Semi-Supervised Tumor Detection in Magnetic Resonance Spectroscopic Images Using Discriminative Random Fields

L. Görlitz[1,2*], B. H. Menze[2*], M.-A. Weber[3], B. M. Kelm[2], and F. A. Hamprecht[2]

[1] Corporate Research, Robert Bosch GmbH, Germany
[2] Multidim. Image Processing, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg
[3] Department of Radiology, German Cancer Research Center (dkfz), Heidelberg

[*] contributed equally

Abstract. Magnetic resonance spectral images provide information on metabolic processes and can thus be used for in vivo tumor diagnosis. However, each single spectrum has to be checked manually for tumorous changes by an expert, which is only possible for very few spectra in clinical routine. We propose a semi-supervised procedure which requires only very few labeled spectra as input and can hence adapt to patient and acquisition specific variations. The method employs a discriminative random field with highly flexible single-side and parameter-free pair potentials to model spatial correlation of spectra. Classification is performed according to the label set that minimizes the energy of this random field. An iterative procedure alternates a parameter update of the random field using a kernel density estimation with a classification by means of the GraphCut algorithm. The method is compared to a single spectrum approach on simulated and clinical data.

1 Introduction

One major challenge in image processing is to exploit spatial correlation in 2-D images. Certain imaging techniques, however, are not only able to record *one* spatially resolved scalar signal, but provide a full vector of different features per pixel. Spectral images are examples of such multidimensional data sets and are in common use, e.g. in satellite remote sensing or non-invasive diagnostics. If the mapped process can be assumed to exhibit some spatial correlation, combining the information of the spectral and spatial dimension will allow for better decisions than the interpretation of one spectrum alone, especially with noisy spectra. Often these two sources of information are processed in a consecutive manner by first analyzing the spectral image spectrum-by-spectrum, and then using the spatial context in a second *post-hoc* step on the label map resulting from the spectrum-wise processing.

Magnetic resonance (MR) spectroscopy is a non-invasive diagnostic method used to determine the relative abundance of specific metabolites at arbitrary locations *in vivo*. Characteristic changes in the spectral pattern can be linked to specific changes of the tissue, providing means for the grading and localization of tumors, e.g. in brain, breast and prostate [9]. Magnetic resonance spectroscopic imaging (MRSI) allows to acquire such spectra on two- or three-dimensional grids. Each spectrum is represented by a vector of several hundred spectral channels and shows a low number of relevant resonance lines, e.g. 5-10 for MR spectra of the brain. When searching for tumorous changes of the spectrum, pattern recognition methods can be applied to evaluate the data in a highly automated fashion and to guide the radiologist to the relevant regions of the spectroscopic image. [6, 18, 11, 16, 23] process the spectral image spectrum by spectrum and in [14] an approach is presented which incorporates spatial information in the classification procedure.

Typically, a limited number of spectra are diagnosed manually by a physician, providing patient-specific, diagnostic information on the tumor. In the following we propose an approach for the detection and localization of brain tumors which uses this information in a flexible, semi-supervised classifier for an adaptive processing of the complete spectral image. It allows to both process spectral information and to exploit the spatial correlation of the data in a coherent, highly adaptable framework (chapter 2). Our approach relies on common chemometric models in the classification of the spectral information and a spatial model, motivated by Bayesian image processing, for the spatial regularization. Seeking for a time-efficient implementation in the clinical setting, we propose an efficient solver based on the GraphCut algorithm in an iterative strategy. Finally the algorithm is tested on simulated and real data, with results shown in chapter 3.

2 Spatio-Spectral Classification Model

The classification of spatio-spectral data can be separated into two tasks: the inference on the spectral signal alone, a learning problem on highly collinear data, and the formulation of a spatial model on the resulting label map combining information from the single-voxel spectral model with a spatial smoothness assumption on the labels.

2.1 Spectral Model

In the following let $X_i = (X_i^1, \ldots, X_i^p)$ represent a p-dimensional spectrum, and Y_i a binary random variable taking values in $\{0, 1\}$, with $Y_i = 0$ for healthy and $Y_i = 1$ for tumorous tissue.

Given appropriate training data, the information of a spectrum X_i can be mapped to low dimensional scores, e.g. to the probabilities of either showing

characteristic tumorous changes of the spectral pattern (with posterior distribution $\pi(Y_i = 1|X_i)$), or to be within the normal range of spectra originating from healthy tissue ($\pi(Y_i = 0|X_i)$). The posterior probability can be estimated with any method, linear (e.g. linear discriminant analysis, partial least squares regression) or nonlinear (e.g. support vector machines, mixture discriminant analysis), parametric or nonparametric, generative or discriminative. A regularization, however, might be indicated, as collinearity between the channels of a spectrum often leads to intrinsic dimensionalities well below the nominal length of the feature vector.

In the current application we have chosen an approach which combines the strong regularization of a chemometric spectral model and the variability of a nonparametric classifier. By assuming a flat prior on the classes ($\pi(Y_i = 0) = \pi(Y_i = 1)$) and by Bayes' rule, it suffices to estimate $\pi(X_i|Y_i = 0)$ and $\pi(X_i|Y_i = 1)$ to predict the most probable assignment of the spectrum X_i. For this, we used a Parzen kernel-density estimator with bandwidth chosen according to Silverman's "rule-of-thumb" [22] on a reduced feature subspace defined by the first two principal components of an external training set.

2.2 Spatio-Spectral Model

All N spectra $\mathbf{X} = \{X_1, \ldots, X_N\}$ lie, by acquisition, on a regular 2-D or 3-D Cartesian grid. The task is to identify each spectrum X_i with either healthy or tumorous tissue. It is assumed that a tumor is significantly larger than the spatial sampling distance, leading to spatial smoothness of the predicted classes. In order to incorporate this smoothness assumption into the spatio-spectral model, a graph-based method was used. Thus the structure of the spectral image is represented by an undirected graph $G = (S, E)$, with vertices S and edges E, with each site $s \in S$ representing a voxel of spectral acquisition and the set of edges E representing the neighborhood relation and therefore, the spatial coupling of the random variables $\{Y_i\}$. In our experiments we chose the set E to be derived from the rectangular 2-D Cartesian acquisition grid, i.e. a 4 neighborhood system, which uses at most pairwise interactions between labels $\mathbf{Y} = \{Y_i\}$ and therefore keeps efficiency in inference and classification.

In order to explicitly model the spatial homogeneity assumption and to exploit the advantage of discriminatory models [19] if enough training data is available, the posterior distribution $p(\mathbf{Y}|\mathbf{X})$ is modeled as a DRF [13] with penalty term given by a parameter-free function. The single-site potential is formulated to reflect the information of the spectral model, and is given by

$$\mathrm{ssp}(\mathrm{Y_i}|\mathrm{X_i}) = -\log \pi(\mathrm{Y_i} = 1|\mathrm{X_i}) \cdot \mathrm{Y_i} - \log \pi(\mathrm{Y_i} = 0|\mathrm{X_i}) \cdot (1 - \mathrm{Y_i}) \qquad (1)$$

and the pair-potential, responsible for the spatial coupling of the labels, is

$$\mathrm{pp}(\mathrm{Y_i}, \mathrm{Y_j}|\mathbf{X}) = \begin{cases} \nu \cdot \gamma(X_i, X_j) \cdot |Y_i - Y_j| & \text{if } \mathrm{X_i} \sim \mathrm{X_j} \\ 0 & \text{else} \end{cases} \qquad (2)$$

where $X_i \sim X_j$ means that X_i and X_j originate from connected vertices, $\gamma(X_i, X_j)$ gives the penalty incurred when Y_i and Y_j are classified to different classes, and ν governs the trade-off between the purely voxel based classification and the spatial smoothness of the label map. For each spectral image $\mathbf{X}$ this induces the following probability distribution on $\{0,1\}^N$, which is an Ising model on $\mathbf{Y}$ given $\mathbf{X}$ [15]:

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z} \exp\left(-\sum_{i=1}^{N} \mathrm{ssp}(\mathrm{Y_i}|\mathrm{X_i}) - \sum_{\mathrm{i,j=1}}^{\mathrm{N}} \mathrm{pp}(\mathrm{Y_i}, \mathrm{Y_j}|\mathbf{X}) \right) \tag{3}$$

and the sought classification is given as the maximum a posteriori (MAP) estimate of this distribution, which corresponds to using the Bayes estimator for the zero-one loss function. According to $\mathrm{ssp}(\mathrm{Y_i}|\mathrm{X_i})$, the spectrum X_i is classified to the most probable class. If two neighboring vertices s_i and s_j are assigned to different classes, a penalty $\gamma(X_i, X_j)$ is incurred, which depends on the similarity of the two spectra X_i and X_j. In contrast to the DRF as used in [13], the penalty is given as a function and not inferred from training data, leading to a significant decrease of the number of parameters to be estimated. In the current model we have chosen the square root of the Perona-Malik tensor [20]

$$\gamma(X_i, X_j) = \frac{1}{\sqrt{|X_i - X_j|^2 + 1}} \tag{4}$$

with $|X_i - X_j|$ denoting a distance between the features of X_i and X_j used in the spectral model, i.e. in this case the Cartesian distance of the projection into the subspace spanned by the first two principal components. The function penalizes the assignment of different labels to neighboring spectra, unless they are very dissimilar. An illustrative example is shown in figure 1. The amount of evidence needed for such a classification is governed by the trade-off parameter ν.

A similar model for object extraction by GraphCut has been proposed by Boykov and Funka-Lea [1]. If $\pi(Y_i = 0) = \pi(Y_i = 1) \quad \forall\, i$ is assumed, this DRF can be seen as a discrete formulation of the active-contours based level-set approach to classification described in [4].

2.3 Semi-Supervised Solution

The posterior distributions $\pi(Y_i|X_i)$ are not known in the beginning, as a non-parametric kernel-density estimator is used to model both class densities. To optimally adapt to different patients, this estimate is obtained in a semi-supervised, patient dependent manner. In clinical practice, a limited number of spectra in the MRSI is always checked and diagnosed by the physician. The resulting labels, which are optimally adapted to the data, are used for the initialization of the estimate of the posterior distributions $\pi(X|Y = 0)$ and $\pi(X|Y = 1)$. To this end, a kernel density estimate is performed for each class separately, in the reduced

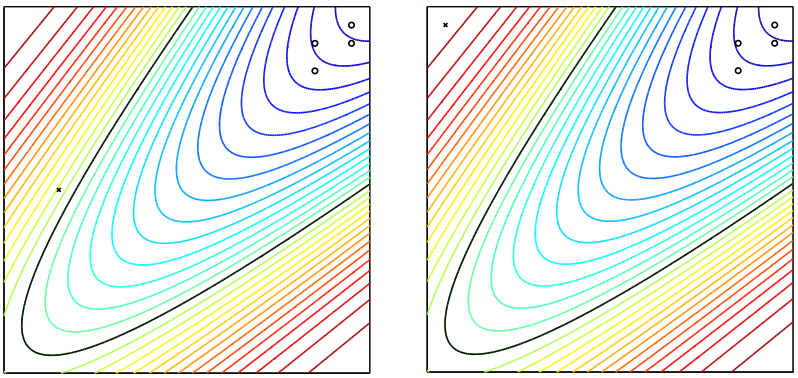

Fig. 1. Point 'x' and its four neighbors 'o' with the posterior distribution in the feature space of the spectral model and decision border indicated; LEFT: label of sample 'x' will flip for low ν, due to different classification of 'o' and weak support for classification; RIGHT: label of 'x' has strong evidence and will not flip, though all neighbors are classified differently.

two-dimensional feature space spanned by the first principal components of the spectra. As the hand-assigned labels should not change in the iteration process the single-site potentials for these spectra are changed to $\mathrm{ssp}(\mathrm{Y_i}|\mathrm{X_i}) = \infty \cdot (1 - \mathrm{Y_i})$ for a tumor label and $\mathrm{ssp}(\mathrm{Y_i}) = \infty \cdot \mathrm{Y_i}$ otherwise.

With this first estimate of the class distributions and with an initial value for ν the classification, corresponding to the maximum probability state of the distribution given in equation (3), can be efficiently calculated by using Bayes theorem to obtain $\pi(Y_i|X_i)$ from $\pi(X_i|Y_i)$ and the GraphCut algorithm [3, 12, 2]. The latter is an instance of the MinCut/MaxFlow-algorithm from graph theory, introduced to image segmentation by Greig et al. [8]. This results in an updated classification of the spectra which, in the next iteration, is used for an update of the kernel-density estimation used to obtain the single-site potentials via Bayes theorem. These two steps are iterated until no spectra changes its classification in subsequent iterations. In our experiments we found that hardly ever more than four iterations were needed until convergence.

This iterative procedure obviously is a version of Dempster's Expectation-Maximization [5] with hard class assignments. It is essential for this approach to start with a good initialization (Fig. 2). Using spectra showing an ambiguous spectral pattern leads to a significantly worse classification result, compared to an initialization with spectra showing a clear pattern for either class.

It is often desirable to show the confidence in the classification. To this end, Gibbs sampling [7] from the posterior distribution $p(\mathbf{Y}|\mathbf{X})$ can be used, a Markov Chain Monte Carlo method [21]. In order to employ a Gibbs sampler, the local

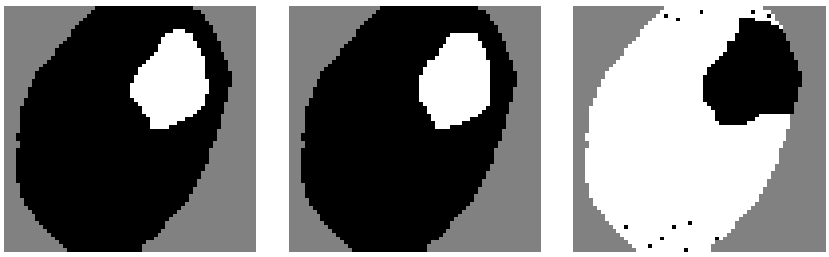

Fig. 2. Simulated data described in chapter 3; LEFT: true classification, MIDDLE: classification after initialization with spectra showing a clear spectral pattern, RIGHT: classification after initialization with spectra with ambiguous spectral pattern.

characteristics have to be known, which can easily be calculated to be

$$p(Y_i|Y_1,\ldots,Y_{i-1},Y_{i+1},\ldots,Y_N,\mathbf{X}) = \frac{1}{\hat{Z}} \exp\left(\mathrm{ssp}(Y_i|X_i) + \sum_{j:j\sim i} \mathrm{pp}(\mathrm{Y_i},\mathrm{Y_j}|\mathbf{X})\right) \tag{5}$$

with $\hat{Z}$ denoting the normalization constant.

3 Experiments

The method was tested both on simulated data providing ground truth for a quantitative analysis, and on real data in order to evaluate the practicability in the clinical setting. The artificial data set consisted of 93 simulated MRSI-data sets from three patients (representing three different tumor geometries) with 16% noise on a 64×64-grid (for details see [10]). The first two principal components were calculated in a leave-"one patient"-out fashion, and all spectra of the hold-out patient were projected onto these directions. To imitate the physician, three spectra, having a posterior probability of at least 95%, were randomly selected per class for initializing the kernel-density estimation. The hyper-parameter ν of the spatio-spectral coupling was optimized in an additional, internal cross-validation loop. For evaluation the spatio-spectral classification was repeated ten times with different initializations. For comparison, the classification without coupling ($\nu = 0$) was also tested.

The algorithm was also tested on 67 MRSI with a spatial resolution of 16×16 acquired from 14 patients under routine protocol during pre-therapeutic diagnostic and follow up on a 1.5T MR scanner at the German Cancer Research Center (dkfz), Heidelberg. Standard signal processing comprised Fourier transformation of the temporal resonance signal, water peak removal and phasing of the spectrum to its real part. Spectra containing artifacts were singled out using the NoN-score [17], and pair potentials in (2) involving these spectra were set to

zero. The spectra were projected onto the first two PCA-directions calculated from an independent, clinically validated set of spectra (for details see [18, 11]).

For the initialization of the algorithm on this clinical data set, two tumorous and two healthy spectra were hand-selected and labeled in each MRSI. The trade-off parameter ν depends on the spatial resolution of the MR scanner and the signal-to-noise ratio of the acquired data. As both can be assumed constant and since no ground truth was available, three MRSI slices from different patients were randomly selected, hand-labeled and ν fixed to the value that gave the smallest cross-validation error. Classification results were compared against the single-voxel results of the external classifier already used in [18].

4 Results

Using the model (3) on the simulated data with the iterative optimization procedure described in chapter 2.3 on the simulated data, a mean accuracy of 98.7% was obtained, with an average true positive rate of 97.5% for tumorous tissue (standard deviation $17.0 \cdot 10^{-4}$) and a true negative rate of 98.8% on healthy tissue (standard deviation $6.6 \cdot 10^{-4}$). The single-voxel classification without spatial regularization reached a mean classification accuracy of 98.2%, with a true positive rate of 94.3% for tumor (standard deviation $114 \cdot 10^{-4}$), and 99.1% for normal spectra (standard deviation of $5.4 \cdot 10^{-4}$). The spatial regularization increased the classification accuracy and reduced the variance, leading to better classification results especially on tumorous tissue. Comparing the average over 500 samples from the posterior distribution (3) with the MAP estimate, calculated via GraphCut, shows the ambiguity of the classification only along the tumor border, indicated by the blurred contours of the tumorous region (Fig. 3, fourth image).
The low SNR of the data led to speckle noise and misclassification in the single-voxel processing (Fig. 3), which was the main reason for the worse performance of this approach. Spectra well within the healthy region which were classified as tumorous (Fig. 3, third image) in the single spectrum approach, were classified correctly if the spatially coupled model was used for classification (Fig. 3, second image).

Adapting the algorithm to the different patients, by using the semi-supervised initialization is of major importance for the good classification results. Using spectra from a different patient in initializing the spectral model often led to disastrous results, to the extent that the tumor is not detected at all or unnecessarily large regions are classified as tumorous (Fig. 4). Using the patient-specific labels assigned by an expert as initialization for the spectral model guarantees an automatic adaption to patient variation and ensures high-quality classification, independent of patient characteristics. From the simulated data, we observe that the main advantages of the spatio-spectral classification are on the one side its ability to adapt optimally to the patient, by using the semi-supervised ini-

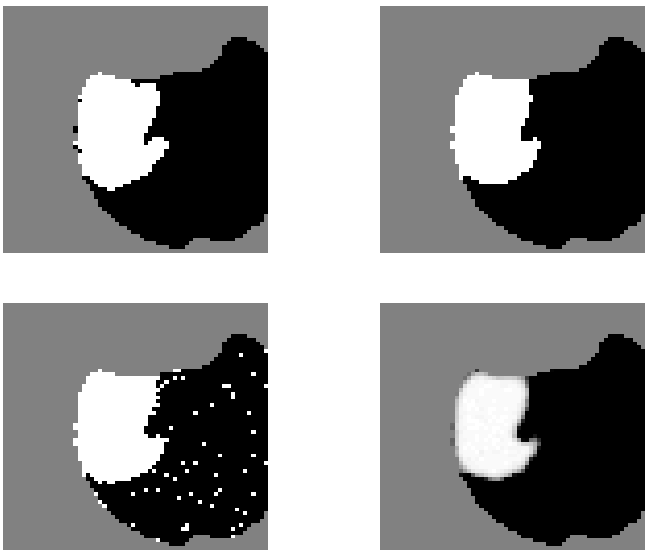

Fig. 3. Top Left: ground truth; Top Right: MAP-estimate for posterior distribution; Bottom Left: single-voxel based classification; Bottom Right: average over 500 samples from the posterior distribution in eq. (3).

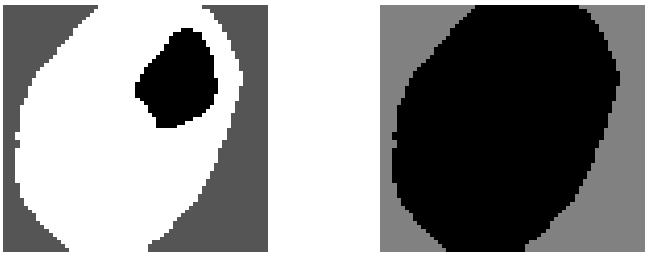

Fig. 4. Left: ground truth; Right: solution of the spatio-spectral model by initializing the density estimate with labeled spectra from a different patient.

tialization, which ensures a highly accurate and reliable classification. On the other hand, it is able to remove isolated misclassifications, depending on the distinctiveness of the spectral pattern. A similar result might be achieved by using morphological operators in a post-hoc processing step, but, in contrast to the proposed approach, all "label islands" will be removed, irrespective of the probability of the voxel belonging to the assigned class. Trading spatial smoothness of the probabilistic result map with the support of a different classification of neighboring spectra by their distance in feature space, is one of the main advantages of the presented approach.

On the real data set a high agreement between the results of the spatio-spectral and single voxel approaches could be observed in those voxels which were assigned a high confidence to one of the classes by the single voxel classifier. Voxels with a less stringent assignment were preferably classified according to their neighborhood by the spatio-spectral classifier. A detailed inspection of these cases showed that a main source for these contradicting results was a low quality of the respective spectra, e.g. caused by a low SNR or showing small shifts of the resonance lines. Visual inspection of the spectrum belonging, for example, to voxel (a) in the second image of figure 5 shows that a low data quality was the most likely reason for an assignment to the "intermediate" class by the single-voxel classifier of [18], while the spectral pattern was in fact "healthy", a label predicted also under a slight spatial regularization (Fig. 5, second image). As a second example, pixel (b) is surrounded by six voxels of tumorous tissue. The spatially regularized classification of the low quality spectrum leads to a distinct "tumor" assignment as opposed to "intermediate" by the single-voxel classifier. Here, the decision for "tumor" is in accordance with the visual inspection of the MR spectrum.

Comparing whole confidence maps of the single-voxel classification (Fig. 5, third image) and the spatio-spectral model (mean over 500 samples from the posterior (3), Fig. 5, fourth image), shows that a consideration of neighborhood information increases the confidence in the assigned labels on low quality data significantly.

Overall, we find that a main advantage of the spatio-spectral classification is its ability to adapt optimally to the individual spectral image, by using the semi-supervised initialization ensuring a highly accurate and reliable classification. Using the patient-specific labels assigned by an expert as initialization for the spectral model guarantees an automatic adaption to patient variation and a high-quality classification, independent of patient characteristics. In addition, the present approach is able to trade local support of a decision with the global support from the labels in its neighborhood. While in a standard post-hoc processing, for example, morphological operators remove all regions below a certain size in the result map, irrespective of the spectral information of the voxels belonging to these areas, the proposed spatio-spectral classification is able to

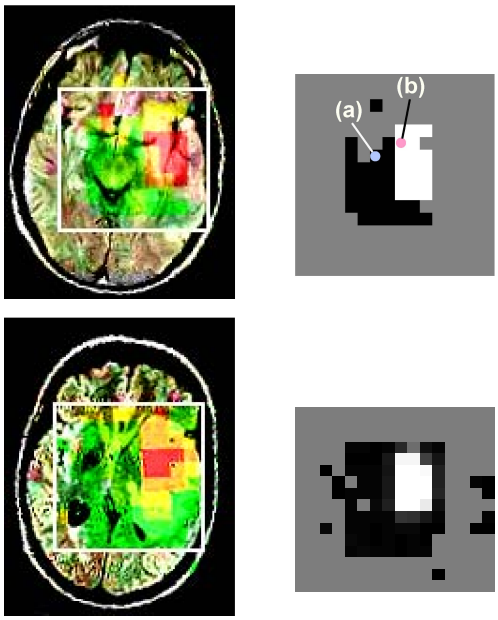

Fig. 5. Top, Bottom Left: Single-voxel classification according to [18], red indicating tumor, green indicating healthy, saturation indicating confidence in respective classification, Top Right: MAP estimate of spatially restricted model (eq. 3) for area indicated by white square in first image; Bottom Right: Average over 500 samples from the posterior distribution (eq. 3) for area indicated by white square in third image

remove isolated misclassifications depending on the distinctiveness of the spectral pattern. On MRSI it is thus able to differentiate between misclassifications resulting from low data quality, often resulting in random class assignments of single spectra, and strongly supported labels of isolated tumor voxels in an otherwise healthy neighborhood.

The model proposed in this chapter proves that adapting to patient characteristics can be efficiently incorporated into spatially regularized models operating both in spatial and spectral dimension of the magnetic resonance spectroscopic image. By using a DRF with a very versatile single-site potential, obtained from a class-wise kernel-density estimate, and a parameter-free penalty function, it is possible to use the few labels generated in standard clinical procedure to segment spectral images with optimal adaption to the patient. An iterative procedure using the GraphCut algorithm was introduced and the necessity of customization to the patient and usage of spatial information was shown.

References

1. Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N–D Image Segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109 – 131, 2006.
2. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124 – 1137, 2004.
3. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222 – 1239, 2001.
4. T. Brox, Y.-J. Kim, J. Weickert, and W. Feiden, "Fully-automated analysis of muscle fiber images with combined region and edge based active contours," in *Bildverarbeitung fr die Medizin 2006: Algorithmen – Systeme – Anwendungen*, 2006, pp. 86 – 90.
5. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likehood from incomplete data via the EM-algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1 – 22, 1977.
6. A. Devos, L. Lukas, J. Suykens, L. Vanhamme, A. Tate, F. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, C. Arús, and S. Van Huffel, "Classification of brain tumours using short echo time ^{1}H MR spectra," *Journal of Magnetic Resonance*, vol. 170, no. 1, pp. 164 – 175, 2004.
7. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721 – 741, 1984.
8. D. Greig, B. Proteous, and A. Seheult, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society, Series B*, vol. 51, no. 2, pp. 271 – 279, 1989.
9. G. Hagberg, "From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods," *NMR in Biomedicine*, vol. 11, no. 45, pp. 148 – 156, 1998.
10. B. M. Kelm and F. A. Hamprecht, "Trading resolution against noise in NMR spectroscopic imaging using conditional random fields," Interdisciplinary Center for Scientific Computing, University of Heidelberg, Tech. Rep., 2007.

11. B. M. Kelm, B. H. Menze, C. M. Zechmann, K. T. Baudendistel, and F. A. Hamprecht, "Automated estimation of tumor probability in prostate MRSI: Pattern recognition vs. quantification," *Magnetic Resonance in Medicine*, vol. 57, pp. 150 – 159, 2007.
12. V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147 – 159, 2004.
13. S. Kumar and M. Hebert, "Discriminative random fields," *International Journal of Computer Vision*, vol. 68, no. 2, pp. 179 – 201, 2006.
14. C. H. Lee, S. Wang, F. Jiao, D. Schuurmans, and R. Greiner, "Learning to model spatial dependency: Semi-supervised discriminative random fields," in *Neural Information Processing Systems 19*, 2007, pp. 793 – 800.
15. S. Z. Li, *Markov random field modeling in computer vision*. London, UK: Springer, 1995.
16. L. Lukas, A. Devos, J. Suykens, L. Vanhamme, F. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, A. Tate, C. Arús, and S. Van Huffel, "Brain tumor classification based on long echo proton MRS signals," *Artificial Intelligence in Medicine*, vol. 31, no. 1, pp. 73 – 89, 2004.
17. B. H. Menze, B. M. Kelm, M.-A. Weber, P. Bachert, and F. A. Hamprecht, "Mimicking the human expert: a pattern recognition approach to score the data quality in MRSI," Interdisciplinary Center for Scientific Computing, University of Heidelberg, Tech. Rep., 2007.
18. B. H. Menze, M. P. Lichy, P. Bachert, B. M. Kelm, H.-P. Schlemmer, and F. A. Hamprecht, "Optimal classification of long echo in vivo magnetic resonance spectra in the detection of recurrent brain tumors," *NMR in Biomedicine*, vol. 19, no. 5, pp. 599 – 609, April 2006.
19. A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naïve Bayes," in *Advances in Neural Information Processing 14*, 2001, pp. 605 – 610.
20. P. Perona and J. Malik, "Scale-Space and edge detection using anistropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629 – 639, 1990.
21. C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Secaucus: Springer, 2004.
22. B. Silverman and B. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall, 1986.
23. A. Tate, C. Majós, A. Moreno, F. Howe, J. Griffiths, and C. Arús, "Automated classification of short echo time in in vivo ^{1}H brain tumor spectra: A multicenter study," *Magnetic Resonance in Medicine*, vol. 49, no. 1, pp. 29 – 36, 2003.