Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model

E. Geremia[1], B. H. Menze[1,2,3], M. Prastawa[4], M.-A. Weber[5], A. Criminisi[6], N. Ayache[1]

[1] Asclepios Research Project, INRIA Sophia-Antipolis, France.
[2] Computer Science and Artificial Intelligence Laboratory, MIT, USA.
[3] Computer Vision Laboratory, ETH Zurich, Switzerland.
[4] Scientific Computing and Imaging Institute, University of Utah, USA.
[5] Diagnostic and Interventional Radiology, Heidelberg University Hospital, Germany.
[6] Machine Learning and Perception Group, Microsoft Research Cambridge, UK.

Abstract. This paper proposes to employ a detailed tumor growth model to synthesize labelled images which can then be used to train an efficient data-driven machine learning tumor predictor. Our MR image synthesis step generates images with both healthy tissues as well as various tumoral tissue types. Subsequently, a discriminative algorithm based on random regression forests is trained on the simulated ground truth to predict the continuous latent tumor cell density, and the discrete tissue class associated with each voxel. The presented method makes use of a large synthetic dataset of 740 simulated cases for training and evaluation. A quantitative evaluation on 14 real clinical cases diagnosed with low-grade gliomas demonstrates tissue class accuracy comparable with state of the art, with added benefit in terms of computational efficiency and the ability to estimate tumor cell density as a latent variable underlying the multimodal image observations. The idea of synthesizing training data to train data-driven learning algorithms can be extended to other applications where expert annotation is lacking or expensive.

1 Introduction

Brain tumors are complex patho-physiological processes representing a series of pathological changes to brain tissue [1]. Increasing effort is invested in modelling the underlying biological processes involved in brain tumor growth [2, 3]. As brain tumors show a large variety of different appearances in multi-modal clinical images, the accurate diagnosis and analysis of these images remains a significant challenge. We show in the example of gliomas, the most frequent brain tumor [4], how a *generative* patho-physiological model of tumor growth can be used in conjunction with a *discriminative* tumor recognition algorithm, based on random regression forests. Applied to real data the random forest is capable of predicting the precise location of the tumor and its substructures. In addition, our model can also infer the spatial distribution of (unobservable)

latent physiological features such as tumor cell densities, thus avoiding the need for expensive patho-physiological model inversion [5].

Generative probabilistic segmentation models of spatial tissue distribution and appearance proved to generalize well to previously unseen images [6–9]. In [6], tumors are modeled as outliers relative to the expected appearance of healthy tissue following a related approach for MS lesion detection [10]. Other methods [7, 8] provide explicit models for the tumor class. For instance, [8] builds a tumor appearance model for channel specific segmentation of the tumor, combining a tissue appearance model with a latent tumor class prior from [9]. Tumor growth models (*e.g.* reaction-diffusion models) have been used repeatedly to improve image registration [11] and, hence, atlas-based tissue segmentation [12]. Similarly, [13] relies on a bio-mechanical tumor growth model to estimate brain tissue loss and displacement. Generative approaches require a detailed formal description of the image generation process and may need considerable modifications when applied to slightly different tasks. These approaches also tend to be computationally inefficient.

In contrast, discriminative techniques focus on modeling the difference between *e.g.* a lesion and healthy tissues, directly [14–16]. A number of recent techniques based on decision tree ensembles have shown strong generalization capabilities and computational efficiency, even when applied to large data sets [17–19]. In [20], for example, a *classification* forest is used for segmenting multiple sclerosis lesions using long-range spatial features. In [15], the authors derived a constrained minimization problem suitable for min-cut optimization that incorporates an observation model provided by a discriminative Probabilistic Boosting Trees classifier into the process of segmentation. For multi-modal brain lesion segmentation, [16] propose a hierarchical segmentation framework by weighted aggregation with generic local image features. Unfortunately, fully supervised discriminative approaches may require large expert-annotated training sets. Obtaining such data is often prohibitive in many clinical applications.

This paper proposes a new way of combining the best of the generative and discriminative world. We use a generative model of glioma [21] to synthesize a large set of heterogeneous MR images complete with their ground truth annotations. Such images are then used to train a multi-variate *regression* forest tumor predictor [20, 22]. Thus given a previously unseen image the forest can perform an efficient, per-voxel estimation of both tumor infiltration density *and* tissue type. The general idea of training a discriminative predictor (a classifier or a regressor) on a large collection of synthetic training data is inspired by the recent success of the Microsoft Kinect for XBox 360 system [23]. This approach has great potential in different domains and especially for medical applications where obtaining extensive expert-labelled is nearly impossible.

2 Learning to estimate tissue cell density from synthetic training data

This section describes the two basic steps of our algorithm: i) synthesizing heterogeneous MR images showing tumors, and ii) training a tumor detector which works on *real* patient images.

2.1 Generative tumor simulation model

The automatic generation of our synthetic training dataset relies on the publicly available brain tumor simulator presented in [21]. It builds on an anisotropic glioma growth model [24] with extensions to model the induced mass-effect and the accumulation of the contrast agents in both blood vessels and active tumor regions. Then, multi-sequence MR images are synthesized using characteristic image textures for healthy and pathological tissue classes (Fig. **??**).

We generate synthetic pathological cases with varying tumor location, tumor count, levels of tumor expansion and extent of edema. The resulting synthetic cases successfully reproduce mass-effect, contrast enhancement and infiltration patterns similar to what observed in the real cases. The synthetic dataset contains 740 synthetic cases. It includes a large variability of brain tumors ranging from very diffusive tumors, showing a large edema-infiltration pattern without necrotic core, to bulky tumors with a large necrotic core surrounded by an enhanced vascularization pattern. For each case, the simulation provides four MR sequences (cf. Fig. 1) which offer different views of the same underlying tumor density distribution.

This synthetic ground truth provides a diverse view of the pathological process including mass-effect and infiltration, but also very detailed annotations for the healthy structures of the brain. The ground truth consists of voxel-wise annotations on the data that are: white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), edema, necrotic tumor core, active tumor rim and blood vessels. Unlike binary annotations which provide a mask for each tissue class, the ground truth consists of a continuous scalar map for each tissue class. Each scalar map provides, for every voxel in the volume, the density of every tissue class.

2.2 Regression forests for estimating tissue cell density

Problem setting We adapt a regression forests similar to the one of [17] to train an estimator of tissue cell densities from visual cues in the multi-channel MR images. For each voxel $\mathbf{v}$, the ground truth provides the density $R_c(\mathbf{v}) \in [0, 1]$ of each tissue class $c \in \mathcal{C}$. The density distribution R is normalized so that it satisfies $\sum_{c \in \mathcal{C}} R_c(\mathbf{v}) = 1$ in every voxel $\mathbf{v}$.

Feature representation To calculate the local image features – both during training and for predictions – we sub-sample or interpolate all images to $1 \times 1 \times 2$

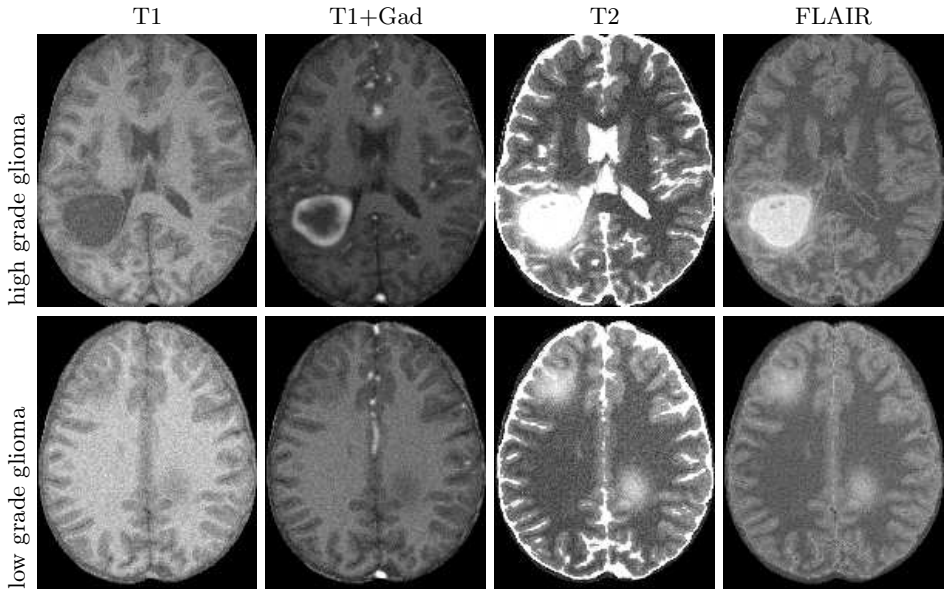

Fig. 1. Synthetic MR images. From left to right: T1, T1+Gad, T2, and FLAIR MR images. Top row: bulky tumor characterized by a large necrotic and a surrounding vascularization pattern. Bottom row: very infiltrating tumor characterized by the extended of the edema.

mm^3 resolution. We perform a skull-stripping and an intensity normalization [25] so that real MR images match the intensity distribution of synthetic MR sequences. Then image features are calculated for each voxel $\mathbf{v}$. Features include local multi-channel intensity (T1, T1+Gad, T2, Flair) as well as long-range displaced box features such as in [20]. In addition we also incorporate symmetry features, calculated after estimating the mid-sagittal plane [26]. In total, every voxel is associated with a 213−long vector of feature responses.

Regression forest training The forest consists of T trees indexed by t. During training observations of all voxels $\mathbf{v}$ are pushed through each of the trees. Each internal node p applies a binary test $t_p = \tau_{low} \leq \theta(\mathbf{v}) < \tau_{up}$ implementing a double thresholding (τ_{low}, τ_{up}) of the visual feature $\theta(\mathbf{v})$ evaluated at voxel $\mathbf{v}$. The voxel $\mathbf{v}$ is then sent to one of the two child nodes based on the outcome of this test. As a result, each node p receives a partition of the input training data $\mathcal{T}_p = \{\mathbf{v}, R(\mathbf{v})\}_p$, composed of a voxel $\mathbf{v}$ and a vector $R(\mathbf{v}) \in [0,1]^{|\mathcal{C}|}$ storing the cell density value for each tissue class. We model the resulting distribution via a multi-variate Gaussian $\mathcal{N}_p(\mu_p, \Gamma_p)$ where μ_p and Γ_p are the mean and covariance matrix of all $R(\mathbf{v}) \in \mathcal{T}_p$, respectively. During training, the parameters (τ_{low}, τ_{up}) of the node test function and the employed visual feature θ are optimized to maximize the information gain. We define the information gain $IG(t_p)$

to measure the quality of the test function t_p which splits $\mathcal{T}_p$ into $\mathcal{T}_p^{left}$ and $\mathcal{T}_p^{right}$. The information gain is defined as $IG(t_p) = -\sum_{k\in\{left,right\}} \omega_k log\rho_k$ with $\omega = |\mathcal{T}_p^k|/|\mathcal{T}_p|$ and $\rho_k = max|eig(\Gamma_k)|$ where eig denotes all matrix eigenvalues. In contrast to the more conventional information gain used in [17], our formulation gives a robust estimate of the dispersion. Indeed, the information gain presented in [17] models the dispersions as $|\Gamma_k|$ which evaluates to 0 in the case a tissue class is missing from the input partition $\mathcal{T}_p$. Our definition of the information gain focuses on the direction showing maximum dispersion, i.e. ρ_k, and ignores the missing information on tissue classes.

At each node p, the optimal test $t_p^* = \arg\max_\Lambda IG(t_p)$ is found by exhaustive search over a random subset of the feature space $\Lambda = \{\tau_{low}, \tau_{up}, \theta\}$. Maximizing the information gain encourages minimizing ρ_p, thus decreasing the prediction error when approximating $\mathcal{T}_p$ with $\mathcal{N}_p$. The trees are grown to a maximum depth D, as long as $|\mathcal{T}_p| > 100$.

After training, the random forest embeds a hierarchical piece-wise Gaussian model which captures the multi-modality of the training data. In fact, each leaf node l_t of every tree t stores the Gaussian distribution $\mathcal{N}_{l_t}$ associated with the partition of the training data arrived at that leaf $\mathcal{T}_{l_t}$.

The employed random regression forest approximates the multi-variate distribution R by a piece-wise Gaussian distribution $\hat{R}$.

Regression forest prediction When applied to a previously unseen test volume $\mathcal{T}_{test} = \{\mathbf{v}\}$, each voxel $\mathbf{v}$ is propagated through all the trained trees by successive application of the relevant binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, estimated cell densities $r_t(\mathbf{v}) = \mu_{l_t}$ are averaged together to compute the forest tissue cell density estimation $r(\mathbf{v}) = (\sum_{t\in[1..T]} r_t(\mathbf{v}))/T$. Note that in each leaf l_t we maintain an estimate of the confidence Γ_{l_t} associated to the cell density estimation μ_{l_t}.

3 Experiments

We conducted two main experiments. First, as a proof of concept, we tested how well the learned forest reproduces the tissue cell densities in the synthetic model. In a second experiment we applied our method to real, previously unseen, clinical images and measured accuracy by comparing the detected and ground truth tumor outlines.

We evaluate the predictions for every test case using two complementary metrics: a segmentation metric and a robust regression metric. The segmentation metric compares binary versions of the physiological maps, independently normalized for each tissue class. The binary masks are obtained by thresholding the prediction and the ground truth at the same value. Then, we evaluate the true postive rate $TPR = TP/(TP + FN)$, the false positive rate $FPR = TP/(TP+FP)$ and the positive predictive value $PPV = TP/(TP+FP)$, where TP, FP, and FN are the number of true positives, false positives, and false neg-

atives, respectively. Finally, we compute the area under the ROC and the one precision-recall curves to measure how well the prediction fits the ground truth.

The robust regression metric evaluates the estimation error between the predicted continuous map and the ground truth. For every tissue class c, we compute the mean over the voxels v of the estimation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$. In order to avoid artificial decrease of the mean error, we make this metric robust by only considering regions of the physiological map showing at least 10% signal in either the prediction or the ground truth.

In both experiments, we used the same forest containing $T = 160$ trees of depth $D = 20$ trained on 500 synthetic cases. The values of these meta-parameters were tuned by training and testing on a different synthetic set.

3.1 Experiment 1: Estimating cell density in synthetic cases

We tested the random forest on a previously unseen synthetic dataset with 240 cases. Results (Fig. 2) show a good qualitative match between predicted and ground truth physiological maps. As a segmentation metric we calculate the true and false positive rates as well as the positive predictive value for each possible threshold jointly on r and R and summarize it through ROC and precision-recall curves. For every tissue class c, we also compute the mean approximation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$ (integrating over voxels with $> .001$ tumor cell density for tumor classes). Results in Fig. 3 show excellent results for WM, GM, CSF. The predicted tumor cell density is in good agreement with ground truth. A systematic bias leads to a slightly larger variance in the error metric due to the small size of the tumor classes compared to the healthy tissue classes.

3.2 Experiment 2: Segmenting tumors in clinical images

We tested the same random forest on 14 clinical cases showing low and high grade glioma (Fig. 5) with T1, T1+Gad, T2 and FLAIR images. None of the clinical cases was used during training. Training was done exclusively on synthetic images. The manually-obtained ground truth consists of a binary tumor mask delineating the tumor+edema region. We calculated the same tumor outline from the predicted continuous physiological masks as done for the synthetic model [21]. Segmentation results (Fig. 4) are in excellent agreement with a state-of-the-art unsupervised multimodal brain tumor segmentation method that also outperformed standard EM segmentation in an earlier study [8]. Note that the method presented in [8] significantly outperformed [6]. Interestingly, in a qualitative evaluation (cf. Fig. 5), the tumor cell density map shows smooth transition between the active rim of the tumor (red) and the edema (green).

4 Conclusions

This paper presented a new generative-discriminative algorithm for the automatic detection of glioma tumors in multi-modal MR brain images. A regression

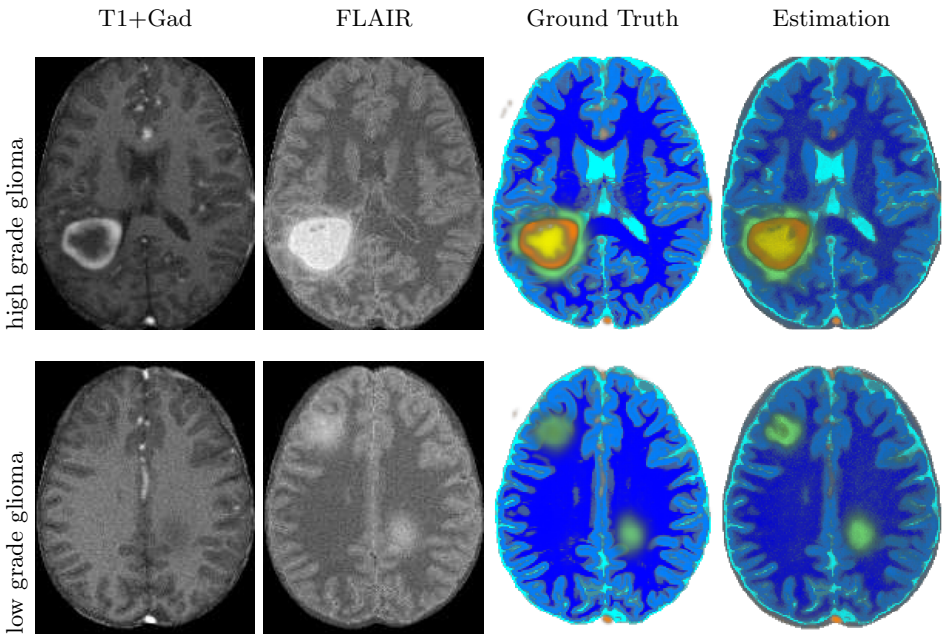

Fig. 2. Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels (orange), and necrotic core (yellow).

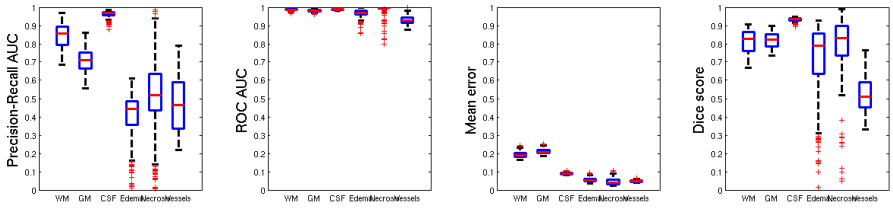

Fig. 3. Evaluation of the predictions on the synthetic dataset for each cell density map . Each label in the x-axis represents a tissue class: WM, GM, CSF, edema, necrotic core, blood vessels, respectively. We show from left to right: the area under the precision-recall curve, the area under the ROC curve, the estimation of the mean prediction error, and the dice score Each point of the ROC and precision-recall curves is built by thresholding the prediction and the ground truth at the same value. The ground truth and the prediction density maps were thresholded at the same value, i.e. 0.3.

forest model was trained on multiple synthetically-generated labelled images.

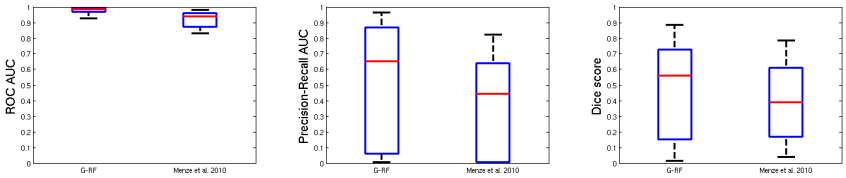

Fig. 4. Evaluation of the predictions on the clinical dataset. Box plots of the area under the ROC curve (left), under the precision-recall curve (right), and the dice score. Comparison of the proposed method (G-RF) with the method presented in [8].

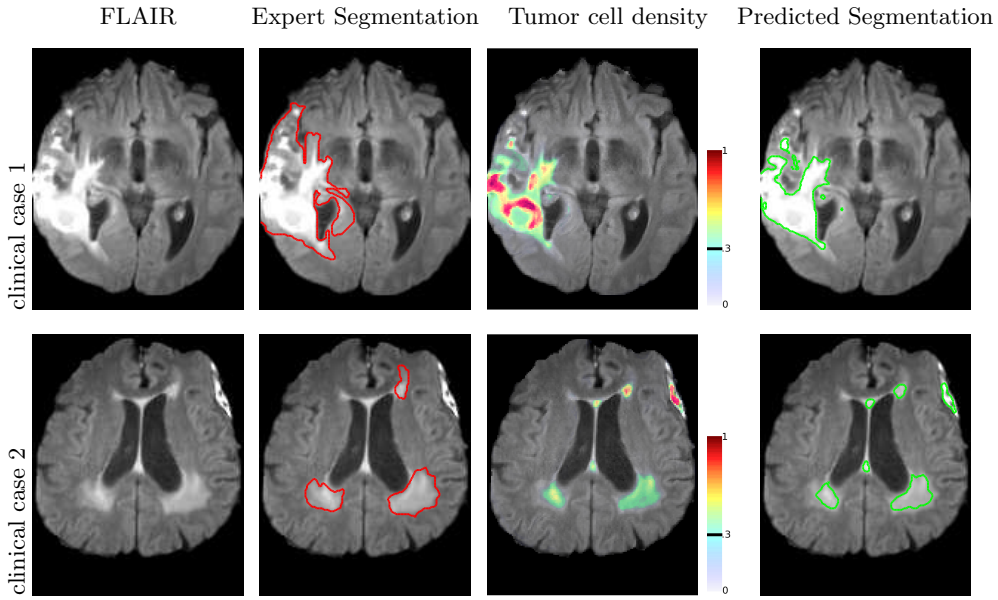

Fig. 5. Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).

Then the system demonstrated to work accurately on previously unseen synthetic cases. It showed promising results on real patient images which led to state of the art tumor segmentation results. Our algorithm can estimate continuous tissue cell densities both for healthy tissues (WM, GM, CSF) as well as tumoral ones.

Acknowledgments

This work was partially funded by the ERC MedYMA grant. We would like to thank Marc-André Weber from the Diagnostic and Interventional Radiology Group in Heidelberg University Hospital, Germany for providing us with the clinical data.

References

1. Angelini, E.D., Delon, J., Bah, A.B., Capelle, L., Mandonnet, E.: Differential MRI analysis for quantification of low grade glioma growth. Medical Image Analysis **16** (2012) 114–126
2. Cristini, V., Lowengrub, J. In: Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach. Cambridge University Press (2010) 185–205
3. Deisboeck, T.S., Stamatakos, G.S. In: Multiscale Cancer Modeling. CRC Press (2010) 359–406
4. Angelini, E.D., Clatz, O., M, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: A review of segmentation, registration, and in silico growth algorithms and their clinical applications. Cur Med Imag Rev **3** (2007) 262–276
5. Menze, B.H., Van Leemput, K., Honkela, A., Konukoglu, E., Weber, M.A., Ayache, N., Golland, P.: A generative approach for image-based modeling of tumor growth. In: Proc IPMI. (2011)
6. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical Image Analysis **8** (2004) 275–283
7. Zou, K.H., III, W.M.W., Kaus, M., Kikinis, R., Jolesz, F.A., Warfield, S.K.: Statistical Validation of Automated Probabilistic Segmentation against Composite Latent Expert Ground Truth in MR Imaging of Brain Tumors. In: Proc MICCAI. (2002) 315–322
8. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Proc MICCAI. 6362, Springer (2010) 151–159
9. Riklin-Raviv, T., Leemput, K.V., Menze, B.H., Wells, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Medical Image Analysis **14** (2010) 654–665
10. Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A.C.F., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging **20**(8) (2001) 677–688
11. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., Cuadra, M.B.: A review of atlas-based segmentation for magnetic resonance brain images. Comp Meth Prog Biomed **104** (2011) 158–164
12. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: Proc MICCAI. LNCS 6891, Springer (2011) 532–540
13. Zacharaki, E.I., Hogea, C.S., Shen, D., Biros, G., Davatzikos, C.: Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. NeuroImage **46** (2009) 762–774
14. Lee, C.H., Wang, S., Murtha, A., Brown, M.R.G., Greiner, R.: Segmenting brain tumors using pseudo-conditional random fields. In: Proc MICCAI. LNCS 5242, Springer (2008) 359–366

15. Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D.: A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI. In: Proc MICCAI. (2008) 67–75
16. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.L.: Efficient multilevel brain tumor segmentation with integrated bayesian model classification. IEEE Transactions on Medical Imaging **27** (2008) 629–640
17. Criminisi, A., Shotton, J., Robertson, D.P., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Proc MICCAI-MCV. LNCS 6533, Springer (2010) 106–117
18. Montillo, A., Shotton, J., Winn, J.M., Iglesias, J.E., Metaxas, D.N., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of CT images. In: Proc IPMI. (2011) 184–196
19. Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D.: Random forest-based manifold learning for classification of imaging data in dementia. In: Proc MICCAI-MLMI. LNCS 7009, Springer (2011) 159–166
20. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images. NeuroImage **57** (2011) 378–90
21. Prastawa, M., Bullitt, E., Gerig, G.: Simulation of brain tumors in MR images for evaluation of segmentation efficacy. Medical Image Analysis **13** (2009) 297–311
22. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical report, Microsoft (2011)
23. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proc CVPR. (2011) 1297–1304
24. Clatz, O., Sermesant, M., Bondiau, P.Y., Delingette, H., Warfield, S.K., Malandain, G., Ayache, N.: Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with mass effect. IEEE Transactions on Medical Imaging **24** (2005) 1334–1346
25. Coltuc, D., Bolon, P., Chassery, J.M.: Exact histogram specification. IEEE TIP **15** (2006) 1143 –1152
26. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3d brain images. IEEE Transactions on Medical Imaging **21** (2002) 122–138