:

$$\theta_m(\boldsymbol{v}_p) = \bar{I}_p^w(\boldsymbol{v}_p + \Delta_r) - b \times \bar{I}_p^{w'}(\boldsymbol{v}_p + \Delta_{r'}) \quad (5)$$

where $\Delta_{\{r, r'\}} \in \mathcal{B}(0, \Phi)$, Φ is the maximal offset and $b \in \{0, 1\}$. By randomly generating many different box sizes w and offsets Δ_r, we obtain a large set of K_a features describing the extended spatial context. $\{w, w', \Delta_r, \Delta_r', b\}$ are generated once and remain similar for each voxel of I_f and I_s.

3. EXTENSION TO A HIERARCHICAL MULTI-SCALE REPRESENTATION

Using single-scale supervoxel decomposition for semi-dense registration is prone to errors. Supervoxel resolutions must be carefully selected since large supervoxels may overlap different tissues or organs contrary to small ones which lead to computationally intensive learning while being not enough representative (huge number of classes with few samples) [7]. Manually setting the scale is cumbersome since it requires exhaustive trials. Even the best compromise may not provide optimal results without relying on multi-scale guidance to accurately match consistent anatomies. We now describe how [5] can be extended in a hierarchical multi-scale fashion.

3.1. Hierarchical multi-scale supervoxel representation

The proposed original extension (Fig.2) is based on a hierarchical multi-scale supervoxel representation [9, 7]. For I_f, it requires a decomposition of Ω_f into a set of $K + 1$ partitions
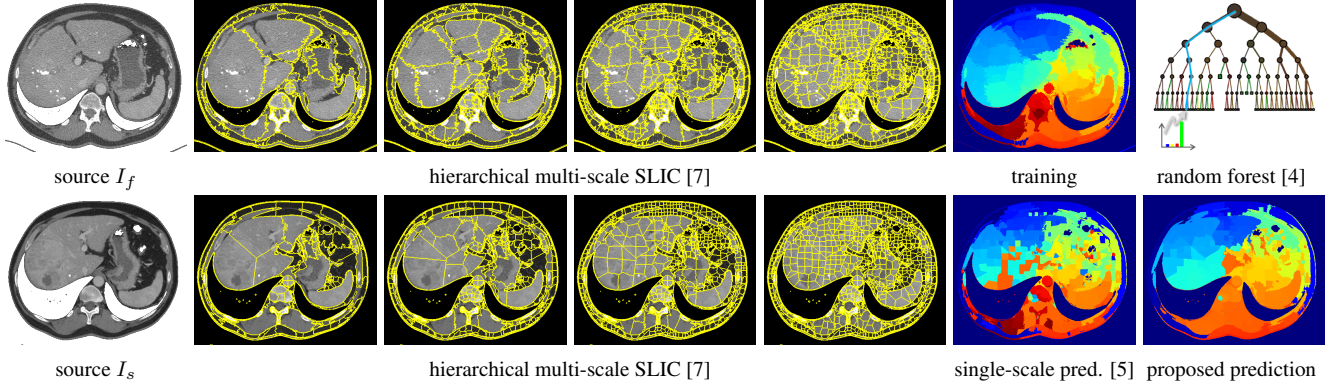
Fig. 2. Proposed extension of automatic semi-dense registration *via* random forest [4] from single [5] to multi-scale supervoxel matching based on a hierarchical supervoxel representation [7]. Colors for training and prediction illustrate supervoxel indexes.

$\mathcal{P}_f^k = \{f_i^k\}$ defined at scales $k \in \{0, \ldots, K\}$ where 0 and K denote resp. coarsest and finest scales. Each partition $\{\mathcal{P}_f^k\}$ is a collection of connected 3D supervoxels $\{f_i^k\}$ built at scale k such that $f_i^k \cap f_{j \neq i}^k = \emptyset$ and $\bigcup_i f_i^k = \Omega_f$. The sequence of partitions $\{\mathcal{P}_f^k\}$ is encoded in the layers of a multi-resolution tree $\mathcal{M}_f = \{\mathcal{M}_f^k\}$ where the layer $\mathcal{M}_f^k$ maps each supervoxel $f_i^k \in \mathcal{P}_f^k$ to a set of child supervoxels $\{f_j^{k+1}\} \subset \mathcal{P}_f^{k+1}$ such that $f_i^k = \bigcup_j f_j^{k+1}$. This multi-scale decomposition is similarly applied to I_s leading to $K+1$ partitions $\mathcal{P}_s^k = \{s_i^k\}$.

This representation is obtained through a recursive coarse-to-fine process. We consider an initial partition $\mathcal{P}_f^0$ (resp. $\mathcal{P}_s^0$) of only one supervoxel. This supervoxel is then split into K_s children supervoxels $\{f_j^1\} \subset \mathcal{P}_f^1$ ($\{s_j^1\} \subset \mathcal{P}_s^1$) using SLIC applied on the region formed by its voxels[1]. Once built, each f_j^1 (s_j^1) is split into K_s children and this procedure is repeated iteratively down to the Kth layer of $\mathcal{M}_f$ ($\mathcal{M}_s$) [7].

3.2. Hierarchical multi-scale supervoxel matching

The first step consists in performing a single-scale supervoxel matching [5] for each scale k independently using context features $\theta(\boldsymbol{v}_p)$. Each supervoxel s_j^k belonging to partition $\mathcal{P}_s^k$ is thus matched to a given supervoxel $h(s_j^k) = f_i^k \in \mathcal{P}_f^k$. Then, the hierarchical matching occurs. Let $\mathcal{A}(f_i^k, d) \in \mathcal{P}_f^{k-d}$ be the ancestor of f_i^k of degree d defined at scale $k - d$. The intuition behind our contribution is that the matching probability obtained *via* RF at scale k should be affected by the matching probabilities arising from the previous scales. We rely in practice on a multi-scale probability aggregation strategy starting from supervoxels of scale K. Their output RF matching probabilities are updated as follows:

$$\hat{p}\big(h(s_j^K) = f_i^K\big) = p\big(h(s_j^K) = f_i^K\big)$$
$$\times \prod_{d=1}^{K} p\big(h(\mathcal{A}(s_j^K, d)) = \mathcal{A}(f_i^K, d)\big) (6)$$

---

[1] implementation of B. Irving, http://maskslic.birving.com.

The supervoxel (s_j^K) to supervoxel (f_i^K) mapping at scale K is finally obtained by maximizing $\hat{p}(h(s_j^K) = f_n^K)$.

$$h(s_j^K) = f_i^K = \arg \max_{f_n^K \in \mathcal{P}_f^K} \hat{p}(h(s_j^K) = f_n^K) \qquad (7)$$

By focusing on finest supervoxels, we get a fine representation not prone to overlap issues. Single-scale matching [5] applied on these supervoxels results in many bad matches (Fig.2) due to too many small-size, non-discriminative classes. Our coarse-to-fine hierarchical guidance alleviates ambiguities by implicitly reducing the matching search space.

4. SEMI-DENSE REGISTRATION RESULTS

Experiments focus on data collected from 25 examination pairs stemming from 6 patients with hepato-cellular carcinoma (HCC). Each pair brings together two CT scans acquired for the same patient at different time points varying from 41 to 407 days. Image pairs are processed both in forward (FW) and backward (BW) which translates in performing learning on I_f (resp. I_s) and prediction on I_s (I_f) to get mapping functions $h_{\text{FW}}(s_j^k) = f_i^k$ and $h_{\text{BW}}(f_i^k) = s_j^k \, \forall k$.

To compare the single (sgl-) [5] and proposed multi-scale (mlt-) supervoxel matching (SM) methods, we rely on a hierarchical supervoxel decomposition with $K = 12$ scales, empirically set to get an accurate description of each scan. sgl-SM is directly applied on the finest scale K whereas mlt-SM applies sgl-SM at each scale independently before gathering ancestor probabilities (Eq.6). Each supervoxel at scale k is split into $K_s = 2$ children defined at scale $k + 1$ allowing a progressive multi-scale guidance. A SLIC compactness [10] of 0.25 with CT intensities rescaled to $[0, 1]$ reaches a good trade-off between compactness and boundary adherence. $K_a = 400$ features are generated for each $\boldsymbol{v}_p$ with a maximal offset of $\Phi = 50$ voxels and box sizes $w \in \{3, 5, 7\}$.

Supervoxel matching allows a straightforward propagation of anatomical labels. To evaluate {sgl, mlt}-SM quantitatively, we provide liver label propagation results through

	sgl-SM	mlt-SM	sgl-SM	mlt-SM
SLIC	without a-priori		with a-priori	
DICE	72.21 ± 9.72	$\mathbf{82.27 \pm 5.00}$	67.82 ± 13.8	$\mathbf{95.21 \pm 6.64}$
sens	72.60 ± 12.7	$\mathbf{83.54 \pm 7.84}$	62.56 ± 17.8	$\mathbf{93.27 \pm 10.1}$
spec	92.91 ± 5.23	$\mathbf{95.28 \pm 3.08}$	95.23 ± 3.81	$\mathbf{99.55 \pm 0.91}$
$\chi(h)$	21.10 ± 7.03	$\mathbf{10.07 \pm 4.49}$	24.33 ± 6.91	$\mathbf{10.27 \pm 4.46}$

Table 1. Quantitative comparisons of RF-based single-scale (sgl-) [5] and hierarchical multi-scale (mlt-) supervoxel matching (SM) *via* liver label propagation and inconsistency averaged over the database (FW and BW). Best results in bold.

DICE, sensitivity and specificity scores comparing liver propagation and ground-truth[2] (GT) masks. Supervoxel correspondences are also assessed through FW/BW inconsistency scores defined for each supervoxel f_i^K and s_j^K similarly to:

$$\chi(f_i^K) = \left\| \mathcal{C}(f_i^K) - \mathcal{C}(h_{\text{BW}}(h_{\text{FW}}(f_i^K))) \right\|_2 \quad (8)$$

with $\mathcal{C}(.)$ the supervoxel centroid. Global inconsistency scores $\chi(h_{\text{FW}})$ (resp. $\chi(h_{\text{BW}})$) for $\{I_f, I_s\}$ are then computed *via* a normalized sum of all $\chi(f_i^K)$ (resp. $\chi(s_j^K)$) weighted by their belonging voxels number, $|f_i^K|$ ($|s_j^K|$).

$\{$sgl,mlt$\}$-SM are applied with two different SLIC decompositions: without and with a-priori liver segmentation awareness to evaluate supervoxels matchings without being disturbed by boundary adherence issues. In this latter case, the first scale is used to separate the liver from other abdominal structures. Multi-scale decompositions are then performed independently for both areas and finally merged.

We present a comparative assessment of $\{$sgl,mlt$\}$-SM in Tab.1. Without a-priori, it reveals better results using mlt-SM which achieves highest DICE (82.3 against 72.2), sens and spec combined with lowest inconsistency (from 22.1 to 10.1 voxels). More significant gains arise with liver segmentation a-priori, especially for DICE whose improvement is ≈27.4%. DICE for sgl is worse with than without a-priori since liver-awarness decomposition generates more supervoxels by construction. Despite additional computational cost, results confirm the significant impact of our multi-scale matching probability aggregation scheme, as highlighted by the accuracy of prediction and liver propagation shown Fig.3.

5. CONCLUSION

Our work focuses on automatic semi-dense image registration using maching learning. We extended random forest (RF) based supervoxel matching from single to multi-scale using a recursive hierarchical supervoxel representation. RF matching probabilities are gathered together with ancestor probabilities acting as a coarse-to-fine guidance. We demonstrate its effectiveness on challenging abdominal CT scans and will further explore this scheme using forward-backward consistency for multi-phases and multi-modal registration.

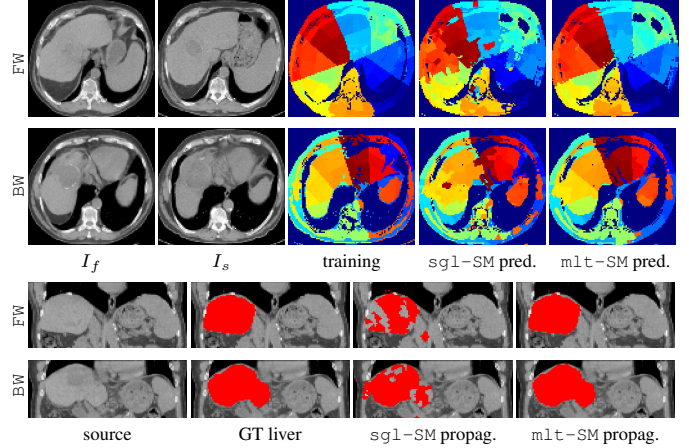[2] provided by Visible Patient, www.visiblepatient.com

Fig. 3. Assessment of $\{$sgl,mlt$\}$-SM with training, prediction (no a-priori), GT and propagated liver (a-priori) masks.

6. REFERENCES

[1] J. A. Schnabel, M. P. Heinrich, B. W. Papież, and J. M. Brady, "Advances and challenges in deformable image registration: From image fusion to complex motion modelling," *Medical Image Analysis*, vol. 33, pp. 145–148, 2016.

[2] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 7, pp. 1153–1190, 2013.

[3] B. Glocker, D. Zikic, and D. R. Haynor, "Robust registration of longitudinal spine CT," in *Medical Image Computing and Computer-Assisted Intervention*, 2014, pp. 251–258.

[4] L. Breiman, "Random Forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] F. Kanavati, T. Tong, K. Misawa, M. Fujiwara, K. Mori, D. Rueckert, and B. Glocker, "Supervoxel classification forests for estimating pairwise image correspondences," *Pattern Recognition*, vol. 63, pp. 561–569, 2016.

[6] D. Zikic, B. Glocker, and A. Criminisi, "Encoding atlases by randomized classification forests for efficient multi-atlas label propagation," *Medical image analysis*, vol. 18, no. 8, pp. 1262–1273, 2014.

[7] P-H. Conze, V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux, "Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans," *International Journal of Computer Assisted Radiology and Surgery*, 2016, published online.

[8] M. P. Heinrich, I. J.A. Simpson, B. W. Papież, M. Brady, and J. A. Schnabel, "Deformable image registration by combining uncertainty estimates from supervoxel belief propagation," *Medical image analysis*, vol. 27, pp. 57–71, 2016.

[9] E. Geremia, B. H Menze, and N. Ayache, "Spatially adaptive random forests," in *IEEE International Symposium on Biomedical Imaging*, 2013, pp. 1344–1347.

[10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, 2012.