

Integrated Spatio-Temporal Segmentation of Longitudinal Brain Tumor Imaging Studies

Stefan Bauer¹(✉), Jean Tessier², Oliver Krieter³, Lutz-P. Nolte¹,
and Mauricio Reyes¹

¹ ISTB, University of Bern, Bern, Switzerland

² F. Hoffmann-La Roche Ltd., Basel, Switzerland

³ Roche Diagnostics GmbH, Penzberg, Germany

stefan.bauer@istb.unibe.ch

Abstract. Consistent longitudinal segmentation of brain tumor images is a critical issue in treatment monitoring and in clinical trials. Fully automatic segmentation methods are a good candidate for reliably detecting changes of tumor volume over time. We propose an integrated 4D spatio-temporal brain tumor segmentation method, which combines supervised classification with conditional random field regularization in an energy minimization scheme. Promising results and improvements over classic 3D methods for monitoring the temporal volumetric evolution of necrotic, active and edema tumor compartments are demonstrated on a longitudinal dataset of glioma patient images from a multi-center clinical trial. Thanks to its speed and simplicity the approach is a good candidate for standard clinical use.

Keywords: Brain tumor · Glioma · Longitudinal studies · Segmentation · Volumetric analysis

1 Introduction

The segmentation of brain tumor images is an important clinical problem. It is necessary for patient monitoring and treatment planning, but it also has applications in clinical drug trials [9], where tumor response to therapy needs to be assessed. Many different automatic or semi-automatic segmentation algorithms have been proposed [5] and while their performance might be still debatable, it is well-accepted that manual segmentations are subject to high intra- and inter-rater variability [12]. This variability is even more influential when analysing longitudinal patient studies where tumor progression or regression should be monitored. In this case, it is more important to have an objective tumor segmentation, which can correctly identify changes over time, than having a very high accuracy at single time points. The RANO (response assessment in neuro-oncology) working group has pointed out, that in addition to the currently applied 2D manual diameter measurements for monitoring tumor growth, in the future it would be desirable to have reliable 3D measurements of volumetric

tumor change [15]. Fully automatic segmentation algorithms are ideally suited for this scenario because they allow for an objective longitudinal assessment of tumor development. Furthermore they allow for an efficient handling of the large multi-modal datasets that are generated in longitudinal studies.

Despite the suitability of automatic methods for longitudinal studies, so far most algorithms for brain tumor segmentation have only been applied to images taken at single time points. Obviously, a standard segmentation algorithm for brain tumor images at single time points (e.g. [4, 6, 8, 16]) could be applied for longitudinal studies, however this would not make use of the full temporal information, thus possibly decreasing robustness. There are only few methods, specifically designed for assessing temporal changes in brain tumor images, which mostly target slowly evolving low-grade gliomas. Konukoglu et al. [11] used a semi-automatic approach based on image registration for change detection. Pohl et al. [13] followed a similar idea, performing semi-automatic segmentation in combination with registration, after which they analyzed local intensity patterns to detect tumor growth. Angelini et al. [2] implemented a histogram mapping, which allowed them to compare intensity difference maps directly after affine registration.

Due to the irregular appearance and fast evolution of high-grade gliomas, registration methods are not well suited for this scenario. We chose a different approach for longitudinal brain tumor segmentation, which is based on supervised classification with integrated 4D spatio-temporal regularization for longitudinal brain tumor segmentation. It offers the possibility to segment tumor and healthy tissues including their subcompartments (necrotic, active, edema region and cerebrospinal fluid (CSF), gray matter (GM), white matter (WM) respectively).

2 Methods

The task is modeled as an energy minimization problem in a spatio-temporal conditional random field (CRF) formulation, where the random field contains cliques with both spatial and temporal links. The integration of spatial and temporal links was conceptually inspired by a work on video segmentation [14]. The energy consists of the sum of singleton potentials and pairwise potentials, which can be seen in the first and second term of Eq. (1), where i and j determine the voxel position in space and time. The optimization problem is solved to yield a segmentation result based on fast linear programming strategies [10].

$$E = \sum_i V(y_i, \mathbf{x}_i) + \sum_{ij} W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) \quad (1)$$

The singleton potentials $V(y_i, \mathbf{x}_i)$ are computed according to Eq. (2), where y_i is the final label output, \tilde{y}_i is the probability function learned from a discriminative classifier, \mathbf{x}_i is the feature vector and δ is the Kronecker- δ function. For the classifier, a 40-dimensional feature vector \mathbf{x}_i is used. It combines the normalized multi-modal intensities with first order textures (mean, variance,

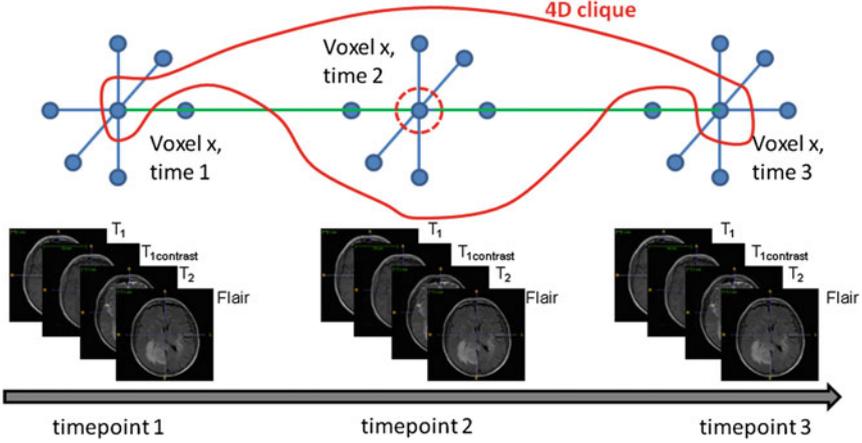


Fig. 1. The regularization is based on 4D spatio-temporal cliques for each voxel. The clique contains information from the local neighborhood of all image modalities at one time point, plus the previous and the subsequent time point.

skewness, kurtosis, energy, entropy) from local patches and statistics of gradient-based intensity differences in a local neighborhood.

$$V(y_i, \mathbf{x}_i) = p(\tilde{y}_i | \mathbf{x}_i) \cdot (1 - \delta(\tilde{y}_i, y_i)) \quad (2)$$

A decision forest classifier is employed because it can efficiently handle multi-label problems and provide posterior probabilities $p(\tilde{y}_i | \mathbf{x}_i)$ as an output [7]. This probabilistic output can be used as a weighting factor in Eq. (2), which allows us to control the degree of regularization depending on the confidence of the classification output.

The pairwise potentials $W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j)$ described in Eq. (3) account for the spatio-temporal regularization. In contrast to standard random field approaches, the cliques do not only model 3D spatial relationships, but 4D spatio-temporal relationships between image voxels, where each time frame is connected to the previous and the subsequent frame as illustrated in Fig. 1.

$$W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) = w_s(i, j) \cdot (1 - \delta(y_i, y_j)) \cdot \exp\left(\frac{-\text{PCD}(\mathbf{x}_i, \mathbf{x}_j)}{2 \cdot \bar{x}}\right) \cdot D(y_i, y_j) \quad (3)$$

In Eq. (3), $w_s(i, j)$ is a weighting function that depends on the “spacing” in each dimension. This means the spatial resolution is taken into account, whereas for the temporal dimension a uniform spacing is chosen because depending on the treatment plan, changes in tumor volume do not necessarily depend on the imaging interval. Different labels of adjacent voxels are penalized by the term $(1 - \delta(y_i, y_j))$, whereas the degree of smoothing is regulated based on the local intensity variation, computed as $\exp\left(\frac{-\text{PCD}(\mathbf{x}_i, \mathbf{x}_j)}{2 \cdot \bar{x}}\right)$ with $\text{PCD}(\mathbf{x}_i, \mathbf{x}_j)$ being a

pseudo-Chebyshev distance and \bar{x} being the respective generalized mean intensity of the relevant modalities. The use of pseudo-Chebyshev distance is motivated by the fact that some modalities better describe certain tissues. In this case the $T_{1\text{contrast}}$ and Flair modalities are the most discriminative to distinguish the borders of individual tumor compartments and we make use of these two modalities only in the PCD term. Prior knowledge for penalizing different tissue adjacencies individually is taken into account by the $D(y_i, y_j)$ term in an empirical way. This allows us to stronger penalize adjacencies of tissues which are less likely to occur (e.g. necrotic and healthy tissue adjacencies).

3 Results

The method has been applied on an image dataset of 6 patients from a multi-center phase 1 clinical drug trial with a well-defined clinical image acquisition and drug administration protocol (part of the patients also had surgical resections). For each patient, there are 6 multi-modal MRI scans available at specific time points over a two-month period before and after drug ingestion (36 multi-modal images in total). This contains two baseline scans and four follow-up scans after treatment with anti-angiogenic therapies. Following the current clinical protocol, we operated only on the structural T_1 , $T_{1\text{contrast}}$, T_2 and Flair MR images. The images of each patient were rigidly registered in order to ensure voxel-to-voxel correspondence and automatically skull-stripped in a pre-processing step. In our case, rigid registration was mandatory because we just aimed at aligning the brains to ensure general correspondence, internal tissue deformations caused by tumor growth or shrinkage were handled by the segmentation algorithm itself. Therefore they did not have to be considered by the registration method, otherwise the results of the volume measurements would be compromised. Additionally, bias-field correction, intensity normalization and denoising with an edge-preserving smoothing filter were performed.

We conducted two different experiments to evaluate the performance of the algorithm. First, we compared the proposed integrated 4D spatio-temporal segmentation to an enhanced version of the standard 3D segmentation from [3] (this was among the best performing methods at the MICCAI BraTS 2012 challenge and is basically the same as the approach presented here without considering temporal links in the regularization). The results were analyzed quantitatively by the overlap of the automatic segmentation result with a manually defined ground-truth using the Dice similarity coefficient and the mean surface distance. The Dice coefficient can range from 0 to 1 with 0 indicating no overlap and 1 indicating perfect overlap. The results are presented in Table 1. The average Dice coefficient increased when the spatio-temporal regularization was used. The increase was statistically significant for the active tumor region and partially also for the edema region (see Table 1). The added benefit of the 4D regularization is even more clear when considering the mean surface distance of the individual tumor compartments to the groundtruth. In almost all cases a clear significant improvement could be observed (see Table 1). The 3D segmentation tended to

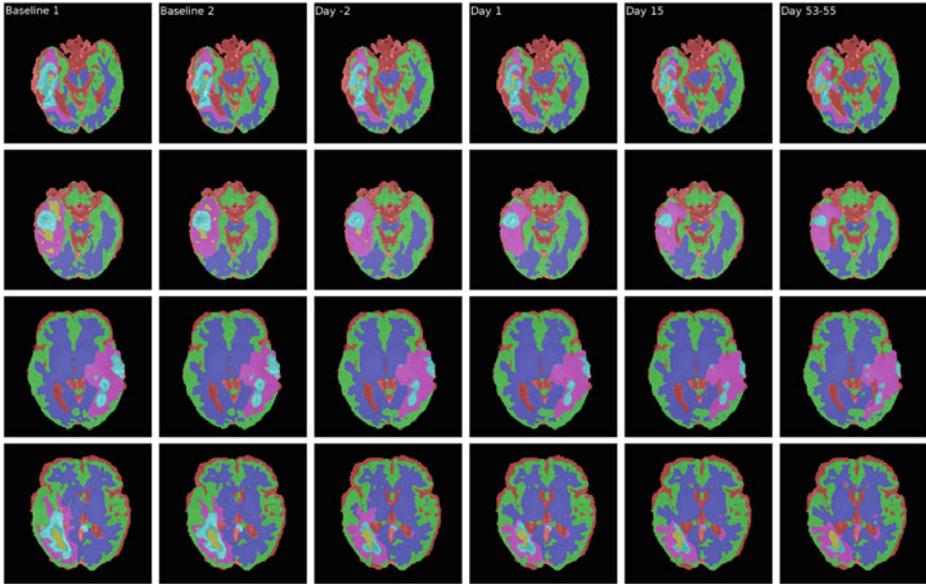


Fig. 2. Segmentation results for patient 2,3,5,6 from top to bottom (color code: red=CSF, green=GM, blue=WM, yellow=necrotic tumor, turquoise=active tumor, pink=edema). Each column shows one time point, starting from the first baseline scan, until the final scan 55 days after the first drug ingestion. (Results for patient 1 and 4 can be found in Figs. 4 and 5.)

produce strong outliers for the segmented volume at some time points. These could be effectively eliminated by the 4D regularization as can also be seen in Fig. 3, where the tumor volume evolution curves are more consistent with the clinical diagnosis for the 4D segmentation than the 3D segmentation. More details about some of the outliers generated by the 3D method in Fig. 3 are shown and further discussed in Figs. 4 and 5

Then, we evaluated the effect of different training datasets for the decision forest. Our standard training data were the training images of high-grade gliomas from the MICCAI 2012 BraTS challenge¹. This is a dataset with a completely different image acquisition protocol than the one used for the longitudinal testing images. We wanted to see if the results improve when training is performed on a rough outline of the tumor on the first baseline image of each patient instead. As expected, it can be observed from Table 1 that the Dice coefficients are always higher when training is performed on the first baseline image of the same patient, but still the algorithm seems to generalize sufficiently well, so that even with training on the completely different BraTS dataset, acceptable results in terms of Dice overlap can be achieved. The rather low Dice coefficients for necrotic tissue can be explained by the fact that the necrotic region is often small

¹ <http://www2.imm.dtu.dk/projects/BRA2012/>

Table 1. Average and standard deviation for overlap and surface distance of individual tumor compartments. Four different cases have been considered: training on BraTS with 3D segmentation, training on BraTS with 4D segmentation, training on the first baseline scan of the same patient with 3D segmentation, training on the first baseline scan of the same patient with 4D segmentation. Cases, where the 4D segmentation yielded a statistically significant improvement ($p < 0.05$) over the 3D segmentation have been marked with an asterisk

	3D BraTS	4D BraTS	3D baseline1	4D baseline1
Dice coefficient				
necrotic	0.18 ± 0.22	0.18 ± 0.21	0.45 ± 0.35	0.45 ± 0.35
active	0.54 ± 0.2	$0.58 \pm 0.18^*$	0.59 ± 0.26	$0.62 \pm 0.24^*$
edema	0.65 ± 0.1	0.66 ± 0.09	0.69 ± 0.13	$0.73 \pm 0.08^*$
Mean surface distance [mm]				
necrotic	15.6 ± 14.9	10.5 ± 9.4	5.5 ± 6.7	$2.1 \pm 2.1^*$
active	4.2 ± 3.2	$2.6 \pm 1.4^{***}$	2.9 ± 3.8	$1.8 \pm 2.9^{**}$
edema	4.7 ± 3.2	$3.5 \pm 1.9^{**}$	4.6 ± 3.7	$2.4 \pm 1.2^{***}$

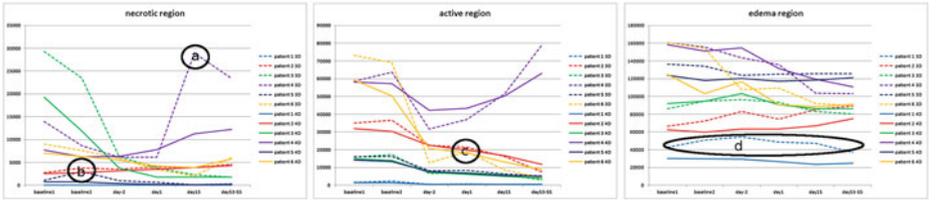


Fig. 3. Trend of the tumor volumes (in mm^3) for necrotic (left), active (center) and edema (right) compartments of all 6 patients. Results are shown with dashed lines for the pure 3D spatial segmentation and with solid lines for the proposed 4D spatio-temporal segmentation. The most prominent outliers of the 3D method are highlighted by black ellipses. Patient 4 showed progressive disease, which could be reliably identified from the computed active tumor region (purple line).

and the Dice coefficient is sensitive to the size of the region. Figure 2 illustrates the segmentation results on an axial slice of four patient images for all time points.

Finally, we compared the trend of the combined necrotic and active tumor volume, which was predicted by the algorithm, to the trend of the gross tumor volume, which had been manually outlined by an expert radiologist on the dynamic contrast enhanced (DCE) images of the same patient (these images were not used for the automatic segmentation method). The trend of the automatically segmented tumor volume is in general agreement with the tumor volumes manually defined on the DCE images, while the absolute values were still showing significant differences (see Fig. 6). The difference in absolute volume could be partially attributed to the fact that the resolution of our structural images was 4 times higher than the resolution of the DCE images, another explanation

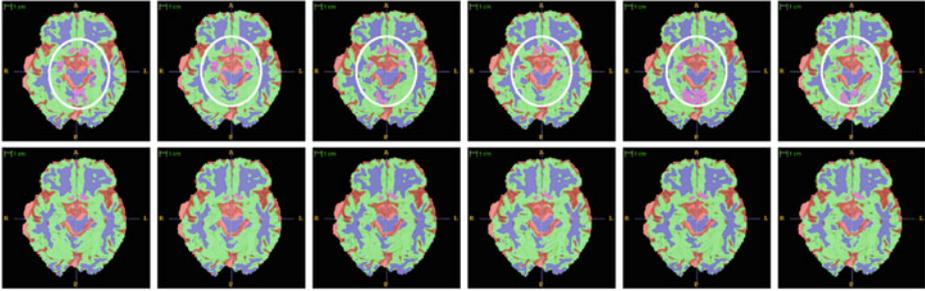


Fig. 4. Illustration of outlier (d) in Fig. 3 (edema region of patient 1). The upper row shows results of the 3D segmentation on one axial slice, the bottom row shows results of the 4D segmentation on the same axial slice. It can be seen that the pink edema region is changing location at every time point if the 3D segmentation method is used. This is not very likely and probably caused by acquisition artefacts. These artefacts are successfully suppressed by using the 4D segmentation method, leading to an edema volume which has a similar trend, but is generally lower than the volume given by the 3D segmentation method, as can be seen from Fig. 3.

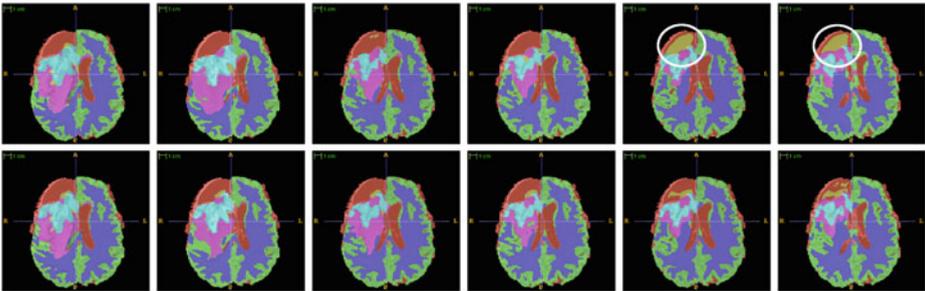


Fig. 5. Illustration of outlier (a) in Fig. 3 (necrotic region of patient 4). The upper row shows results of the 3D segmentation on one axial slice, the bottom row shows results of the 4D segmentation on the same axial slice. It can be seen that due to artefacts during image acquisition, a large fluid filled cavity is wrongly classified as being a necrotic tumor region (yellow) at the last two acquisition time points if the 3D segmentation method is used.

could be that structural images and functional DCE images contain different information.

Computation time on a multi-core CPU with 2.67 GHz was approximately 30 min for a 4-dimensional patient dataset, which translates to 5 min per single time point. This is in the range of the fastest state-of-the art algorithms for brain tumor segmentation.

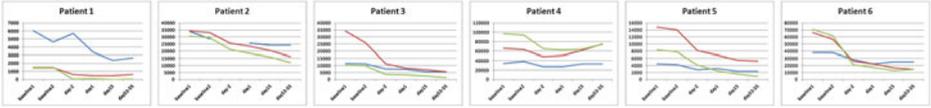


Fig. 6. Trend of the combined necrotic and active tumor volume, which was determined manually on the DCE images (blue line), and which was determined automatically from the four structural modalities, either with training on the BraTS data (red line) or with training on a rough manual tumor outline on the first baseline scan of each patient (green line). Results are shown in mm^3 for all 6 patients. Disease progression in patient 4 could be reliably identified.

4 Discussion and Conclusion

We presented a fully automatic method for integrated spatio-temporal segmentation of longitudinal brain tumor studies. The method is clinically oriented and can be easily used on the standard structural MRI modalities. To the best of our knowledge, this is the first automatic segmentation method, which is dedicated to longitudinal assessment of tumor progression or regression in high-grade glioma patients. It has many potential applications in radiology, oncology and clinical trials because it can eliminate the problem of non-objectiveness and user bias during evaluation and diagnosis.

We have demonstrated that a spatio-temporal segmentation has advantages over an independent treatment of all time points, exhibiting an increased robustness. This is specifically important for cases, where a large number of outliers occur at one time point due to either imaging artefacts or differences in image appearance. Such problems can be effectively handled by the temporal links of the cliques, yielding smoother and more informative curves for the volume trend. We acknowledge that this might lead to some bias by temporal smoothing, but that is outweighed by the improved robustness against outliers as shown qualitatively in Figs. 4, 5 and quantitatively by Dice scores and surface distances in Table 1. The presented method has the potential to allow for a more reliable diagnosis and assessment of tumor progression or regression. We have also shown that the trends of tumor volume evolution over time can be well captured by both intra- and inter-patient training, but obviously the results for intra-patient training are still more accurate.

Finally, we have been able to show that the automatic results for the longitudinal trend of the gross tumor volume, obtained from four structural imaging modalities only, correlated well with the longitudinal trend of the volumes, manually determined based on the DCE images. Additionally, the patient who clearly showed progressing disease could be reliably identified (see patient 4 in Fig. 3 center and in Fig. 6). The results obtained for the longitudinal tumor evolution were also in general agreement with clinical results reported by Ananthnarayan et al. [1] for patients exposed to treatment with an anti-angiogenic compound. Regarding the recommendations of the RANO group for assessing tumor evolution, the

proposed approach aligns well with the vision to consider automatically detected volumetric changes over time for an effective assessment of brain tumor response to therapy in the future.

Acknowledgements. We would like to thank Roche for providing the image data including the manual measurements. This research was partially funded by the Swiss Institute for Computer Assisted Surgery (SICAS), the Swiss Cancer League and the Bernese Cancer League.

References

1. Ananthnarayan, S., Bahng, J., Roring, J., Nghiemphu, P., Lai, A., Cloughesy, T., Pope, W.B.: Time course of imaging changes of GBM during extended bevacizumab treatment. *J. Neuro-Oncology* **88**(3), 339–347 (2008)
2. Angelini, E., Delon, J., Bah, A.B., Capelle, L., Mandonnet, E.: Differential MRI analysis for quantification of low grade glioma growth. *Med. Image Anal.* **16**(1), 114–126 (2012)
3. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. In: Menze, B., Jakab, A., Bauer, S., Reyes, M., Prastawa, M., Van Leemput, K. (eds.) *Miccai Brats Workshop*. Miccai Society, Nice (2012)
4. 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: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III*. LNCS, vol. 6893, pp. 354–361. Springer, Heidelberg (2011)
5. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* **58**(13), R97–R129 (2013)
6. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.: Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Trans. Med. Imaging* **27**(5), 629–640 (2008)
7. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation. manifold learning and semi-supervised learning. Tech. rep., Microsoft Research (2011)
8. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: GLISTR: glioma image segmentation and registration. *IEEE Trans. Med. Imaging* **31**(10), 1941–1954 (2012)
9. Henson, J.W., Ulmer, S., Harris, G.J.: Brain tumor imaging in clinical trials. *AJNR. Am. J. Neuroradiol.* **29**(3), 419–424 (2008)
10. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: setting the state of the art with primal-dual strategies. *Comput. Vis. Image Underst.* **112**(1), 14–29 (2008)
11. Konukoglu, E., Wells, W., Novellas, S., Ayache, N., Kikinis, R., Black, P., Pohl, K.: Monitoring slowly evolving tumors. In: *IEEE ISBI 2008*, pp. 812–815. IEEE (2008)
12. Mazzara, G.P., Velthuisen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int. J. Radiat. Oncol. Biol. Phys.* **59**(1), 300–312 (2004)

13. Pohl, K.M., Konukoglu, E., Novellas, S., Ayache, N., Fedorov, A., Talos, I.F., Golby, A., Wells, W.M., Kikinis, R., Black, P.M.: A new metric for detecting change in slowly evolving brain tumors: validation in meningioma patients. *Neurosurgery* **68**(1 Suppl Operative), 225–233 (2011)
14. Wang, Y., Loe, K.F., Wu, J.K.: A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 279–289 (2006)
15. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., Degroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., Tsien, C., Mikkelsen, T., Wong, E.T., Chamberlain, M.C., Stupp, R., Lamborn, K.R., Vogelbaum, M.A., van den Bent, M.J., Chang, S.M.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J. Clin. Oncol.: Official J. Am. Soc. Clin. Oncol.* **28**(11), 1963–1972 (2010)
16. Zikic, D., et al.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 369–376. Springer, Heidelberg (2012)