

Semi-automatic Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans Using Random Forests and Supervoxels

Pierre-Henri Conze, François Rousseau, Vincent Noblet, Fabrice Heitz, Riccardo Memeo, Patrick Pessaux

▶ **To cite this version:**

Pierre-Henri Conze, François Rousseau, Vincent Noblet, Fabrice Heitz, Riccardo Memeo, et al.. Semi-automatic Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans Using Random Forests and Supervoxels. 6th International Workshop on Machine Learning in Medical Imaging (MLMI 2015), Oct 2015, Munich, Germany. pp.212-219, 10.1007/978-3-319-24888-2_26 . hal-01214060

HAL Id: hal-01214060
https://hal.science/hal-01214060

Submitted on 9 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-automatic liver tumor segmentation in dynamic contrast-enhanced CT scans using random forests and supervoxels

Pierre-Henri Conze[1], François Rousseau[2], Vincent Noblet[1], Fabrice Heitz[1], Riccardo Memeo[3] and Patrick Pessaux[3]

[1] ICube, Université de Strasbourg, CNRS, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Strasbourg, France
[2] Institut Mines-Télécom, Télécom Bretagne, INSERM, LATIM, Brest, France
[3] Department of Hepato-Biliary and Pancreatic Surgery, Nouvel Hôpital Civil, Institut Hospitalo-Universitaire de Strasbourg, Strasbourg, France

Abstract. Pre-operative locoregional treatments (PLT) delay the tumor progression by necrosis for patients with hepato-cellular carcinoma (HCC). Toward an efficient evaluation of PLT response, we address the estimation of liver tumor necrosis (TN) from CT scans. The TN rate could shortly supplant standard criteria (RECIST, mRECIST, EASL or WHO) since it has recently shown higher correlation to survival rates. To overcome the inter-expert variability induced by visual qualitative assessment, we propose a semi-automatic method that requires weak interaction efforts to segment parenchyma, tumoral active and necrotic tissues. By combining SLIC supervoxels and random decision forest, it involves discriminative multi-phase cluster-wise features extracted from registered dynamic contrast-enhanced CT scans. Quantitative assessment on expert groundtruth annotations confirms the benefits of exploiting multi-phase information from semantic regions to accurately segment HCC liver tumors.

1 Introduction

Hepato-cellular carcinoma (HCC) is the most common type of liver cancer and the third most frequent cause of cancer-related death. Pre-operative locoregional treatments (PLT) tend to downstage HCC tumors by necrosis. Standard evaluation scores (RECIST, mRECIST, EASL or WHO) used to predict the response to PLT do not provide fully satisfactory results [1]. A more efficient HCC patient follow-up is reached through tumor necrosis (TN) rate which provides more significant correlation with survival rates.

To overcome inter-expert variability induced by visual qualitative assessment, we present a computed-aided diagnosis method for TN rate computation. Assessing TN remains an open issue due to a wide variability of shape, size, location, contour aspect and intensity of HCC tumors. It requires the segmentation of healthy liver parenchyma as well as tumoral active and necrotic areas (Fig. 4).

For this task, dynamic contrast-enhanced images provide discriminative information since HCC is characterized by arterial enhancement followed by venous washout in response to contrast agent injection [2].

Dynamic contrast-enhanced images have been recently exploited for HCC liver tumor segmentation via extraction of *multi-phase voxel-wise* features used in level sets [3], k-means [4] or graph cuts [5]. These multi-phase features capture the dynamic in response to agent injection and tend to build a full perfusion model from low temporal resolution data (Fig. 3*b*). However, to reach a better accuracy, user interaction appears necessary since arterial enhancement and venous washout depend on contrast agent kinetic and injection protocol [2]. In this direction, tumor extraction has been covered in an interactive perspective for single-phase images [6,7] by relying on supervised ensemble learning with *single-phase voxel-wise* features including spatial characteristics. Although more appropriate, such strategy is voxel-wise as in [3,4,5] and therefore requires a significant amount of interaction while relying on a limited spatial context

In this work, we propose to exploit robust *multi-phase cluster-wise* features extracted from registered multi-phase contrast-enhanced CT scans by combining clustering and supervised ensemble learning. Performing interactive learning and prediction on semantic regions allow tumor segmentation to take advantage of discriminative dynamic information at an extended spatial extent. Moreover, it ensures weak interactions efforts for practitioners. This method applied to TN rate estimation is a key step toward accurate PLT response assessment.

2 Methodology

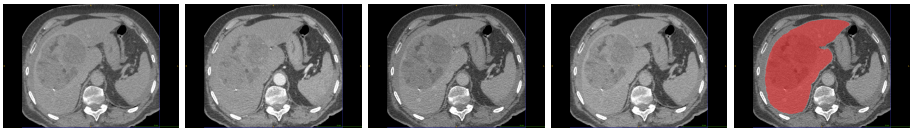
2.1 Traditional approach: *single-phase voxel-wise* random forest

In an interactive setting for image segmentation, the user usually defines manually a set of K labeled voxels $\boldsymbol{S} = \{\boldsymbol{v}_k, c(\boldsymbol{v}_k)\}_{k\in\{1,...,K\}}$ where $c(\boldsymbol{v}_k) \in \{c_1, ..., c_N\}$ is the label at voxel $\boldsymbol{v}_k$ with N the total class number. The set $\boldsymbol{S}$, referred as *training data*, is used to build a voxel/label mapping model whose aim is to predict the label $c(\boldsymbol{v})$ of each test voxel $\boldsymbol{v}$. In the context of ensemble learning methods, random decision forests [8] have grown in popularity due to their ability to offer a unified framework for many machine learning tasks [9].

A random decision forest consists of T independent trees made of both *internal nodes* which split input data according to binary tests and *terminal nodes* which reach all together a final data partition. At each internal node, the split sends voxels to left and right children nodes. For this task, the associated binary test focuses on a randomly subset $\hat{\theta}(\boldsymbol{v})$ of the visual features $\theta(\boldsymbol{v})$ extracted for input voxels $\boldsymbol{v}$ and halve the input dataset according to:

$$h(\boldsymbol{v}, \theta) = \begin{cases} \text{true, if } \tau_{low} < \hat{\theta}(\boldsymbol{v}) < \tau_{up} \\ \text{false, otherwise} \end{cases} \quad (1)$$

where $\hat{\theta}(\boldsymbol{v})$ is compared to thresholds τ_{low} and τ_{up}. In the context of tumor segmentation, methods relying on random decision forest such as [7,10] usually focus on one single image and therefore involve *single-phase* visual features $\theta(\boldsymbol{v})$.



$\tilde{I}_{\texttt{WI}}$ $\tilde{I}_{\texttt{AR}}$ $I_{\texttt{EV}}$ $\tilde{I}_{\texttt{LV}}$ liver mask

Fig. 1. Registered multi-phase input data ($\tilde{I}_{\texttt{WI}}$, $\tilde{I}_{\texttt{AR}}$, $I_{\texttt{EV}}$ and $\tilde{I}_{\texttt{LV}}$) obtained from dynamic contrast-enhanced CT scans with associated liver segmentation mask.

During training, each tree takes the *training voxel* set $\boldsymbol{S}$ as input and optimizes its own internal nodes ($\{\tau_{low}, \tau_{up}, \hat{\theta}(\boldsymbol{v})\}$) via information gain maximization [9] to obtain the most discriminative binary tests with respect to $\boldsymbol{S}$. After this optimization, each leaf node l_t of the t^{th} tree receives a partition $\boldsymbol{S}_{l_t}$ of the training data $\boldsymbol{S}$ and produces an entire class probability distribution: $\mathtt{p}_{l_t}(c_i|\boldsymbol{S})$ $\forall i \in \{1, ..., N\}$. To predict the label $c(\boldsymbol{v})$ of a given *test voxel* $\boldsymbol{v}$ with associated single-phase features $\theta(\boldsymbol{v})$, $\boldsymbol{v}$ is injected into each optimized tree which makes it reaches a leaf node l_t per tree following split rules. For each label c_i, we get:

$$\mathtt{p}(c(\boldsymbol{v}) = c_i) = \frac{1}{T}\sum_{t=1}^{T}\mathtt{p}_{l_t}(c(\boldsymbol{v}) = c_i|\boldsymbol{S}) = \frac{1}{T}\sum_{t=1}^{T}\frac{|\{\boldsymbol{v}_k, c(\boldsymbol{v}_k)\} \in \boldsymbol{S}_{l_t} \mid c(\boldsymbol{v}_k) = c_i|}{|\boldsymbol{S}_{l_t}|} \tag{2}$$

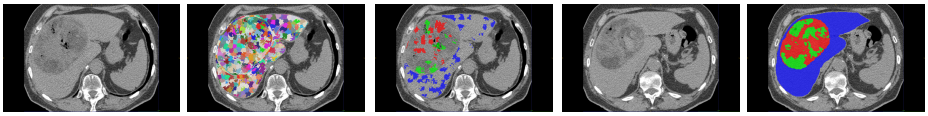
The final predition of $c(\boldsymbol{v})$ corresponds to the label c_i maximizing $\mathtt{p}(c(\boldsymbol{v}) = c_i)$:

$$c(\boldsymbol{v}) = \arg\max_{c_i} \mathtt{p}(c(\boldsymbol{v}) = c_i) \tag{3}$$

2.2 Proposed methodology: *multi-phase cluster-wise* random forest

From voxels to semantic regions. In the traditional approach (Sec. 2.1), voxels are acting without inter-dependencies which may result in a lack of spatial consistency regarding classification results. Features are sometimes explicitly related to spatial context [7,10] but spatial extent remains limited. *A-posteriori* regularization techniques such as conditional random field (CRF) [11] would more accurately introduce spatial constraints but it strongly increases computational complexity. Moreover, such voxel-wise learning-based approaches require a significant amount of interaction for the end-user to get a large enough training set. To reduce interaction efforts and intrinsically introduce spatial consistency, we perform random forest on training and test *semantic regions* instead of voxels.

Exploiting multi-phase input data. Since practitioners focus HCC diagnosis on the association of both arterial hypervascularity and venous washout [2], we propose to take full advantage of *multi-phase* contrast-enhanced CT scans. In practice, for a given examination, a contrast agent is injected to the patient and CT scans are acquired at different phases : before injection (WI) but also after at arterial (AR), early venous (EV) and late venous (LV) phases. Each examination



(a) 3D SLIC supervoxels (b) user selection (c) our 3D segmentation

Fig. 2. Semi-automatic tumor segmentation with *multi-phase cluster-wise* random forest: by performing user interaction (b) and learning on 3D supervoxels (a), it segments parenchyma (*blue*), active (*red*) and necrotic (*green*) areas of the 3D liver volume (c).

consists in a set of images $\boldsymbol{S}_I = \{I_{\texttt{WI}}, I_{\texttt{AR}}, I_{\texttt{EV}}, I_{\texttt{LV}}\}$ which are warped with respect to I_{EV} since I_{EV} exhibits greater inter-class contrasts than other acquisitions. The set of registered multi-phase images (Fig. 1) is $\boldsymbol{S}_{\tilde{I}} = \{\tilde{I}_{\texttt{WI}}, \tilde{I}_{\texttt{AR}}, I_{\texttt{EV}}, \tilde{I}_{\texttt{LV}}\}$ where $\tilde{.}$ denotes warped scans. To obtain $\boldsymbol{S}_{\tilde{I}}$ from $\boldsymbol{S}_I$, we apply a symmetric diffeomorphic non-rigid registration based on the variational formulation of [12]. A liver segmentation mask (Fig. 1) is assumed to be available at EV phase.

Proposed protocol. Our interactive tumor segmentation method translates in *multi-phase cluster-wise* random forest classification for TN rate estimation. $N = 3$ classes are considered: parenchyma, active and necrotic areas.

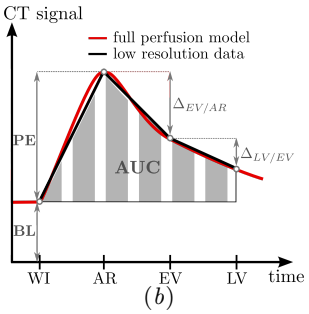
First, the liver volume is over-segmented using a 3D extension of the simple linear iterative clustering (SLIC) superpixel algorithm [13]. Starting from all the voxels belonging to the liver mask in $I_{\texttt{EV}}$, SLIC provides a set of K_R 3D compact clusters $\boldsymbol{R} = \{\boldsymbol{r}_i\}_{i\in\{1,\ldots,K_R\}}$ (Fig. 2a). Then, the interactive training cluster selection occurs. Instead of brushing strokes on many voxels, the practitioner has only to select and label a subset of $\boldsymbol{R}$ (Fig. 2b). We claim that such interaction is more suitable for clinical practice due to its simplicity and its possible integration in multi-examination training. It results in a *training* cluster set $\boldsymbol{S} = \{\boldsymbol{r}_j, c(\boldsymbol{r}_j)\}_{j\subset\{1,\ldots,K_R\}}$ combining cluster $\boldsymbol{r}_j$ with groundtruth labels $c(\boldsymbol{r}_j) \in \{c_1, c_2, \ldots, c_N\}$. A random forest is then built based on the training set $\boldsymbol{S}$ via a training procedure for which internal nodes are optimized with respect to *multi-phase cluster-wise* visual features $\theta(\boldsymbol{r}_j)$ assigned to each cluster $\boldsymbol{r}_j$.

Test clusters $\boldsymbol{r}$ ($\boldsymbol{r} \in \boldsymbol{R}\backslash\boldsymbol{S}$) are then propagated into the forest to get a label prediction $c(\boldsymbol{r})$ (Fig. 2c) based on their own *multi-phase cluster-wise* features. Finally, the TN rate τ is computed as follows: $\tau = \frac{\sum_r |\boldsymbol{r}|.\mathbf{1}_{c(\boldsymbol{r})=c_1}}{\sum_r |\boldsymbol{r}|.[\mathbf{1}_{c(v)=c_0}+\mathbf{1}_{c(v)=c_1}]}$ where $|\boldsymbol{r}|$ is the number of voxels in cluster $\boldsymbol{r}$, c_0 and c_1 active and necrosis labels.

Multi-phase cluster-wise features. The accuracy of our method is related to the ability of the features to discriminate the different tissues. We assign to each semantic region, $\boldsymbol{r}$, 20 *multi-phase cluster-wise* visual features divided into 3 groups (Fig. 3a). The two first groups introduce spatial characteristics in terms of intensity and gradient magnitude at the cluster spatial extent. In both cases and for each phase, mean and standard deviation values are computed among all voxels of $\boldsymbol{r}$. Thus, we quantify intrinsic intensity, visual homogeneity, textural information and texture-scale repartition. The third group fully exploits dynamic

Related to	*Multi-phase cluster-wise* features	Nb
Spatial intensity	mean intensity including BL	4
	standard deviation	4
Spatial gradient	mean gradient magnitude	4
	standard deviation	4
Multi-phase	peak enhancement (PE)	1
	area under enhancement curve (AUC)	1
	inter-phase diff. $\Delta_{EV/AR}$, $\Delta_{LV/EV}$	2

(a)



(b)

Fig. 3. *Multi-phase cluster-wise* features involved into random forest (a) to discriminate the different tissues and illustration of dynamic features (b).

contrast-enhanced data by combining several multi-phase intensity information averaged among the voxels of $\boldsymbol{r}$: dynamic features used in [3,4,5] including peak enhancement (PE) and area under curve (AUC) (Fig. 3b) to which is added inter-phase intensity differences $\Delta_{EV/AR}$ and $\Delta_{LV/EV}$ since PE only compares WI and AR. The baseline pre-contrast (BL) (Fig. 3b) is taken into account with intensity features. These dynamic features discriminate clusters based on their own dynamic characterized by arterial enhancement and venous washout.

3 Results

Evaluation on clinical data. The protocol has been tested on data collected from 7 examinations $\{e_1, e_2, \ldots, e_7\}$ performed on patients with HCC. Each examination results in a set of dynamic contrast-enhanced CT scans including $I_{\texttt{WI}}$, $I_{\texttt{AR}}$ and $I_{\texttt{EV}}$ with additional $I_{\texttt{LV}}$ for e_1 and e_2. For each e_i, 6 equally reparted 2D axial slices have been selected in $I_{\texttt{EV}}$ to cover the tumor spatial extent and labeled by 4 experts in hepato-digestive surgery to reach groundtruth (GT) segmentation masks delimitating parenchyma, active and necrotic tissues. Manual segmentation is performed on temporally averaged images over all phases. It results in a database of 42 slices with associated fused GT masks obtained by fusing all the expert annotations with STAPLE [14]. This database has been created since, to our knowledge, no such data is freely available.

Each examination e_i is processed independently. For each one, we extract among all the 3D SLIC clusters those overlapping the annotated 2D axial slices. Among these clusters, we identify those whose intersection with annotated slices has a predominant GT label (consensus over at least 95% of voxels). This predominant label is assigned to the 3D cluster as its own GT label. Finally, 1/3 of these labeled 3D SLIC clusters are randomly selected (to simulate user interaction) and used to train our *multi-phase cluster-wise* random forest (*MpCl*-RF). Once trained, the forest is employed to classify all the remaining 3D SLIC clusters. We provide comparisons with *single-phase voxel-wise* (*SpVx*), *single-phase cluster-wise* (*SpCl*) and *multi-phase voxel-wise* (*MpVx*) random forest (RF). To

make comparisons possible, training for *Sp/MpVx* focus on voxels belonging to clusters selected for training in *Sp/MpCl*. Each forest contains $T = 100$ trees.

Features differ for each method. Cluster-wise strategies assign to clusters the features of Fig. 3 for *MpCl* and a modified set for *SpCl* since multi-phase features are ousted and spatial intensity and gradient ones only focus on $I_{\texttt{EV}}$. In voxel-wise, *MpVx* assigns to voxels intrisic and spatially averaged (3^3 windows) intensity and gradient at each phase as well as PE, AUC, $\Delta_{EV/AR}$ and $\Delta_{LV/EV}$ whereas *SpVx* only involves intrisic and averaged intensity and gradient in $I_{\texttt{EV}}$.

SpVx, *SpCl*, *MpVx* and *MpCl* are quantitatively assessed via TN rate error $\Delta\tau = |\tau - \tau_{GT}|$ comparing estimated and GT TN rates and *DICE* coefficients between obtained and fused GT masks for parenchyma, active and necrotic tissues: $DICE_{\texttt{prcm}}$, $DICE_{\texttt{activ}}$ and $DICE_{\texttt{necro}}$. To be less sensitive to the variability due to RF random aspects, results for each e_i are averaged over 10 realizations.

Discussion. We present in Tab. 1 a comparative assessment of the 4 methods through $\Delta\tau$ and *DICE* coefficients averaged over the whole database. Segmentation results are displayed on Fig. 4 with corresponding fused GT mask (one slice per examination). The comparative study reveals better results using *MpCl*-RF. It obtains the smallest $\Delta\tau$ with 5.26 and the highest *DICE* for parenchyma, active and necrotic areas with 74.4, 71.9 and 93.3. In comparison, *MpVx*-RF reaches equivalent $DICE_{\texttt{necro}}$ but less accurate $\Delta\tau$, $DICE_{\texttt{activ}}$ and $DICE_{\texttt{prcm}}$. According to visual results (Fig. 4), *MpVx*-RF allows an obvious worst spatial consistency while requiring more interaction efforts compared to *MpCl*-RF.

The comparison between single and multi-phase approaches confirms that exploiting multi-phase CT images instead of one single scan significantly improves the results with gains of *DICE* around 8.6 (11.2), 8.1 (6.8) and 3.6 (6.9) resp. for cluster and voxel-based strategies. The significant differences between *Sp/MpCl*-RF tend to justify the use of robust multi-phase features when making RF working on clusters since a classification error on one single cluster can have a strong impact on the accuracy. In addition, the averaged maximum inter-expert variability computed over the GT masks provided by the 4 experts is 21.2 ± 15.2 in TN rate. This significant variability reinforces the necessity to work toward

Table 1. Quantitative comparisons of *single-phase voxel-wise* (*SpVx*), *single-phase cluster-wise* (*SpCl*), *multi-phase voxel-wise* (*MpVx*) and the proposed *multi-phase cluster-wise* (*MpCl*) random forest (RF) through TN rate error and *DICE* coefficients averaged over the whole database. Best results are emphasized in bold.

methods	*SpVx*-RF	*SpCl*-RF	*MpVx*-RF	***MpCl*-RF**
$\Delta\tau$	6.40 ± 2.85	9.13 ± 4.78	6.60 ± 3.32	$\mathbf{5.26} \pm 3.90$
$DICE_{\texttt{activ}}$	54.3 ± 17.2	65.8 ± 15.3	65.5 ± 12.4	$\mathbf{74.4} \pm 12.6$
$DICE_{\texttt{necro}}$	65.0 ± 21.6	63.8 ± 25.8	71.8 ± 17.6	$\mathbf{71.9} \pm 19.5$
$DICE_{\texttt{prcm}}$	80.5 ± 13.1	89.7 ± 4.90	87.4 ± 9.00	$\mathbf{93.3} \pm 3.08$

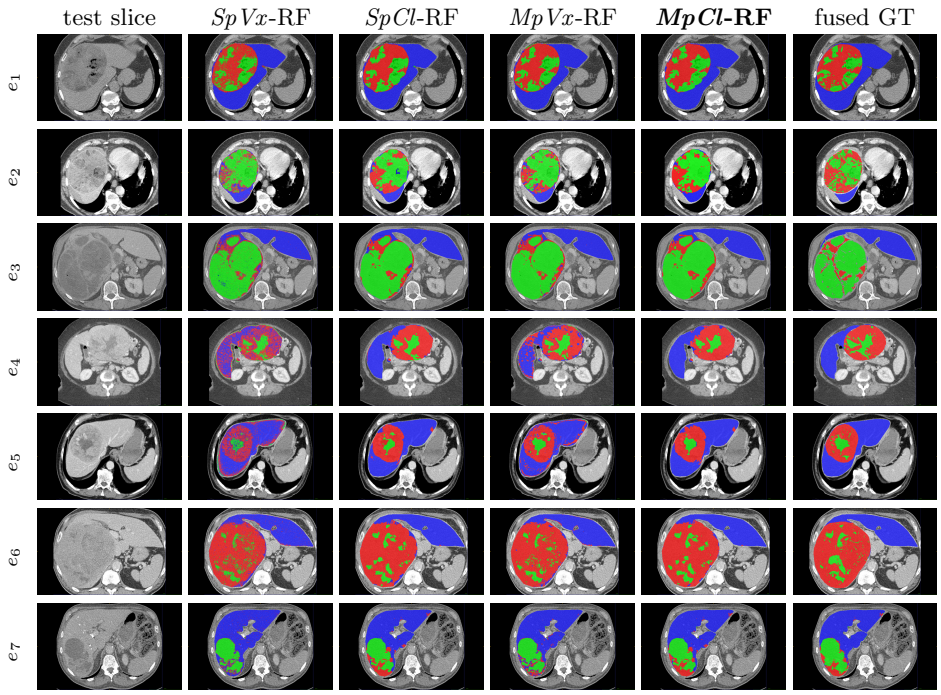


Fig. 4. HCC liver tumor segmentation results via *single-phase voxel-wise* (*SpVx*), *single-phase cluster-wise* (*SpCl*), *multi-phase voxel-wise* (*MpVx*) and the proposed *multi-phase cluster-wise* (*MpCl*) random forest (RF) with groundtruth (GT) masks. Parenchyma, active and necrotic areas are respectively in blue, red and green.

an efficient computed-aided diagnosis method dedicated to TN rate estimation in order to overcome the bias induced by visual qualitative assessment.

4 Conclusion

In this work, we addressed the semi-automatic evaluation of pre-operative locoregional treatments (PLT) response for hepato-cellular carcinoma (HCC). Toward this goal, we proposed a method that estimates the tumor necrosis (TN) rate from dynamic contrast-enhanced CT scans. While ensuring weak interaction efforts for practitioners, it accurately segments parenchyma, active and necrotic tissues. Our approach applies random forest on supervoxels and involves robust multi-phase cluster-wise features. Quantitative assessment on clinical data confirms the benefits of exploiting dynamic information extracted from multi-phase images at a cluster spatial extent. Multi-examination learning would deserve further investigation to make the proposed strategy becoming fully automatic. More generally, it could be easily applied to other tumor types, organs and modalities.

Acknowledgments. This work received the financial support from Fondation Arc, `www.fondation-arc.org`.

References

1. Ronot, M., Bouattour, M., Wassermann, J., Bruno, O., Dreyer, C., Larroque, B., Castera, L., Vilgrain, V., Belghiti, J., Raymond, E., et al.: Alternative response criteria (Choi, EASL and mRECIST) versus RECIST1.1 in patients with advanced hepatocellular carcinoma treated with *Sorafenib*. The Oncologist (2014)
2. Ronot, M., Vilgrain, V.: Hepatocellular carcinoma: Diagnostic criteria by imaging techniques. Best Practice & Research Clinical Gastro-enterology **28**(5) (2014)
3. Lee, J., Cai, W., Singh, A., Yoshida, H.: Estimation of necrosis volumes in focal liver lesions based on multi-phase hepatic CT images. In: Virtual Colonoscopy & Abdominal Imaging. Computational Challenges & Clinical Opportunities. (2011)
4. Raj, A., Juluru, K.: Visualization and segmentation of liver tumors using dynamic contrast MRI. In: Conference of Engineering in Medicine and Biology. (2009)
5. Fang, R., Zabih, R., Raj, A., Chen, T.: Segmentation of liver tumor using efficient global optimal tree metrics graph cuts. In: Abdominal Imaging. Computational and Clinical Applications. Springer (2012) 51–59
6. Shimizu, A., Narihira, T., Furukawa, D., Kobatake, H., Nawano, S., Shinozaki, K.: Ensemble segmentation using Adaboost with application to liver lesion extraction from a CT volume. In: Workshop on 3D Segmentation in the Clinic. (2008)
7. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for ms lesion segmentation in multi-channel mr images. In: Medical Image Computing and Computer-Assisted Intervention. (2010)
8. Breiman, L.: Random Forests. Machine learning **45**(1) (2001) 5–32
9. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends in Computer Graphics and Vision **7**(2–3) (2012)
10. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic detection and segmentation of kidneys in 3D CT images using random forests. In: Medical Image Computing and Computer-Assisted Intervention. (2012)
11. Bauer, S., Nolte, L.P., Reyes, M.: Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Medical Image Computing and Computer-Assisted Intervention. (2011) 354–361
12. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision (2005)
13. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(11) (2012) 2274–2282
14. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging **23**(7) (2004) 903–921