Hierarchical Probabilistic Gabor and MRF Segmentation of Brain Tumours in MRI Volumes

Nagesh K. Subbanna[1], Doina Precup[2], D. Louis Collins[3], and Tal Arbel[1]

[1] Centre for Intelligent Machines, McGill University, Canada
[2] School of Computer Science, McGill University, Canada
[3] McConnell Brain Imaging Centre, McGill University, Canada

Abstract. In this paper, we present a fully automated hierarchical probabilistic framework for segmenting brain tumours from multispectral human brain magnetic resonance images (MRIs) using multiwindow Gabor filters and an adapted Markov Random Field (MRF) framework. In the first stage, a customised Gabor decomposition is developed, based on the combined-space characteristics of the two classes (tumour and non-tumour) in multispectral brain MRIs in order to optimally separate tumour (including edema) from healthy brain tissues. A Bayesian framework then provides a coarse probabilistic texture-based segmentation of tumours (including edema) whose boundaries are then refined at the voxel level through a modified MRF framework that carefully separates the edema from the main tumour. This customised MRF is not only built on the voxel intensities and class labels as in traditional MRFs, but also models the intensity differences between neighbouring voxels in the likelihood model, along with employing a prior based on local tissue class transition probabilities. The second inference stage is shown to resolve local inhomogeneities and impose a smoothing constraint, while also maintaining the appropriate boundaries as supported by the local intensity difference observations. The method was trained and tested on the publicly available MICCAI 2012 Brain Tumour Segmentation Challenge (BRATS) Database [1] on both synthetic and clinical volumes (low grade and high grade tumours). Our method performs well compared to state-of-the-art techniques, outperforming the results of the top methods in cases of clinical high grade and low grade tumour core segmentation by 40% and 45% respectively.

1 Introduction

Worldwide, it is estimated that roughly 238,000 cases of brain tumours are diagnosed every year [2]. One of the primary diagnostic evaluation tools for brain tumours is magnetic resonance imaging (MRI) of the brain to evaluate the size of the tumour and its proximity to critical structures of the brain. Brain tumours present significant challenges to traditional segmentation techniques due to the wide variability in their appearance in terms of shape, size, position within the brain, their intensity variability and heterogeneity caused by swelling (edema), the presence of cysts, tumour type, etc. Moreover, the image acquisition parameters, scanner type, imaging artefacts, and pre-processing steps also greatly affect the appearance of tumours in MRI. The tumour boundaries appear differently in different MRI contrasts (e.g. T1w, T2w, or FLAIR images) commonly used in clinical contexts leading to non-trivial ambiguities both in terms of the

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 751–758, 2013.
© Springer-Verlag Berlin Heidelberg 2013

modelling and with regards to expert ground truth labelling required for training and testing. Finally, depending on the tumour type, the contrast between the boundaries and the surrounding healthy tissue can often be quite weak. These factors greatly impede the ability of generative techniques to properly model and predict tumour appearance.

Over the years, a number of techniques have been successfully devised to segment brain tumours automatically. These include atlas-based techniques [4], [5], [6], [9] which register brain tumour volumes to healthy brain atlases, and classify tumour regions as outliers. Generative techniques [3], [13] model the brain tumour intensity patterns during training and segment tumours in test images based on these models. Discriminative techniques [7], [8], [10], [11] on the other hand, avoid modelling intensity distributions and instead determine boundaries between tumour and non-tumour classes based on local intensity and/or some regional information. Overall, the literature suggests that it is advantageous to embed both local and global information into the framework in order to localise tumours and segment their boundaries correctly. Global information can be provided through position frequency techniques which offer a convenient way of coarsely detecting and locating tumours [12] based on learning texture patterns over images. Markov Random Field (MRF) approaches have been devised to model local information to segment tumours [13], however they have primarily been used to perform spatial regularisation through a class prior. The premise of this paper is that optimal tumour segmentation can be achieved through the careful design and combination of coarse Gabor texture based segmentation and a refined MRF model into a single probabilistic framework that leverages the advantages of both techniques. Preliminary work based on a similar idea was presented in [15] at MICCAI BRATS 2012, but this paper presents substantial improvements including a customised Gabor filterbank to optimally separate tumour, and non-tumour, an improved MRF to include inter-class transition probabilities, and also model intensity differences between classes.

In this paper, we develop a fully automated, hierarchical probabilistic framework for segmenting brain tumours from multimodal brain MRI. At the first stage, the goal is to coarsely segment the tumour (and associated edema) from the surrounding healthy tissue using texture based features. Here, specialised Gabor functions are developed to optimally separate the tumour class from the surrounding healthy tissues during training. A Bayesian classification framework is developed, based on the combined space Gabor decomposition, resulting in tumour/non-tumour probabilities. In the second stage, the boundaries and details of the segmentation are refined through an adapted probabilistic graphical MRF model, designed to separate the edema from the main tumour. This customised MRF differs from standard MRFs in that it is not simply a smoothing operator through a prior on class labels. In addition to taking voxel intensities and class labels into account, it also models the intensity differences between neighbouring voxels in the likelihood model and considers the likelihood of transitions between classes.

The entire proposed framework is trained and tested on the publicly available, MICCAI 2012 Brain Tumour Segmentation Challenge (BRATS) database [1]. On-line segmentation statistics (e.g. Dice overlap metrics) are provided. In comparison with other participants, our method outperforms the top methods from the competition in the cases of clinical high grade and low grade tumour core by 40% and 45% respectively, and performs roughly as well as the top methods in other categories.

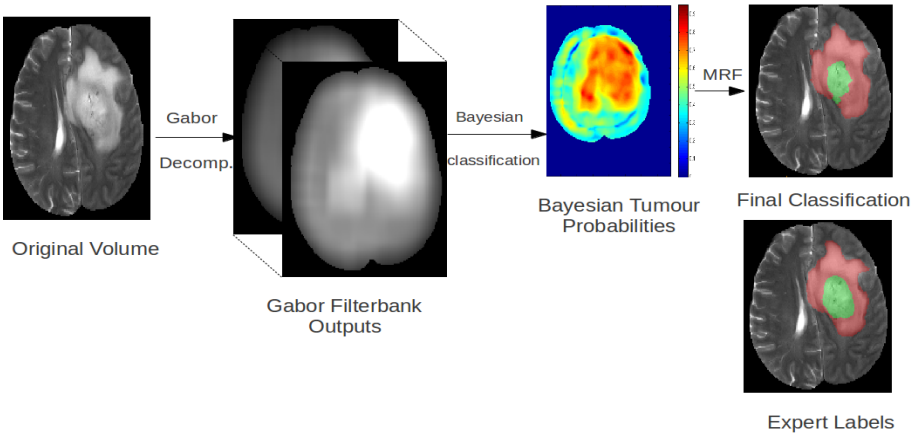

Fig. 1. Flowchart displaying the various stages of the classification technique. In the MRF classification and expert labels, red label represents edema, and green represents tumour.

2 Proposed Framework

We develop a hierarchical probabilistic brain tumour segmentation approach, using two stages. In the first stage, multiwindow Gabor decompositions of the multi-spectral MRI training images are used to build multivariate Gaussian models for both the healthy tissues and the tumour (including core and edema tissue). A Bayesian classification framework using these features is used to obtain initial classification results. In the second stage, Gaussian models are built for healthy tissues (i.e. grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF)) as well as for tumour tissues (e.g. core tissues, edema) from intensity distributions acquired from the training dataset. A Markov Random Field is trained to classify all these types of tissues. A flowchart of the process is shown in Fig. 1. We now present in detail the two stages.

2.1 Stage 1: Multiwindow Gabor Bayesian Classification

Training: The data consists of MRI intensity volumes in different contrasts (T1, T1c (T1-post gado-contrast), T2 and FLAIR). Hence, at each voxel, we have a 4-dimensional vector containing the intensity in each contrast. Each contrast $\mathbf{f}$ of each volume is processed using multiwindow, 2D Gabor transforms of the form suggested by [16]. We use a set of R window functions $g_r, r = 1 \ldots R$ of the form:

$$g_r[x, y; a, b, n_1, n_2, m_1, m_2, \sigma_{x_r}, \sigma_{y_r}] = e^{-((x-n_1 a)^2/{\sigma_{x_r}}^2 + (y-n_2 a)^2/{\sigma_{y_r}}^2)} e^{-j2\pi(m_1 bx + m_2 by)/L}, \tag{1}$$

where L is the total number of voxels in the slice under consideration, x and y are space coordinates within the slice, a and b are the magnitude of the shifts in the spatial and frequency domains respectively, $n_{1,2}$ and $m_{1,2}$ are the indices of the shifts in the

position and frequency domains respectively, σ_{x_r} and σ_{y_r} are variance parameters of the r-th window. In our experiments we chose b such that there are 6 equally spaced orientations between 0 and π radians (sufficient in practice) and $a = 1$. Let $\mathbf{G}$ be the Gabor matrix whose columns are generated by picking all possible shift values for both a and b for all the R windows with every x and y represented in each column. The filter bank coefficients $\mathbf{c}$ are obtained by convolving each contrast volume slice by slice with the Gabor filter bank $\mathbf{G}$. We use the same $\mathbf{G}$ matrix for all contrasts. It was proved in [18] that if both $g_r[\cdot]$ and its remapping window are positive definite[1], then any $g_r[\cdot]$ can be used and perfect reconstruction is possible. Here, both $g_r[\cdot]$ and its remapping window are Gaussian (but with different parameters), and are positive definite.

Each voxel in the training volumes belongs to one of two classes: tumour or healthy tissue. We estimate the window function parameters (σ_{x_r} and σ_{y_r}) that will maximize the distance between the two classes. More formally, let $\{\mathbf{f}_t\}$ and $\{\mathbf{f}_h\}$ be the sets of voxels belonging to the tumour class and the healthy class respectively. The corresponding tumour coefficients $\mathbf{c}_t$ in the combined space are obtained by a convolution of the Gabor filters centred at the tumour voxels in question. Similarly, the $\mathbf{c}_h$ are obtained with Gabor filters centred at the non-tumour voxels in question. Ideally, the coefficients of the tumour and healthy class should be as different as possible. To achieve this goal, we solve the following optimisation problem:

$$(\sigma_x, \sigma_y) = \arg\max_{\sigma_x, \sigma_y} \sum_{j,k} \mid c_j - c_k \mid, \forall c_j \in \{\mathbf{c}_t\}, \forall c_k \in \{\mathbf{c}_h\} \tag{2}$$

where σ_x and σ_y are the vectors containing the R σ_{x_r} and σ_{y_r}. We solve this optimisation using simulated annealing during training.

Classification: Each test volume is decomposed into its multiwindow Gabor filter bank outputs, $\mathbf{I}^G$, using the convolution at each voxel described above. The class of each voxel i, C_i is then estimated using Bayesian classification:

$$P(C_i \mid \mathbf{I}_i^G) \propto P(\mathbf{I}_i^G \mid C_i) P(C_i), \tag{3}$$

where $\mathbf{I}_i^G$ is the set of Gabor coefficients of voxel i.

2.2 Markov Random Field Classification

The main purpose of the second stage is to remove false positives, distinguish the different sub-types of tumour tissue (e.g. core vs edema), and refine the boundaries of the tumour. The proposed Markov Random Field (MRF) framework differs in several important ways from standard MRF approaches. First, the model is designed specifically to model the differences in intensity between a voxel and its neighbours probabilistically, in order to preserve the correct tumour boundaries. Our MRF uses significantly larger clique sizes than in standard models, which typically use only pairs of voxels. The prior models are conditioned on all possible class configurations within the neighbourhood. More precisely, we consider that the label of voxel i, C_i, is probabilistically inferred through an intensity vector, $\mathbf{I}_i$, and cliques involving the voxel and some of

[1] Positive definite windows are windows whose discrete Fourier transform is real and positive.

its adjacent voxels in the neighbourhood N_i. For all voxels j in the neighbourhood N_i, $j \in N_i$, there exists a corresponding class vector $\mathbf{C}_j$ and corresponding set of intensity vectors $\mathbf{I}_j$. The energy at voxel i has to be minimised to infer the optimal classification for $P(C \mid \mathbf{I})$ as follows: The energy of voxel i is given by:

$$U(C_i \mid \mathbf{I}_i) = -[\log P(C_i) + \log P(\mathbf{I}_i \mid C_i) + \sum_{j \in N_i} \log P(\Delta \mathbf{I}_{i,j} \mid C_i, \mathbf{C}_j)] + \alpha m(\mathbf{C}_j, C_i), \tag{4}$$

where $P(C_i)$ is the prior probability of class C_i, $P(\mathbf{I}_i \mid C_i)$ models the likelihood of C_i given the intensity of voxel i, $P(\Delta \mathbf{I}_{i,j} \mid C_i, \mathbf{C}_j)$ models the difference in intensity between i and voxels in the j-th clique for classes C_i and $\mathbf{C}_j$, $m(\mathbf{C}_j, C_i)$ is the potential of transitioning from C_i to $\mathbf{C}_j$ and α ($\alpha = 1$ here) is a weighting parameter.

Training: During training, the volumes are non-linearly registered to a brain tissue atlas, masking out the tumour region using the experts classification labels. The registration allows us to generate separate labels for grey matter, white matter, and cerebrospinal fluid. The core and edema are superimposed from the expert labels. We consider an 8-neighbourhood around the voxel in the axial plane as well as the corresponding voxels in the slices above and below. The neighbourhood N_i consists of all size 2, 3, and 4 cliques that contain voxel i. When only 2 tumour class labels are available for training (i.e. "Two class Gabor-MRF"), we chose to model the single voxel clique likelihood $P(\mathbf{I}_i|C_i)$ as a Gaussian mixture model (GMM) with 8 component Gaussian mixtures, for both the tumour core and edema classes due to the heterogeneity of the regions. However, when various class labels are available for the tumour core (e.g. necrotic core, enhancing tumour, solid tumour), we chose to use single multivariate Gaussian distributions for each class instead of Gaussian mixture models. The healthy classes are all modelled as multivariate Gaussians. The differences between intensities of the different classes are modelled as multivariate Gaussian distributions for all class combinations in both cases. The class transition probabilities are extracted from the frequency of co-occurrence in the training volumes.

Classification: The Gabor-Bayesian probabilities are used as priors for the tumour and the edema classes, with an exponential decay from the initially classified tumour regions. The tumour areas are masked out and healthy atlases are registered to the remaining regions to get the prior probabilities of healthy tissues in the non-tumour regions. Iterated conditional modes (ICM) [19] are used to minimise the total energy as the initial classification provides a good starting point for the optimisation.

3 Experiments and Results

The framework was trained and tested on the publicly available MICCAI 2012 Brain Tumour Segmentation Challenge datasets[1]. Here 20 high grade and 10 low grade real tumour volumes are available for training with 2 class (core and edema) and 4 class (necrotic core, enhancing tumour, solid tumour, and edema) labels. In addition, 25 high grade and 25 low grade synthetic volumes with 2 class labels are also available. The final test set consisted of 11 high grade and 4 low grade real data sets, and 10 high grade

and 5 low grade synthetic volumes. The algorithm was trained on both the 2 class and 4 class labels for real glioma cases, as well as on the 2 class synthetic data, separately for both high and low grade tumours. As the system is set up to test 2 class labelling: i.e., only core and edema, for the results of the 4 class case, the 3 tumour labels were merged to create a single tumour core label. All the resulting test labels were uploaded onto the website and the statistics were provided automatically. Testing of the algorithm took slightly more than an hour per case on a Dell Optiplex 980 I7 machine.

Qualitative Results: Fig. 2 shows the results of our algorithm on a slice from a low grade tumour and a high grade tumour against the experts segmentation, along with the corresponding unlabelled T1c and FLAIR slices. Visually, in both cases, it can be seen that our results are comparable to the experts' labelling.

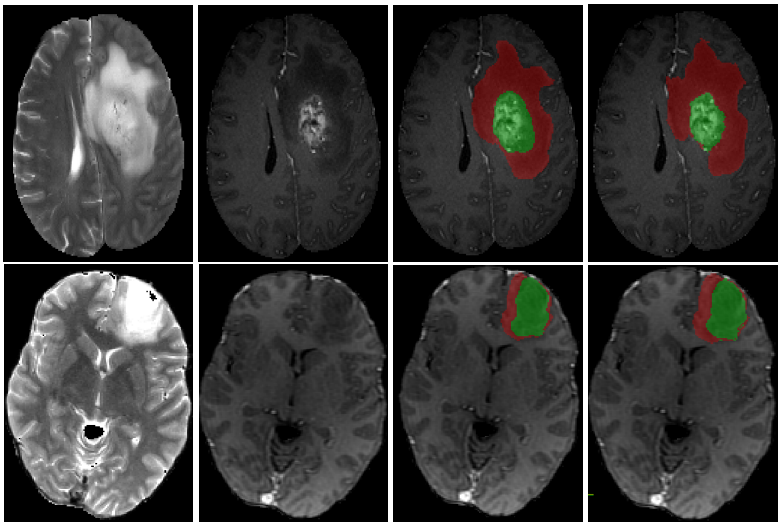

Fig. 2. Row 1, case HG0011, Row 2, case LG0015. (a) The unlabelled FLAIR slice, (b) the unlabelled T1C slice (c) expert labelling and (d) our algorithm labels, (red = edema, green = tumour). Our algorithm's labels corresponds closely with the experts' labels.

Quantitative Results: Table 1 shows the results of our technique on both real glioma and synthetic tumour volumes. Dice similarity coefficient was used to compare results, as in the BRATS challenge. Our algorithm outperforms the winning algorithm by about 40% in the case of high grade (HG) tumour cores and about 45% in the case of low grade (LG) tumour cores, and has statistics comparable to the winner for edema. Table 2 shows that our statistics are comparable to the winners with Dice averages of around 0.8 overall for LG tumours (core and edema) and 0.85 for HG. The 4 class Gabor-MRF outperforms the 2 class Gabor-MRF mainly due to the heterogeneity of the tumour core. Modelling the core as a GMM with 8 modes in the 'two class case' is proven to be less effective than training on each sub-class separately, in the four class case, even with a simple model for each sub-class.

Table 1. Comparison of the segmentation results of the proposed method and the methods participating in the BRATS Challenge as presented on the website [1] for both edema and tumour core for clinical cases using average Dice similarity coefficient values. The final results of our method are found in the "Four class Gabor-MRF", where the method was trained and tested on the 4 class tumour labels and the 3 core labels were merged to create a single core label prior to uploading. The "Two class Gabor-MRF", where the algorithm was trained and tested on a single tumour core class, is shown for comparison. The winners of the challenge are highlighted in bold.

Method	HG Edema	HG Tumour Core	LG Edema	LG Tumour Core
Shin. et. al	0.038	0.144	0.061	0.232
Bauer. et. al.	0.536	**0.512**	0.179	0.332
Zikic. et. al.	**0.598**	0.476	**0.324**	**0.339**
Subbanna. et. al.	0.166	0.248	0.14	0.245
Xiao. et. al.	0.539	0.337	0.279	0.224
Zhao. et. al.	0.003	0.058	0	0
Two class Gabor-MRF	0.613	0.641	0.268	0.241
Four class Gabor-MRF	0.56	0.74	0.240	0.49

Table 2. Comparison of the segmentation results of the proposed method "Two class Gabor-MRF" and the methods participating in the BRATS Challenge as presented on the website [1] against the experts' labels for both the edema and the tumour core for the synthetic cases in the BRATS Challenge [1] using Dice similarity coefficient values. The winners of the challenge are highlighted in bold. As may be observed, our technique has a performance comparable to the winners of the challenge. In the case of synthetic tumour volumes, both our algorithm and the winners have Dice values of around 0.85.

Method	HG Edema	HG Tumour Core	LG Edema	LG Tumour Core
Shin. et. al	0.312	0.284	0.213	0.072
Bauer. et. al.	0.785	0.779	0.746	**0.858**
Zikic. et. al.	**0.85**	**0.869**	**0.749**	0.842
Subbanna. et. al.	0.696	0.398	0.645	0.42
Xiao. et. al.	0.343	0.414	0.1	0.469
Zhao. et. al.	0	0	0	0
Two class Gabor-MRF	0.877	0.841	0.772	0.832

4 Discussions, Future Work and Conclusion

The results show that our method performs very well in the task of segmenting brain tumours and edema in both synthetic and real clinical multimodal brain MRI. It outperforms other techniques in segmenting tumour cores. Its main strength is the hierarchical approach that first coarsely segments the tumour using a customised Gabor decomposition and then refines the segmentation using an adapted MRF. We hope to enhance our technique by improving the second stage to recover from any errors made during the first stage by using a more flexible hierarchical MRF. Currently, the accuracy of the second stage is dependent on obtaining a reasonable classification from the first stage. We are working on solving the optimal cluster separation analytically. Finally, although

the results of the experiments on this dataset are promising, the automated approaches are all tested against the subjective ground truth labels of one set of clinicians. Further experimentation on data acquired from multiple centres would be desirable.

Acknowledgement. The authors wish to thank the organisers of MICCAI BRATS Challenge 2012 [1] for providing the data, tumour labels and online validation tools.

References

1. http://www2.imm.dtu.dk/projects/BRATS2012/data.html
2. Ferlay, J., et al.: Estimates of worldside burden of cancer in 2008. In: GLOBOCAN 2008 (2008)
3. Corso, J., et al.: Efficient Multilevel Brain Tumour Segmentation with Integrated Bayesian Model Classification. IEEE Trans. Med. Imag. 27(5), 629–640 (2008)
4. Kaus, M., et al.: Adaptive Template Moderated Brain Tumour Segmentation in MRI. In: Workshop Fuer Bildverarbeitung Fur Die Medizin, pp. 102–105 (1999)
5. Prastawa, M., et al.: A brain tumor segmentation framework based on outlier detection. Med. Image Ana. 8(3), 275–283 (2004)
6. 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: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010)
7. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: 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)
8. 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 3D MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 67–75. Springer, Heidelberg (2008)
9. Moon, N., et al.: Model based brain and tumor segmentation. In: ICPR, vol. 1, pp. 528–531 (2002)
10. Bauer, S., et al.: Segmentation of Brain Tumour Images Based on Integrated Hierarchical Classification and Regularisation. In: BRATS MICCAI (2012)
11. Parisot, S.: Graph-based Detection, Segmentation and Characterisation of Brain Tumours. In: CVPR, pp. 988–995 (2012)
12. Mishra, R.: MRI based brain tumor detection using wavelet packet feature and Artificial Neural Networks. In: Int. Conf. and Work. on Emerging Trends in Tech., pp. 656–659 (2010)
13. Bauer, S., et al.: Atlas-Based Segmentation of Brain Tumor Images Using a Markov Random Field-Based Tumor Growth Model and Non-Rigid Registration. In: IEEE EMBS, pp. 4080–4083 (2010)
14. Farias, G., et al.: Brain Tumour Diagnosis with Wavelets and Support Vector Machines. In: Int. Conf. Intell. Systems and Knowledge Engg., pp. 1453–1459 (2008)
15. Subbanna, N., et al.: Probabilistic Gabor and Markov Random Fields Segmentation of Brain Tumours in MRI Volumes. In: BRATS MICCAI (2012)
16. Zibulski, M.: Discrete multiwindow Gabor-type transforms. IEEE Trans. Sig. Proc. 45(6), 1428–1442 (1997)
17. Jain, A., et al.: Unsupervised Texture segmentation using Gabor filters. Patt. Rcgn. 24(12), 1167–1186 (1991)
18. Subbanna, N., et al.: Existence Conditions for Non-Canonical Discrete Multiwindow Gabor Frames. IEEE Trans. Sig. Proc. 55(10), 5113–5117 (2007)
19. Duda, R., et al.: Pattern Classification. John Wiley and Sons (2000)