A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI

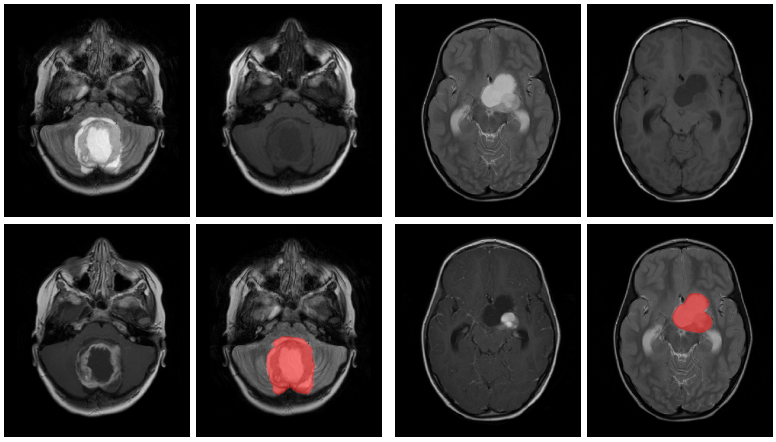
Michael Wels[1,4], Gustavo Carneiro[2], Alexander Aplas[3], Martin Huber[4], Joachim Hornegger[1], and Dorin Comaniciu[2]

[1] Chair of Pattern Recognition, University Erlangen-Nuremberg, Germany
michael.wels@informatik.uni-erlangen.de
[2] Siemens Corporate Research, IDS, Princeton, NJ, USA
[3] Institute of Radiology, University Medical Center, Erlangen, Germany
[4] Siemens, CT SE SCR 2, Erlangen, Germany

Abstract. In this paper we present a fully automated approach to the segmentation of pediatric brain tumors in multi-spectral 3-D magnetic resonance images. It is a top-down segmentation approach based on a Markov random field (MRF) model that combines probabilistic boosting trees (PBT) and lower-level segmentation via graph cuts. The PBT algorithm provides a strong discriminative observation model that classifies tumor appearance while a spatial prior takes into account the pair-wise homogeneity in terms of classification labels and multi-spectral voxel intensities. The discriminative model relies not only on observed local intensities but also on surrounding context for detecting candidate regions for pathology. A mathematically sound formulation for integrating the two approaches into a unified statistical framework is given. The proposed method is applied to the challenging task of detection and delineation of pediatric brain tumors. This segmentation task is characterized by a high non-uniformity of both the pathology and the surrounding non-pathologic brain tissue. A quantitative evaluation illustrates the robustness of the proposed method. Despite dealing with more complicated cases of *pediatric brain tumors* the results obtained are mostly better than those reported for current state-of-the-art approaches to 3-D MR brain tumor segmentation in adult patients. The entire processing of one multi-spectral data set does not require any user interaction, and takes less time than previously proposed methods.

1 Introduction

Detection and delineation of pathology, such as cancerous tissue, within multi-spectral brain magnetic resonance (MR) volume sequences is an important problem in medical image analysis. For example, a precise and reliable segmentation of brain tumors present in the childlike brain is regarded critical when aiming for the automatic extraction of diagnostically relevant quantitative or more abstract findings. This may include the volume of the tumor or its relative location. Once these findings are obtained they can be used both for guiding computer-aided diagnosis and therapy planning as well as for traditional decision making. However, the manual labeling of volumetric data is usually time consuming, which has the potential to delay clinical workflow, such that there is a need for fully automatic



(a) (b)

Fig. 1. Two different cases of pediatric brain tumors exhibiting heterogeneous shape and appearance. Columns (a) and (b) show axial slices of the typically acquired pulse sequences (row-wise from left to right: T2-weighted, T1-weighted, and T1-weighted after contrast enhancement) and the expert annotated ground-truth overlaid to the T2-weighted pulse sequence. Please view in color.

segmentation tools in this particular context. Furthermore, manual annotations may vary significantly among experts as a result of individual experience and interpretation.

As multi-spectral 3-D magnetic resonance imaging (MRI) is the method of choice for the examination of neurological pathology such as brain cancer in pediatric patients, automatic approaches first have to be capable of dealing with the characteristic artifacts of this imaging modality: Rician noise, partial volume effects, and intra-/inter-scan intensity inhomogeneities. Second and more importantly, they have to be robust enough to handle the heterogeneous shape and appearance of pediatric brain tumors in different patients (see Fig. 1).

In this paper, we propose a fully automatic solution based on a novel top-down segmentation scheme that uses a statistical model of the pathology appearance as a constraint for a sub-sequent optimization problem. The statistical model is provided by a machine learning technique that is able to work with high-dimensional feature vectors allowing to encode characteristic voxel contexts. The optimization problem itself is stated as a search for an MAP estimate of the most-likely binary image segmentation, which permits efficient computation of a solution by means of a max-flow/min-cut optimization procedure and is optimal in terms of Bayesian classification theory.

Approaches in the field of MR brain tumor segmentation rarely rely on pure data-driven approaches due to the complexity in terms of tumor shape and appearance of the segmentation task. The vast majority of methods make use of domain knowledge using different types of representation and combine it with low-level imaging techniques. Fletcher-Heath et al. [1] use unsupervised fuzzy clustering followed by 3-D connected components with an intermediate step incorporating knowledge about the usual distribution of cerebral spinal fluid and location of the ventricular system. Gering et al. [2] use trained parametric statistical models for intensity distributions of non-pathologic brain tissue to detect

model outliers on the voxel level that are considered tumor voxels in a multi-layer Markov random field framework. In a similar manner Prastawa et al. [3] detect outliers based on refined intensity distributions for healthy brain tissue initially derived from a registered probabilistic atlas, which introduces structural domain knowledge. Registration is also used in combination with voxel intensities in the adaptive template-moderated classification algorithm by [4]. More recent approaches try to enrich low level segmentation techniques, like level set evolution [5] or hierarchical clustering [6], by using supervised machine learning on higher dimensional feature sets associated with each image voxel. These feature sets are capable of representing a more general variety of domain knowledge on different levels of abstraction. In a similar manner we make use of the recently proposed technique of probabilistic boosting trees (PBT) [7] for supervised learning, which has proven its robustness and its capability for efficient training and classification in numerous applications [8, 9]. The probability estimates provided by PBT are then used to constrain the highly efficient computation of minimum cuts [10] for image segmentation based on a Markov random field (MRF) prior model. It takes into account both coherence of classification labels as well as multi-spectral intensity similarities within voxel neighborhoods.

To the best of our knowledge, this is the first paper giving an integrated formulation for combining PBT classification and computation of minimum cuts. Opposed to [5, 6] there is no involvement of a time consuming bias field correction step in data pre-processing. In the case of [6] this seems to be done by FAST [11], which relies on an HMRF-EM segmentation approach. In the presence of abnormal tissue types this requires the determination of the number of different intensity regions expected within each scan. Furthermore, the inherent low level segmentation might bias the final segmentation result. In contrast we build discriminative models, i.e. PBTs, whose generalization capabilities are strong enough to implicitly handle those intra-patient intensity non-uniformities. Moreover, we apply our method to the more complicated task of segmenting pediatric brain tumors where not only pathology underlies significant variation in shape and appearance but also the non-pathological "background", which is caused by ongoing myelination of white matter during maturation [12].

2 Discriminative Model-Constrained Graph Cuts Segmentation

Our segmentation method relies on the integrated formulation of an objective function that is subject to optimization via the efficient graph cuts algorithm [10]. In the following we derive this objective function from the general MAP framework for image segmentation.

In general, the problem of segmenting an image can be stated as the search for an MAP estimate of the most likely class labels given appropriate prior and observation models in terms of probability density functions. Let $\mathcal{S} = \{ 1, 2, \ldots, N \}$, $N \in \mathbb{N}$, be a set of indices to image voxels. At each index $s \in \mathcal{S}$ there are two random variables: $y_s \in \mathcal{Y} = \{ +1, -1 \}$ and $\boldsymbol{x}_s \in \mathcal{X} = \mathbb{R}^M$, $M \in \mathbb{N}$. The former, y_s, denotes the unobservable binary segmentation of voxel s into fore- and background, whereas the latter, $\boldsymbol{x}_s$, states the observable vector of associated features that are assumed to be causally linked to the underlying class labels $y \in \mathcal{Y}$ by a unified observation model defined by a probability density

function $p(\boldsymbol{x}|y)$ for $\boldsymbol{x} \in \mathcal{X}$. The emergence of the class labels themselves is described by a prior model $p(y)$. The segmentation task at hand can now be stated as the search for an MAP estimate

$$\boldsymbol{Y}^* = \underset{\boldsymbol{Y}}{\operatorname{argmax}}\, p(\boldsymbol{Y}|\boldsymbol{X}) \tag{1}$$

where $p(\boldsymbol{Y}|\boldsymbol{X})$ is the joint posterior probability over the image domain $\mathcal{S}$ with $\boldsymbol{Y} = (y_s)_{s\in\mathcal{S}}$ and $\boldsymbol{X} = (\boldsymbol{x}_s)_{s\in\mathcal{S}}$. Using the Bayes rule, and assuming a uniform distribution $p(\boldsymbol{X})$, we have:

$$\boldsymbol{Y}^* = \underset{\boldsymbol{Y}}{\operatorname{argmax}} \ln p(\boldsymbol{X}|\boldsymbol{Y}) + \ln p(\boldsymbol{Y}). \tag{2}$$

To concretize this optimization problem a region-specific probability term and an appropriate prior need to be identified. In our method $p(\boldsymbol{X}|\boldsymbol{Y})$ is provided by a PBT classifier. The machine learning technique of PBT recursively groups boosted ensembles of weak classifiers to a tree structure during learning from expert annotated data. Training a PBT resembles inducing a multivariate binary regression tree as the final strong classifier

$$H(\boldsymbol{x}) = \sum_{t=1}^{T} h_t(\boldsymbol{x}) \tag{3}$$

generated within each inner node for a feature vector $\boldsymbol{x}$ through a combination of real-valued contributions $h_t(\boldsymbol{x})$ of $T \in \mathbb{N}$ weak classifiers asymptotically approaches the additive logistic regression model [13]

$$H(\boldsymbol{x}) \approx \frac{1}{2} \ln \frac{p(y=+1|\boldsymbol{x})}{p(y=-1|\boldsymbol{x})}. \tag{4}$$

Accordingly, at each inner node v of the resulting PBT there are current approximations of the posterior probability $\tilde{p}_v(y|\boldsymbol{x})$. During classification those values are used to guide tree traversing and combined propagation of posteriors in order to get a final approximation $\tilde{p}(y|\boldsymbol{x})$ of the true posterior probability $p(y|\boldsymbol{x})$ at the tree's root node.

As mentioned above, assuming $\boldsymbol{X}$ to be distributed uniformly, and also to be independently and identically distributed, we have $p(\boldsymbol{x}|y) \propto p(y|\boldsymbol{x})$ and therefore $p(\boldsymbol{X}|\boldsymbol{Y}) \approx \prod_{s\in\mathcal{S}} \tilde{p}(y_s|\boldsymbol{x}_s)$ in (2).

The feature vectors $\boldsymbol{x}$ used for PBT classification consist of individual multispectral intensities, inter-spectrality intensity gradients, and 2-D Haar-like features [14, 15] computed on an intra-axial 2-D context surrounding the voxel of interest. The Haar-like features are derived from a subset of the extended set of Haar-like feature prototypes by [16] and are represented only implicitly in memory by so-called (Rotated) Integral Images. This allows for fast re-computation of the features with respect to a given voxel when they are actually assessed. We decided on 2-D Haar-like features in contrast to the full set of 3-D Haars because of their lower computational cost and memory requirements. Also, as we intend to capture a discriminative representation of the full context, and not only of local edge characteristics at the central voxel, Haar-like feature values

are computed according to the given prototypes on every valid origin and scale within the chosen voxel context.

For the prior distribution we assume a Markov random field prior model

$$p(\boldsymbol{Y}) \propto \exp(-U(\boldsymbol{Y}|\frac{1}{\lambda})) \tag{5}$$

formulated, according to the Hammersly-Clifford Theorem, as a Gibbs distribution with energy function

$$U(\boldsymbol{Y}|\frac{1}{\lambda}) = \frac{1}{\lambda}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{N}_s} V_{st}(y_s, y_t) \tag{6}$$

where $\frac{1}{\lambda}$ with $\lambda \in]0.0, +\infty[$ controls the relative influence of the spatial prior over the observation model and $\mathcal{N}_s$ states the neighborhood of voxel s. Inspired by [17] the interaction potentials are

$$V_{st}(y_s, y_t) = \exp\left(-\frac{1}{2}\sum_{l=1}^{L}\frac{(i_{s_l} - i_{t_l})^2}{\sigma_l^2}\right)\cdot\frac{\delta(y_s, y_t)}{dist(s,t)} \tag{7}$$

where vectors $(i_{s_1}, \ldots, i_{s_L})$ and $(i_{t_1}, \ldots, i_{t_L})$ denote the observed intensities at voxels s and t taken from $L \in \mathbb{N}$ aligned input pulse sequences and

$$\delta(y_s, y_t) = \begin{cases} 1 & \text{if } y_s \neq y_t \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

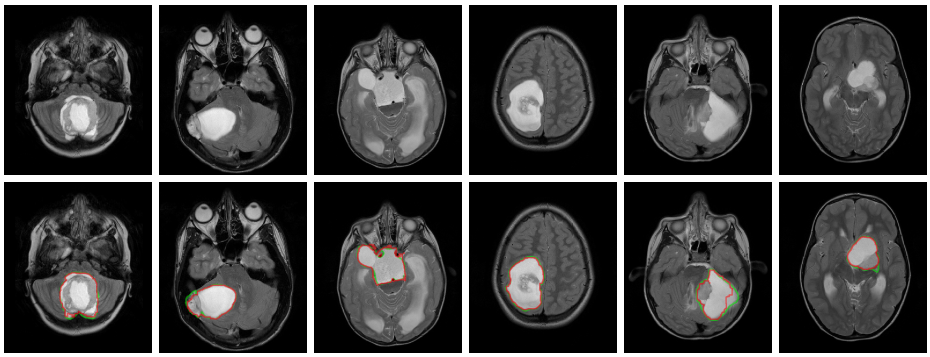
The function $dist(s,t)$ denotes the physical distance between voxels s and t, which varies when working on image volumes with anisotropic voxel spacing. The model emphasizes homogeneous segmentations among neighboring voxels but weights penalties for heterogeneity according to intensity similarities of the voxels involved.

With the equality

$$\begin{aligned}\boldsymbol{Y}^* = \underset{\boldsymbol{Y}}{\operatorname{argmax}} \sum_{s\in\mathcal{S}} \ln\tilde{p}(y_s|\boldsymbol{x}_s) - \frac{1}{\lambda}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{N}_s} V_{st}(y_s, y_t) = \\ \underset{\boldsymbol{Y}}{\operatorname{argmin}}\, \lambda\cdot\left(\sum_{s\in\mathcal{S}} -\ln\tilde{p}(y_s|\boldsymbol{x}_s)\right) + \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{N}_s} V_{st}(y_s, y_t)\end{aligned} \tag{9}$$

the initial maximization problem can be transformed into a minimization problem that is in a suitable form for optimization by the graph cuts algorithm [10]. Note that the reciprocal of the regularization parameter in (6) can equivalently be used to weight the influence of the observation model over the prior model. Given (9), setting up the graph and defining the edge capacities for the associated max-flow/min-cut problem instance is straightforward. Details can be found in [17] with the difference that we do not use additional hard constraints to predetermine individual class labels of certain voxels.[5]

[5] In order to compensate for PBT misclassifications due to cerebral spinal fluid/cyst intensity ambiguities, $\tilde{p}(y_s|\boldsymbol{x}_s)$ in (9) is weighted by the likelihood of the observed multi-spectral intensities at voxel s given y. The PDFs for that are estimated during segmentation via histograms by understanding the hard voxel classification for $\tilde{p}(\boldsymbol{y}_s = 1|\boldsymbol{x}_s) > 0.5$ as intermediate segmentation that is close to the final result.



(a) (b) (c) (d) (e) (f)

Fig. 2. Segmentation results obtained by leave-one-patient-out cross validation. The first row shows selected slices of the T2-weighted pulse sequences of the six available patient data sets. The second row shows the associated segmentation results (red) and the ground-truth segmentation (green) overlaid on the T2-weighted pulse sequence. Please view in color.

3 Experimental Setting and Results

For quantitative evaluation of the proposed method there were six multi-spectral expert annotated data sets of pediatric patients aged from 1 year and 5 months to 15 years and 10 months available—among them four pilocytic astrocytomas, one pilomyxoid astrocytoma, and one anaplastic astroblastoma. Each scan consists of three 3-D images acquired at different pulse sequences (T2-weighted, T1-weighted, and T1-weighted after contrast enhancement). The resolution is $512 \times 512 \times 20$ with a voxel spacing of 0.45mm×0.45mm×6.0mm. Where necessary due to patient movement during image acquisition the pulse sequences were co-aligned by means of the MedINRIA affine registration tool (www-sop.inria.fr/asclepios/software/MedINRIA). All the sequences were further pre-processed by the following pipeline: skull stripping by the Brain Extraction Tool (BET) [18], gradient anisotropic diffusion filtering (www.itk.org), and inter-scan intensity standardization by Dynamic Histogram Warping (DHW) [19]. Note that all of the pre-processing steps involved, including co-alignment, can be performed fully automatically without any user interaction.

The PBT voxel classifiers built were restricted to a maximum depth of 10 with 10 weak classifiers per tree node. The 2-D voxel context considered was of size 11×11 on volumes down-sampled to a voxel spacing of 2.0mm×2.0mm×6.0mm. The graph cuts optimization, using Vladimir Kolmogorov's publicly available implementation [10], is carried out on the original image resolution with $\mathcal{N}_s$ defined to be a standard 6-neighborhood on the 3-D image lattice. The standard deviation $(\sigma_1, \ldots, \sigma_L)$ for the interaction potentials in (7) was estimated offline as "camera noise" within manually delineated homogeneous regions throughout the patient volumes. A leave-one-out cross validation on the patient data sets and their accompanying PBT models yielded best average segmentation scores in terms of the Jaccard coefficient $(TP/(TP + FP + FN)$ where TP, FP, and FN denote the number of true positive, false positive, and false negative voxels, respectively) for $\lambda \in [0.1, 0.5]$ such that finally $\lambda = 0.2$ was chosen for computing the results depicted in Fig. 2. In order to remove small regions of false positive

voxels only the largest connected component of the graph cuts result is considered to be the final segmentation. With Jaccard coefficients of 0.78 ± 0.17 the segmentation results are better than those published by [5] (0.60) and, except for one case, in a similar range as those of [6] (0.86) who all work with adult patient data sets and partly on four pulse sequences [6]. However, comparability of results is limited because of different characteristics between the data sets used by the mentioned scientists, e.g., pediatric patients versus adult patients, additional usage of more expressive pulse sequences, presence of necrotic tissue within the tumors, restriction to a certain histological type of tumor, etc.

It takes about 1–2 minutes to process one of the multi-spectral MRI volumes in a non-optimized C++ implementation of our segmentation method on a Fujitsu Siemens Computers notebook equipped with an Intel Pentium M 2.0 GHz processor and 2 GB of memory. With the same hardware as above training one classifier takes about 4 hours. Preprocessing the images takes about 3 minutes so a total amount of 5 minutes is needed for the processing of one patient data set. In terms of total processing time our method is therefore faster than the method of [6] who claim to be fastest among current approaches to fully automatic MRI brain tumor segmentation.

4 Conclusions

The contribution of this paper is two-fold: starting from the well-known MAP framework for image segmentation we have derived a constrained minimization problem suitable for max-flow/min-cut optimization that incorporates an observation model provided by a discriminative PBT classifier into the process of segmentation. Secondly, we successfully applied the method to the difficult problem of fully automatic pediatric brain tumor segmentation in multi-spectral 3-D MRI. The experimental results obtained are mostly better than those recently published for fully automatic brain tumor segmentation in adult patients.

As the proposed method relies on the observed intensities as a very strong indicator for tumor appearance it is to some extent sensitive to symptoms affecting this feature. In the case of a hydrocephalus where circulation of CSF has nearly come to a still stand it is virtually impossible to distinguish voxels within the cystic portion of the tumor from voxels within the ventricular system from solely the intensities. This may cause the method to generate false positives and false negatives (see Fig. 2 (e)) in some cases.

In the future we will consider an extended use of prior knowledge to overcome this issue. This knowledge would have to cover not only direct tumor characteristics like shape and location, but also indirect characteristics of the surrounding non-pathological brain tissue. The goal is to detect, to segment, and to identify most types of pathological tissue that occur within pediatric brain tumors.

Acknowledgements. This work has been partially funded by the EU project Health-e-Child (IST 2004-027749). We would like to thank A. Rossi from G. Gaslini Children's Research Hospital in Genoa, Italy, for providing the MRI data used in our experiments.

References

1. Fletcher-Heath, L., Hall, L.O., Goldgof, D.B., Murtagh, F.R.: Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. Artif. Intell.

Med. **21**(1–3) (2001) 43–63
2. Gering, D.T., Grimson, W.E.L., Kikinis, R.: Recognizing deviations from normalcy for brain tumor segmentation. In: Int. Conf. Med. Image Comput. and Comp.-Assist. Interv., Tokyo, Japan. (Sept. 2002) 388–395
3. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Med. Image Anal. **8**(3) (Sept. 2004) 275–283
4. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated segmentation of MR images of brain tumors. Radiology **218** (2001) 586–591
5. Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M., Murtha, A.: 3D variational brain tumor segmentation using a high dimensional feature set. In: Proceedings of the Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop (11th IEEE Int. Conf. Comp. Vis., Rio de Janeiro, Brazil). (Oct. 2007) 1–8
6. Corso, J.J., Yuille, A., Sicotte, N.L., Toga, A.: Detection and segmentation of pathological structures by the extended graph-shifts algorithm. In: Int. Conf. Med. Image Comput. and Comp.-Assist. Interv., Brisbane, Australia. (Oct. 2007) 985–993
7. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: IEEE Int. Conf. Comp. Vis., Beijing, China. (Oct. 2005) 1589–1596
8. Tu, Z., Zhou, X., Comaniciu, D., Bogoni, L.: A learning based approach for 3D segmentation and colon detagging. In: Europ. Conf. Comp. Vis., Graz, Austria. (May 2006) 436–448
9. Carneiro, G., Georgescu, B., Good, S., Comaniciu, D.: Automatic fetal measurements in ultrasound using constrained probabilistic boosting tree. In: Int. Conf. Med. Image Comput. and Comp.-Assist. Interv., Brisbane, Australia. (Oct. 2007)
10. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE T. Pattern Anal. **26**(9) (Sept. 2004) 1124–1137
11. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE T. Med. Imag. **20**(1) (Jan. 2001) 45–57
12. Murgasova, M., Dyet, L., Edwards, A.D., Rutherford, M.A., Hajnal, J.V., Rueckert, D.: Segmentation of brain MRI in young children. In: Int. Conf. Med. Image Comput. and Comp.-Assist. Interv., Copenhagen, Denmark. (Oct. 2006) 687–694
13. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Ann. Stat. **28**(2) (1998) 337–407
14. Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T.: Pedestrian detection using wavelet templates. In: IEEE Comp. Soc. Conf. Comp. Vis. and Pat. Recog., San Juan, Puerto Rico. (June 1997) 193–199
15. Viola, P., Jones, M.: Robust real-time object detection. In: 2nd International Workshop on Statistical Learning and Computational Theories of Vision, Vancouver, Canada. (July 2001) 1–25
16. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In Michaelis, B., Krell, G., eds.: DAGM-Symposium. Volume 2781 of Lect. Notes in Comp. Sci., Springer (Sept. 2003) 297–304
17. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Int. J. Comput. Vision **70**(2) (2006) 109–131
18. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3) (2002) 143–155
19. Cox, I.J., Hingorani, S.L.: Dynamic histogram warping of image pairs for constant image brightness. In: IEEE Int. Conf. on Image Proc., Vol. II, Washington, D.C., USA. (Oct. 1995) 366–369