Segmenting Brain Tumors using Alignment-Based Features

Mark Schmidt, Ilya Levner, Russell Greiner
University of Alberta
Department of Computing Science
Edmonton, Alberta, T6G 2E8, CANADA
schmidt@cs.ualberta.ca

Abstract

Detecting and segmenting brain tumors in Magnetic Resonance Images (MRI) is an important but time-consuming task performed by medical experts. Automating this process is a challenging task due to the often high degree of intensity and textural similarity between normal areas and tumor areas. Several recent works have explored aligning a spatial 'template' image in order to incorporate spatial anatomic information, but it is not obvious how this alignment should be used. This work quantitatively evaluates the performance of 4 different types of Alignment-Based (AB) features encoding spatial anatomic information for use in supervised pixel classification. This is the first work to (1) compare several types of AB features, (2) explore ways to combine different types of AB features, and (3) explore combining AB features with textural features in a learning framework. We considered situations where existing methods perform poorly, and found that combining textural and AB features allows a substantial performance increase, achieving segmentations that very closely resemble expert annotations.

1 Introduction

Radiation Oncologists and many other medical experts currently spend a substantial portion of their time segmenting medical images, and in particular, labeling brain tumors and associated edema in MRI. This labeling is especially important when planning radiation therapy, and new treatment techniques such as Intensity-Modulated Radiation Therapy will soon drastically increase the segmentation workload [1]. The introduction and acceptance of these types of new technology will thus depend on reducing the manual time associated with detecting and segmenting tumors.

Over the past 15 years, there has been an immense amount of research focusing on semi-automatic and fully automatic methods to perform brain tumor detection and/or segmentation in MRI, but the tools used in practice still rely on significant manual interaction. There have been manually engineered automatic systems designed to perform this task, but even the most effective of these address simplified cases (such as the high-contrast 'enhancing' pixels [2]), do not achieve sufficiently accurate results for difficult cases (ie. an average score across patients of 0.597 in [3], see Table 2), and depend highly on the local imaging hardware available to the designer. This is primarily due to the complex nature of the data, and the difficulty in translating complex anatomic knowledge into a series of operations that have good generalization properties.

There has been significant exploration of this problem from a Machine Learning (ML) perspective, by formulating the problem as a pixel classification task (illustrated in Figure 1). A Machine Learning solution would remove dependencies on the specific imaging hardware used (as it could be trained from anywhere), and could potentially have better generalization properties than manually engineered systems for this task. However, Machine Learning approaches applied to this problem have had limited success. This is a result of the pixel intensities violating the 'independent and identically distributed' (iid) assumption within and between images, and of the high similarity that can be observed between some normal pixels and some tumor pixels. The iid assumption is violated due to the spatial nature of the data, and the corruption of MRI intensities by noise, inter-slice intensity variations, an intra-volume intensity bias field, and differences in the intensities recorded for identical tissues between volumes. The latter problem is the most severe, and proposed ML methods often require training data from the image to be segmented (termed *patient-specific* training) [4].

Some recent ML systems have incorporated techniques to correct (or reduce) these effects to allow *inter-patient* training [5, 6, 7], making the methods fully automatic. However, disambiguating normal and tumor regions with similar or identical intensity properties remains an unsolved

problem. Incorporating multi-scale or textural information can aid in discriminating tumors from normal regions [4, 8], but these types of approaches have limited success since they do not account for a pixel's spatial anatomic location. Recent systems have thus explored the use of a spatially aligned 'template' image to incorporate spatial information [5, 6, 7, 9, 10]. However, it is not obvious how best to use the alignment. Each of the Alignment-Based methods so far have only explored a single type of Alignment-Based (AB) information, and moreover, none have used learning to combine AB features with textural information during classification.

In this work, we seek to quantify the performance of different types of AB features for performing automatic brain tumor segmentation as a pixel classification problem. This is the first work to explore (and compare) multiple types of AB features (we quantified 4 different types), and the first work that explores combining multiple types of AB features. Moreover, this is also the first work that explores using Machine Learning to combine AB features with textural features. Our empirical results show that this gives a substantial improvement over other proposed feature sets, producing for the first time a fully automatic system that may actually be used in practice.

Section 2 will briefly introduce the preprocessing steps we used, outlining our spatial alignment and intensity normalization steps that allow *inter-patient* training and the use of AB features. Section 3 presents the AB and textural features we examined, while Section 4 outlines our segmentation method. Section 5 presents our quantitative experimental results.

2 Preprocessing

The preprocessing phase has two purposes: It makes the intensities of identical tissue types more consistent within and between images, and spatially aligns the images with a template image. This was done in a processing pipeline as in [5]. We used Statistical Parametric Mapping implementations for the spatial registration and resampling stages [11], used the T1 single subject template from [12], and used existing methods for the intensity normalization stages. Our pipeline consisted of:

1. Noise Reduction (Non-Linear Filtering [13]).

2. Inter-Slice Intensity Variation Correction (Weighted Regression [14]).

3. Intra-Volume Intensity Bias Field Correction (Non-parametric Nonuniform intensity Normalization [15]).

4. Alignment of the different modalities (Maximization of Normalized Mutual Information).
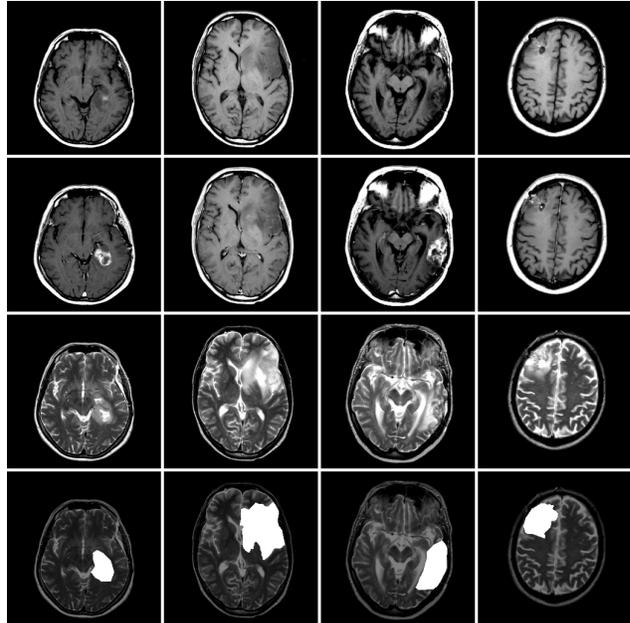
Figure 1. Examples of Input and Desired Output for 4 patients. Top to bottom: T1 images, T1 images after contrast agent injection, T2 images, Manual Annotations overlaid on T2 images.

5. Linear Alignment of the modalities with the template (*Maximum a Posteriori* formulation).

6. Non-Linear Warping of the modalities with the template (*Maximum a Posteriori* formulation).

7. Resampling of the voxels to the template coordinate system (β-Splines).

8. Inter-Volume Intensity Standardization (Weighted Regression [14]).

Creating this pipeline required a great deal of experimentation. Many steps in this preprocessing pipeline represent open research problems, and performing these steps can be complicated significantly by the presence of large tumors (as opposed to relatively small lesions in [5]). We selected methods for steps 1-4 and 7 that are not affected by the presence of tumors. However, template alignment (5-6) can be affected by the presence of large tumors, since they will not match corresponding regions in the template. To take this into account, we increased the regularization parameter (prior) used for these steps, which produced acceptable alignments without significantly distorting the abnormal areas. For (8), we used our Intensity Standardization method [14], that uses left-to-right symmetry to confer robustness to areas of abnormality. Examples of the input and output of this pipeline are shown in Figure 2.
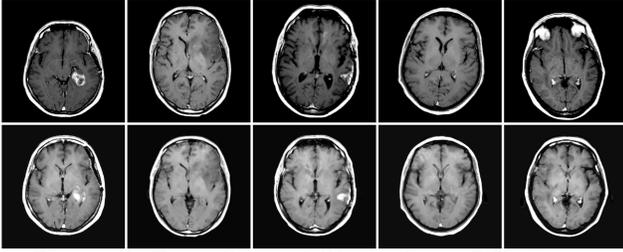
Figure 2. Top: Central Slices from 5 Patients before preprocessing. These images are misaligned, the brains have different shapes, and there are significant intensity variations between images. Bottom: Central Slices from the same volumes after using our preprocessing pipeline for intensity and spatial normalization to reduce these (and other) effects.

3 Features Extraction

We evaluated 4 different types of AB features (shown in Figure 3). The first AB feature reflected the fact that by definition brain tumors only occur in the brain: it was a binary mask of the template image's brain area, that was smoothed to reflect uncertainty in the exact position of the brain boundary in spatially aligned patients. This was obtained from [11], and we hypothesized that smoothing could potentially be more effective than the distance transform used in [9], since it means that pixels inside the brain but near the edge will receive decreased values. This is clearly advantageous compared to performing a separate automatic brain extraction step as in [5], which could potentially remove tumor pixels near the edge of the brain.

The second type of AB features were spatial likelihoods for the 3 normal tissue types obtained from [12]. This type of AB feature was shown to enhance segmentation in [5]. The third type of AB feature was the average intensity maps from a set of individuals aligned with the template coordinate system (obtained from [12]). We hypothesized that these could enhance performance since a large difference from the expected intensity value at a pixel location could indicate abnormality.

The final AB feature was a characterization of left-to-right symmetry. Tumors are typically asymmetric around this axis while normal areas are typically symmetric. The spatial alignment of the template allowed us to use the template's known line of symmetry as an approximation of the line of symmetry in an aligned image. We thus characterized left-to-right symmetry by subtracting the intensity value of the pixel on the opposite the side of the line of symmetry from the pixel's own intensity value (recall that the intensities have already been normalized).

To characterize pixel neighborhood and textural properties, we used a multi-scale feature representation. This consisted of including, for each feature, the pixel-level feature
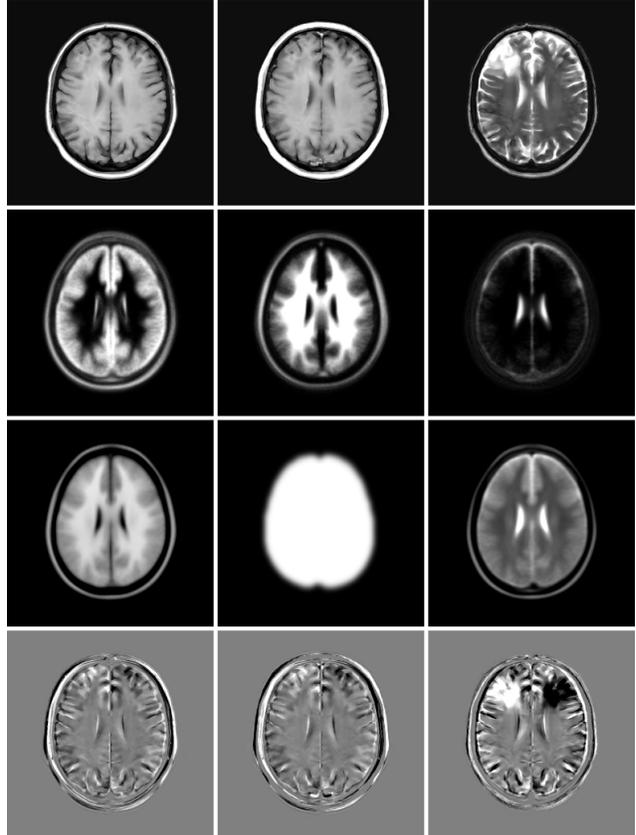
Figure 3. AB Features. Row 1: Intensity and Spatially Normalized Input Images. Row 2: Normal Tissue Spatial Priors. Row 3: Expected Intensity Spatial Maps and Smoothed Spatial Brain Mask (middle). Row 4: Left-to-right Symmetry (scaled).

value, and the values after filtering the feature image with two Gaussian filters. For our experiments, we used a standard deviation of 2.25 for the first filter, and 4.25 for the second (the in-plane size of the pixels was $1mm$ by $1mm$ after resampling). Figure 4 shows examples of the feature images generated from this process. These multi-scale features directly encode local and neighborhood intensity information, but their differences form differences of Gaussians (similar to the Laplacian of Gaussian filter [16]), and thus these feature also implicitly encode local image gradient information.

4 Segmentation

For pixel classification, we used a soft-margin Support Vector Machine (SVM) trained using the SVMlight optimization strategy [17]. Since this task can involve millions of training and testing instances with a relatively small feature set, we used the linear kernel, which allows more efficient training than other kernels and allows much more
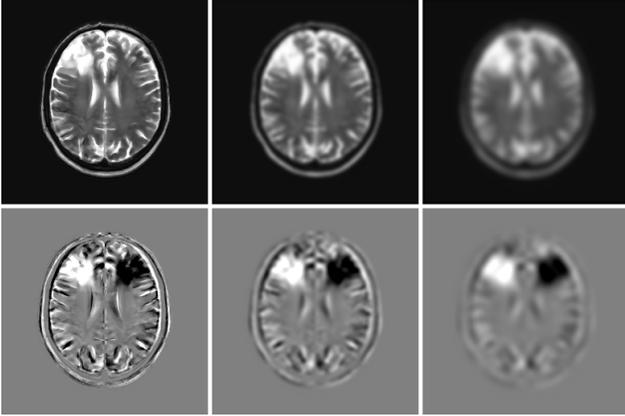
Figure 4. Examples of Multi-Scale Features. Top row: T2 image. Bottom Row: T2 left-to-right symmetry.
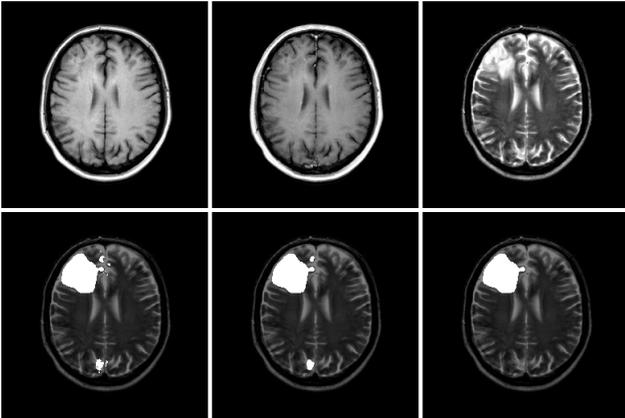
Figure 5. Segmentation Process. Top: Input Images. Bottom: left to right: SVM pixel classification results with the (PAS) feature set, the median root image of this result, the selected largest cluster.

efficient classification of new instances (through the use of the primal formulation).

To take into account dependencies in the labels of spatially adjacent pixels, we post-processed the classification results by repeatedly applying a median filter to the predicted discrete class labels, until convergence to the median root of the classification image. This removes isolated labels, while preserving (and smoothing) the edges of the segmented structures. From the median root, we selected the largest connected cluster as the final segmentation result (although the most effective feature sets tended to produce a single connected segmentation). Figure 5 shows an example of this segmentation process.

5 Results

In our experiments, the task was to produce a binary segmentation of the (heterogeous) tumor and edema region, with ground truth defined by an expert radiologist. Our data set consisted of 10 patients with brain tumors covering 4 different types of tumors. Moreover, one patient had a surgical cavity where the tumor had recently been removed (the correct label for the cavity was edema), and the data came from two different MRI machines. This diverse scenario, combined with the fact that we performed *inter-patient* testing (training on 9 patients and testing on 1), makes this a very difficult (but practical) scenario, where existing methods proposed to perform this task will perform poorly. Since we sought to quantify the effectiveness of the different AB features, the preprocessing and segmentation steps were kept identical, and only the feature set used for pixel classification was varied. To quantify the quality of the segmentations, we used the Jaccard similarity measure between the ground truth and the segmentation results in terms of true positives (tp), false positives (fp), and false negative (fn) pixels of the image:

$$J = \frac{tp}{tp + fp + fn} \qquad (1)$$

The feature sets evaluated in this work consisted of an intensity-only set (), and this feature set augmented with all combinations of the 4 types of AB features: The spatial brain mask (B), the normal tissue probabilities (P), the average intensities (A), and left-to-right symmetry (S). We measured each of these at the pixel-level and at multiple scales. The average test scores for the 10 patients from these feature sets are shown in Table 1. Since the segmentation scores for some patients over the different feature sets were generally higher than for others, we adopted a Student's t test of *paired samples* [18], and assumed a difference with probability 0.05 or less was significant.

Among the pixel-level sets with 1 AB feature, the (P) set performed significantly better than the () set. This was not surprising and agrees with [5]. Our proposed (A) set also performed significantly better than the () set, but (B) and (S) did not. Combining any of P, A, or S gave higher scores than each individually, and combining all three, (PAS), significantly outperformed the best individual type of AB feature, (P).

For each feature set, the scores of the multi-scale features were significantly higher than the corresponding pixel-level features. This is supported by the literature for the () set, but learning to combine multi-scale AB features has not been previously performed. In the multi-scale case, (P) did not significantly improve on the multi-scale (), whereas (A) and (S) did. Combining any of P, A, or S again resulted in higher scores than any individually for the multi-scale case,

Table 1. Average Jaccard scores for the different feature sets.

Feature Set	Pixel-Level	Multi-Scale
()	0.470	0.601
(B)	0.499	0.600
(P)	0.602	0.693
(A)	0.592	0.695
(S)	0.490	0.721
(BP)	0.574	0.705
(BA)	0.578	0.701
(BS)	0.581	0.726
(PA)	0.606	0.719
(PS)	0.621	0.727
(AS)	0.619	0.725
(BPA)	0.581	0.720
(BPS)	0.589	0.728
(BAS)	0.587	0.728
(PAS)	0.627	0.732
(BPAS)	0.593	0.732

and the combination of all three, (PAS), significantly outperformed the best individual type of AB feature at multiple scales, (S). Note that symmetry seems to be more effective when measured at multiple scales. This is likely due to the robustness to minor asymmetries that could be conferred through the addition of the coarse-scale features.

Our results support the hypothesis that multiple AB features can be advantageous, and that AB features can be combined with textural features to offer improved results. Our second experiment compared this idea to several recently proposed feature sets. We evaluated the intensity-based feature set of [19] (IO), the von Neumann intensity neighborhood of radius 6 proposed in [4] (vNN), the use of the intensities and the 3 normal tissue spatial priors as in [5] (IPr), the Diagonalized Nearest Neighbor Pattern Matching features at 4 scales as in [10] (using a database of 25 normal brains) (DNN), and the multi-scale illumination- and rotation-invariant MR8 textures from [20] $(MR8)$. The average results from these feature sets are compared to the multi-scale (PAS) feature set in Table 2. The multi-scale (PAS) set significantly outperforms each of the other feature sets except the $(MR8)$ textures, where the probability of observing the difference was 0.10 (as opposed to 0.05 or less). Interestingly, the $MR8$ textures seemed to make different mistakes than the other methods and due to this, note also that we could not establish that the $MR8$ features significantly outperformed the multi-scale () feature set.

Table 2. Comparison between feature sets proposed in the literature and the most effective set evaluated in this work (in bold).

Feature Set	Average Score
IO	0.470
vNN	0.598
IPr	0.602
DNN	0.545
MR8	0.636
(PAS)	0.732
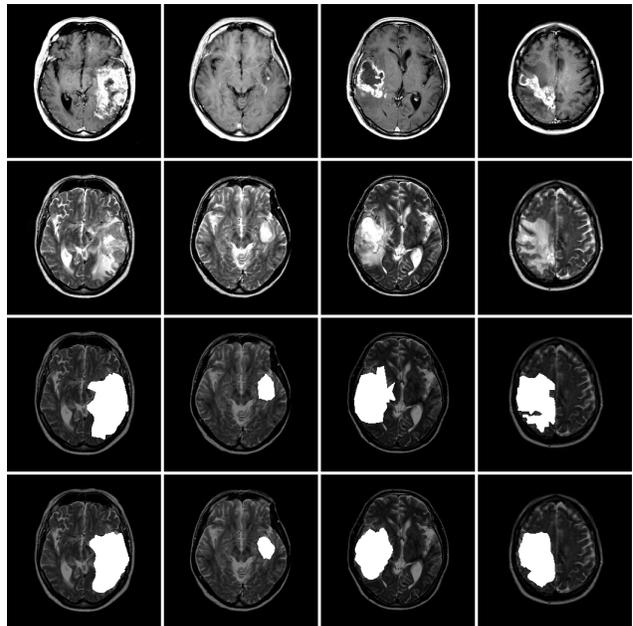
Figure 6. Comparison of Manual and Automatic Segmentation for 4 patients. Row 1: T1 image after contrast agent injection. Row 2: T2 image. Row 3: Expert Segmentation. Row 4: Segmentation with the (PAS) feature set.

6 Discussion

Further exploring the anomalous $MR8$ feature set is an interesting direction of future research. It is possible that more accurate results could be obtained by computing the $MR8$ texture features for the P, A, and S features, although this would constitute a much higher computational expense than our multi-scale method (48 filtering operations for each pixel-level feature instead of 2). The difference in the trend observed across patients for the $MR8$ features might also indicate that these textural features might complement other proposed methods of texture characterization. Another di-

rection of future research could be the use of feature selection. For example, it may not be beneficial to use left-to-right symmetry at the pixel level due to areas of minor asymmetry. Our results might also be improved by replacing the median root filter with a Markov Random Field as used in [7] or a Level Set algorithm as used in [6]. Finally, the use of other AB features might also improve results, and we are currently exploring the exploitation of a large database of normal brains for this purpose.

Contribution: This paper presented a fully automatic method to segment brain tumors and edema. While previous ML-based systems proposed for this task have explored a single type of AB information, this paper evaluated 4 different types of AB features, and combinations of these different types of AB features. This was also the first work that explored using Machine Learning to combine AB features with textural features. A very challenging *inter-patient* testing scenario with multiple tumor types was used, where existing methods are not adequate. Our results indicate that the use of multiple types of AB features can offer a significant increase in performance ($p > 0.95$), and that a further significant performance increase can be gained by using multi-scale AB features ($p > 0.95$). We showed that this latter technique significantly outperforms other recently proposed feature sets ($p = 0.90$), and gives results that are quantitatively and qualitatively very similar to expert annotations.

References

[1] B.S. The, S.Y. Woo, and E.B. Butler. Intensity Modulated Radiation Therapy (IMRT): A New Promising Technology in Radiation Oncology. *Oncologist*, 4(6):433–442, 1999.

[2] M.C. Clark, L.O. Hall, D.B. Goldgof, R. Velthuizen, F.R. Murtagh, and M.S. Silbiger. Automatic tumor segmentation using knowledge- based techniques. *IEEE Transactions on Medical Imaging*, 17:238–251, April 1998.

[3] L.M. Fletcher-Heath, L.O. Hall, D.B. Goldgof, and F. Reed Murtagh. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine*, 21:43–63, 2001.

[4] C. Garcia and J.A. Moreno. Kernel based method for segmentation and modeling of magnetic resonance images. *Lecture Notes in Computer Science*, 3315:636–645, Oct 2004.

[5] A.P. Zijdenbos, R. Forghani, and A.C. Evans. Automatic "pipeline" analysis of 3-d mri data for clinical trials: application to multiple sclerosis. *IEEE Transactions on Medical Imaging*, 21(10):1280–1291, Oct 2002.

[6] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig. A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 8(3):275–283, September 2004.

[7] D.T. Gering. *Recognizing Deviations from Normalcy for Brain Tumor Segmentation*. PhD thesis, MIT, 2003.

[8] S. Dickson and B.T. Thomas. Using neural networks to automatically detect brain tumours in MR images. *International Journal of Neural Systems*, 4(1):91–99, 1997.

[9] M.R. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz, and R. Kikinis. Automated segmentation of MR images of brain tumors. *Radiology*, 218:586–591, 2001.

[10] D.T. Gering. Diagonalized nearest neighbor pattern matching for brain tumor segmentation. *R.E. Ellis, T.M. Peters (eds), Medical Image Computing and Computer-Assisted Intervention*, 2003.

[11] Statistical parametric mapping, http://www.fil.ion.bpmf.ac.uk/spm/, Online.

[12] Icbm view: an interactive web visualization tool for stereotaxic data from the icbm and other projects, http://www.bic.mni.mcgill.ca/icbmview/, Online.

[13] S.M. Smith and J.M. Brady. Susan - a new approach to low level image processing. *Int. Jounral of Computer Vision*, 23(1):45–78, May 1997.

[14] M. Schmidt. A method for standardizing mr intensities between slices and volumes. Technical report, University of Alberta, 2005.

[15] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17:87–97, February 1999.

[16] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.

[17] T. Joachims. Making large-scale svm learning practical. In B. Scholkopf, C.J.C. Burges, and A.J. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1999.

[18] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press, 1998.

[19] J. Zhang, K. Ma, M.H. Er, and V. Chong. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. *International Workshop on Advanced Image Technology*, pages 207–211, 2004.

[20] M. Varma and A. Zesserman. Classifying images of materials: Achieving viewpoint and illumination independence. *Lecture Notes in Computer Science*, 2352:255–271, 2002.