Comparative Analysis of Discriminant Wavelet Packet Features and Raw Image Features for Classification of Meningioma Subtypes

Hammad Qureshi[a], Nasir Rajpoot[a], Roland Wilson[a], Tim Nattkemper[b] and Volkmar Hans[c] *

[a]Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK
[b]Applied Neuroinformatics, University of Bielefeld, 33615 Bielefeld, Germany
[c]Institute of Neuropathogy, Evangelisches Krankenhaus, 33617 Bielefeld, Germany

Abstract. The idea of multiresolution analysis has been around for over two decades now. In this paper, we explore a multiresolution analysis based technique for histopathological image classification and compare it with raw image analysis. The principle idea for the former is to derive an optimal wavelet representation, called Adaptive Discriminant Wavelet Packet Transform (ADWPT), of the images in order to achieve the largest discrimination power. Our results show that the textural features combined with ADWPT yield a significant improvement in classification accuracy.

1 Introduction

One of the most important issues that has hampered the solution to the problem of image classification has been efficient pattern representation. Many researchers have argued that to achieve good image classification results, first an efficient representation of the image classes under consideration must be derived. It is widely accepted that the human visual system performs a transformation of the input image into a spatial frequency decomposition before the brain actually performs the recognition task. The idea of optimal representation has been researched widely and various techniques have been proposed. In this paper, we present an optimal, in terms of its discriminatory power, wavelet decomposition for the problem of histopathological image classification and demonstrate that it performs better than image statistics based analysis.

Histopathological diagnosis of tumours, especially of the brain and spinal cord, even today requires decision making by human experts. Manual diagnosis and decision making is hampered by two limitations: First, reviewing histological slides by humans is time consuming and the human experts are not always available. Secondly, although a lot of effort has been made to exactly define diagnostic criteria for all tumour entities within the World Health Organization (WHO) Classification of Tumours [1] but the inter-rater variability is still considerable (see e.g. [2]). This consequently influences further therapy regimens greatly and hence a bias is introduced. Due to the fast progress in digital image retrieval and analysis technologies, machine based decision making may be used to substantially support histopathologists by providing more objective diagnostic parameters and allow for high-throughput analysis. A first step, however, towards developing new algorithms for image classification (i.e., 'diagnosis') is to test whether the automated technique is able to reproduce human assignment of single tumour samples to diagnostic classes. In order to develop such a technique, we have focussed on meningiomas which are tumours of the brain and the spinal cord arising from cells of the surrounding (i.e. meningeal) coverings of the brain. Meningiomas account for 20% of all brain tumours and exist in three different grades of malignancy (WHO Grad I-III), most being benign (over 80%), but some showing an increased propensity to recurrence and rare cases being malignant. Most benign WHO Grade I meningiomas belong to one of four subtypes: Meningiothelial, fibroblastic, transitional, and psammomatous. Figure 1 shows the four meningioma subtypes and their salient features. Correct histopathological diagnosis can be made in most cases by a trained human expert, i.e. a neuropathologist. Therefore, this tumour is well suited for testing diagnostic properties of new machine based pattern recognition algorithms used in histopathological analysis.

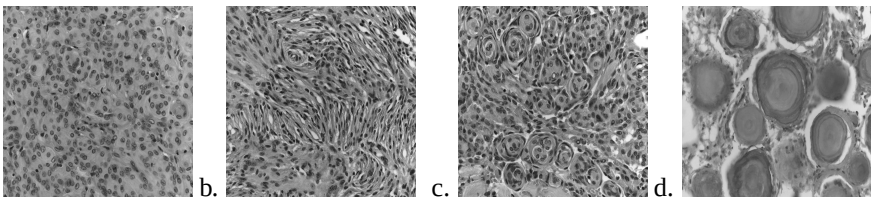

Figure 1. Meningioma Images at 40x magnification for each subtype a. Meningiothelial (cells form synctium), b. Fibroblastic (spindle shaped cells in collagen-rich matrix), c. Transitional (cells form whorls with psammoma bodies), d. Psammomatoes (high number of psammoma bodies)

Numerous studies [5–7] have successfully demonstrated the utility of wavelets and other transforms for image classification. Lessman et al. [3] studied the problem of content-based visualization of meningioma images to aid in

*contact author: hammad@dcs.warwick.ac.uk

characterization of the database contents using the wavelet transform. In our previous work [4], we presented the results of preliminary investigation into meningioma subtype classification using wavelet packets. Saito and Coifman [8] showed how wavelet packet transform may be used for local features extraction using relative entropy as the criterion for basis selection. Rajpoot [9] developed discriminant wavelet packets for texture classification. Support Vector Machines (SVMs) have been employed frequently in the literature to perform pattern classification. In [10], Nattkemper et al. demonstrated the ability of SVM by proving that they perform better than decision trees and nearest neighbor (k-NN) classifiers in breast tumor classification.

This paper presents a technique called the Adaptive Discriminant Wavelet Packet Transform (ADWPT) for feature extraction and image classification and compares its performance with classification using Haralick [11] features from raw images. The AWPT subbands can be thought of as descriptors [12] which can be used for image compression but they may also be employed as features for texture-based image classification. Our experimental results demonstrate the strength of an ADWPT based image representation for meningioma subtype classification.

2 Materials and Methods

The image acquisition process involved the analysis of the routinely stained histopathological slides on a Zeiss Axioskop 2 plus microscope with a Zeiss Achroplan 40x/0,65 lens to obtain 1300×1030 pixels, 24 bit, true color RGB pictures. Representative images of the four meningioma subtypes considered in this study are shown in Figure 1. These images were converted from their RGB values to 8-level grayscale images for analysis and cropped to 1024×1024 pixels. Each image is further divided into 4 parts of 512×512. In this paper, we compare two approaches first using raw image features and then wavelet packet features (refer Figure 3) to classify amongst patterns.

2.1 Raw Image Based Analysis

In image based analysis, we obtain statistical features described in section 2.3 after dividing a given image into non-overlapping slabs of 32 by 32, 64 by 64 and 256 by 256 pixels. Each slab is taken as a separate entity for features extraction. There were a total of 256 features for 32×32, 64 features for 64×64 and 4 features for 256×256. A divided image is shown in Figure 2a.

2.2 ADWPT-based Analysis

The first step to any statistical image classification technique is the computation of features to be fed into a classifier. It is of paramount importance that the features acquired are relevant and characterize the underlying image classes. As opposed to the ordinary wavelet transform, the ADWPT adaptively decomposes both the high frequency and the low frequency components of a signal at multiple resolutions. ADWPT is based upon the wavelet packet transform [13]. The computation of ADWPT decomposition is explained in the following sections.

2.2.1 Adaptive Discriminant Wavelet Packet Transform (ADWPT)

The first stage of our technique is the computation of the best wavelet packet basis in terms of the discrimination power of the four meningioma subtypes. This is important because our aim is to obtain a representation of the image which is effective in localizing the spatial frequencies and comparing the most different aspects of the images.

Let X be a set of all the WPT subbands at all possible resolutions. For an ADWPT based image classification, the problem can be posed as the selection of an optimal combination of subbands from a subset Y of X such that (*a*) it yields a complete basis for any given image and (*b*) it has the largest discriminatory power of all other combinations that can yield a complete basis. The aim of ADWPT basis selection process then is to obtain Y of cardinality n which achieves the best classification performance such that $n < N$ and $Y \subset X$. The process is divided into two steps:

Full Wavelet Packet Transform (FWPT)

The wavelet transform is computed by applying a highpass and lowpass filter upon the input signal to acquire the high frequency and low frequency subbands. The wavelet packet transform is performed iteratively on every subband to a certain level to obtain the high and low frequency coefficients at each stage. Without loss of generality, the transform for a 1D discrete signal $\mathbf{x} = \{x_i\}, i = 0, 1, ..., N-1$ may be computed as follows:

$$S_0^0(i) = x_i \quad (1)$$

$$S_j^{2p}(i) = \sum_k g_{k-2l} S_{j-1}^p(k) \quad (2)$$

$$S_j^{2p+1}(i) = \sum_k h_{k-2l} S_{j-1}^p(k) \quad (3)$$

where $i = 0, ..., N-1$, $l = 0, ..., 2^{-j}N-1$, g_n and h_n are lowpass and highpass filters respectively, $j = 1, 2, ..., J$; $J = log_2 N$, $S_j^p(l)$ is the transform coefficient corresponding to the wavelet packet function having relative support size 2^j, frequency $p2^j$ and is located at $l2^j$. Hence, j, p and l are regarded as the scale, frequency and position indices of the wavelet packet function. The transform is invertible if appropriate dual filters $\tilde{g}_n$ and $\tilde{h}_n$ are applied on the synthesis side.

A simple wavelet transform decomposes only the low level frequencies iteratively whereas wavelet packets involves decomposing all the subbands. The wavelet packet decompositions are maintained in a quadtree structure, with the parent being the original subband or image and the children being the wavelet decompositions of the parent. First the image is decomposed into its respective subbands and then each subband is decomposed further until a predefined maximum depth of the tree is reached. This results in a combinatorial explosion of possible wavelet packet bases that can be used to completely represent the image and its called the Full Wavelet Packet bases. The next stage is the selection of the best bases which is achieved with the help of dynamic programming.

Computation of Discrimination Power

First, a pseudo probability density function (pdf) is obtained for each subband using the normalized energy for the subband coefficients. A psuedo probability density function is computed by dividing the square of a coefficient by the sum of the square of the coefficients in a subband. The next step is to compute the average pseudo probability density functions (apdf) by summing the pseudo pdfs pairwise of the training images belonging to the same class and dividing them by 2.

$$\mathcal{A}(X_a, X_b) = (x_i^a + x_i^b)/2$$

X_a and X_b are the pseudo pdfs of two samples belonging to the same class. The process is repeated for all the subbands of the training images and it is important to note that the average of two sample images is computed per iteration. The objective is to acquire a basic model of localized frequency values for each class so that the difference between the classes may be estimated. Then the discriminating power of each subband is obtained, using the Hellinger distance between different classes, as follows.

$$\mathcal{D}(X, Y) = \sum_i (\sqrt{x_i} - \sqrt{y_i})^2$$

where x_i and y_i are the normalized energy of the ith subband coefficient for two training images of different classes. This distance is used as the discriminatory power $\mathcal{P}$ in the best basis selection process. The pairwise distance between the 4 different kinds of textures is obtained based upon the following formulae.

$$\mathcal{P}_k = \sum_{i=1}^{3} \sum_{j=i+1}^{4} \mathcal{D}(X_i^k, X_j^k)$$

where i and j represent the different class indices.

Best Basis Selection

1. Compute the J-level full wavelet packet tree decomposition.
2. Initialize $j = J - 1$.
3. For all $0 \leq p < 2^j$, $0 \leq q < 2^j$, do the following:

   a If $\mathcal{P}_{p,q}^j < max[\mathcal{P}_{2p,2q}^{j+1}, \mathcal{P}_{2p,2q+1}^{j+1}, \mathcal{P}_{2p+1,2q}^{j+1}, \mathcal{P}_{2p+1,2q+1}^{j+1}]$ keep the four child subbands at depth $j+1$ where $\mathcal{P}_{p,q}^j$ represents the discrimination power of a node at position i, j

   b otherwise keep the parent at depth j and remove the child subbands.

4. Decrement j by 1.
5. If $j < 0$, then stop, otherwise goto step 3.

One important fact to mention here is that when we apply the ADWPT on different training or input data, in most cases we get the same decomposition. Figure 2b shows the decomposition obtained for four different data sets which implies that the decomposition remains consistent even when different subsets of texture data are used.

2.3 Statistical Features Extraction

Once the ADWPT of the images and the divided raw images are obtained, the next step is the computation of the statistical features. Gray co-occurrence based features proposed by Haralick [11] are used in our technique. We use texture correlation, contrast, homogeneity and energy.

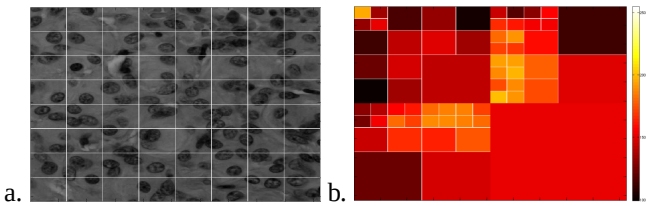

Figure 2. a. Raw image divided into 64x64 pixel slabs, b. Best basis decompositions for 4 different data sets with color denoting the discrimination power of the subbands

2.4 Classification using Support Vector Machines

SVM is a supervised classifier, which approximates the decision surfaces of the theoretical Bayes classifier. SVM has found broad area of applications since its invention in 1995 by Vapnik [14]. SVM uses various kernels to map the input space into a higher dimensional feature space to make the non-linear hyperplane linear. To achieve this without increasing computational complexity a kernel trick is employed.

In the training phase, the statistical features based on the gray level co-occurrence matrix of the subbands (in case of ADWPT, image at each resolution is called a subband) and slabs (in case of image blocks) are used as input vectors. Multiple SVMs are trained and feature vectors are used as training data. The Matlab version of LibSVM provided by Chang and Lin [15] was used in our analysis. The results were obtained using a Gaussian SVM. Figure 3 depicts the algorithms employed in this study.

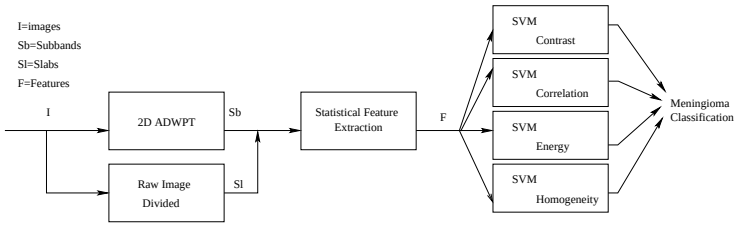

Figure 3. Block diagram of our technique

3 Results and Discussion

The experimental setup consisted of employing the Daubechie 8 tap filter for obtaining the ADWPT upto four levels. The averaged results of the various test runs are shown in the Table 1. The results presented have been cross-validated with four different test trial runs using different training and testing data sets. A total of 64 length features set is derived using ADWPT and the raw image based analysis yields feature sets of length 256 (32×32 slabs), 64 (64×64 slabs) and 4 (256×256 slabs).

In all 320 image samples were available as input data for training and testing. 80 image samples for each subtype of meningioma were available for experimentation. The data is divided into five batches per meningioma subtype with each batch containing 16 images. During the training and testing phase a leave one out mechanism is observed which means that 1 batch is kept for testing while the rest are used for training.

Feature	F_r	M_r	P_r	T_r	$Overall_r$	F_a	M_a	P_a	T_a	$Overall_a$
Contrast	32.8	29.7	93.8	28.1	46.1	90.7	62.5	98.5	76.6	82.1
Correlation	0	0	0	100	25	75	68.7	96.9	73.4	78.5
Energy	54.7	34.4	89.1	29.7	51.9	56.2	73.4	100	82.8	78
Homogeneity	46.8	39	89.1	23.5	49.6	54.7	67.2	96.9	81.3	75

Table 1. Raw image based (denoted by subscript r) vs. ADWPT features (denoted by subscript a) %age classification accuracy results (correct classification %age averaged over 4 trials) of meningioma images using SVM and gray-level co-occurrence features (F=Fibroblastic,M=Meningiothelial,P=Psammomatous,T=Transitional)

Table 1 shows the average results for various trial runs using statistical features over raw images and ADWPT based representation. The results for 32×32 and 256×256 slabs were also obtained but there was no significant improvement

in performance, hence, only the results for 64×64 are shown. Our results show that ADWPT performed consistently better than raw image features. It is also notable that for the psammomatous subtype both techniques perform consistently well. The primary reason for this, is that the psammomatous texture is significantly different from the other textures included in the study since it contains dispersed psammoma bodies and therefore, is most easily differentiable. It would be important to comment upon the performance of the statistical features computed and why they were chosen for analysis. The idea is to capture the degree of contrast, correlation, energy and the degree of homogeneity of pixels in case of image slabs and wavelet coefficients in case of wavelet packet subbands. From the results it can be seen that the contrast feature performs well for ADWPT representation and not too well for raw image features. The main reason for this is that with wavelet representation the whole image is analyzed at different scales and frequencies represented by each subband. Whereas incase of raw image features the image is divided and features are computed for each slab separately which means only local features embodied by the slab are analyzed. The ADWPT representation performs better as each correlation feature derived from it represents the whole image information at various scales and frequencies whereas in case of raw textural analysis only correlation features at a certain location in the image are computed (due to this very reason the contrast feature is not performing well in case of raw image analysis). Hence, ADWPT provides us with a powerful tool to find inherent textural properties which cannot be achieved with raw image analysis.

It is important to note here that all these features are taken separately and results computed exclusively. If they are combined in some novel way such as envisaged in our previous work [4], the results may improve. However, in this paper our focus has been to demonstrate that the classification accuracy improves significantly when ADWPT based image representation is used instead of raw images.

According to Table 1, there are certain meningioma subtypes for which the classification accuracy is not as good as it is in the case of other meningioma subtypes. Particularly fibroblastic and meningiothelial subtypes have been found difficult to differentiate as they resemble each other to a certain extent. Furthermore, the statistical features being employed at this instance have only been able to capture properties that are better able to differentiate between the subtypes showing better accuracy. Our future work would focus on improving these results by using various more sophisticated features, experimenting with different classifiers and deriving better more optimal representations using the ADWPT.

References

1. P. Kleihues & W. K. Cavenee. *World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Nervous System*. IARC Press, 2000.
2. P. Burger. "What is an oligodendroglioma?" *Brain Pathol (2002)* **12**, pp. 257–259, 2002.
3. B. Lessmann, V. Hans, A. Degenhard et al. "Feature space exploration of pathology images using content-based database visualization." In *Proceedings SPIE Medical Imaging*. 2006.
4. H. Qureshi, N. Rajpoot, K. Masood et al. "Classification of meningiomas using discriminant wavelet packets and learning vector quantization." In *Proceedings of Medical Image Understanding and Analysis*. 2006.
5. M. Unser & M. Eden. "Multiresolution feature extraction and selection for texture segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11(7)**, pp. 717–728, 1989.
6. A. M. P. Michael M. Leung. "Scale and rotation invariant texture classification." In *Record of The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers*. 1992.
7. R. Porter & N. Canagarajah. "Gabor filters for rotation invariant texture classification." In *Proceedings of 1997 IEEE International Circuits and Systems*. 1997.
8. N. Saito & R. R. Coifman. "On local feature extraction for signal classification." In *Applied Analysis (O. Mahrenholtz and R. Mennicken, eds.), special issue of Zeitschrift fur Angewandte Mathematik und Mechanik, Akademie-Verlag, Berlin*. 1996.
9. N. Rajpoot. "Local discriminant wavelet packet basis for texture classification." In *Proceedings SPIE Wavelets X, San Diego, California*. 2003.
10. T. W. Nattkemper, B. Arnrich, O. Lichte et al. "Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods." *Artificial Intelligence in Medicine* **34(2)**, pp. 129–139, 2005.
11. R. Haralick. "Statistical and structural approaches to texture." In *Proceedings of the IEEE*. May.
12. N. Rajpoot, R. Wilson, F. Meyer et al. "A new basis selection paradigm for wavelet packet image coding." In *Proceedings IEEE International Conference on Image Processing (ICIP)*. 2001.
13. R. R. Coifman & M. V. Wickerhauser. "Entropy-based algorithms for best basis selection." *IEEE Transactions on Information Theory* **38**, pp. 713–718, 1992.
14. V. Vapnik. *Statistical Learning Theory*. Springer N.Y., 1998.
15. C.-C. Chang & C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at url-http://www.csie.ntu.edu.tw/ cjlin/libsvm.