A Novel Feature Extraction Method – Wavelet-Fourier Analysis and Its Application to Glaucoma Classification

Yufeng Zheng[a,*], *Edward A. Essock*[a,b]

[a]Dept. of Psychological & Brain Sciences, Univ. of Louisville, Louisville, KY,
[b]Dept. of Ophthalmology & Visual Sciences, Univ. of Louisville.
{yufeng.zheng, essock}@louisville.edu

Abstract

A novel wavelet-Fourier analysis (WFA) is presented to extract features from retinal nerve fiber layer (RNFL) data and classify eyes as "glaucomatous" or "healthy". Thickness estimates from 143 eyes from 72 people had been obtained with a scanning laser polarimeter (GDx-VCC). For each scan, a double-hump curve was defined by the RNFL thickness obtained within 64 radial sectors at a fixed distance from the optic disc. A discrete wavelet transform (DWT) was applied to these one-dimensional data, and then a Fourier transform (FFT) was applied to the DWT detail coefficients. The union of DWT approximation coefficients and FFT amplitudes was used to provide the final feature vectors. Principle component analysis (PCA) was employed for dimensionality reduction. Finally, Fisher's linear discriminant function (LDF) was used as a classifier. Experimental results with 84 normal eyes and 59 glaucomatous eyes show that the WFA method has the best classification performance in terms of sensitivity/specificity and ROC area, which are 0.775/0.965 and 0.941, compared to a FFT-based analysis (an earlier method developed in our lab, 0.755/0.958 and 0.918) and the NFI (a standard metric provided by the manufacturer, 0.671/1.0 and 0.918).

1. Introduction

Glaucoma often leads to characteristic damage of retinal ganglion cells. Therefore, strategies for glaucoma detection focus on detecting the functional or structural changes associated with ganglion cell disruption. The long processes of the ganglion cells are accessible by imaging through the eye's pupil and several methods of assessing the anatomical integrity of ganglion cells have been developed based on measuring the thickness of the retinal nerve fiber layer (RNFL) of the eye at each of a large array of points across the back of the eye. These computer-assisted imaging technologies [1,2,9] offer promise for detecting and assessing glaucomatous disruption. One such device, the scanning laser polarimeter (GDx-VCC, Laser Diagnostic Technologies, Inc., San Diego, CA) infers RNFL thickness based on a change in the polarization of light exiting the eye, with the amount of polarization retardation proportional to pointwise RNFL thickness. A new version of this device (GDx-VCC) measures and compensates for individual differences in an eye's extraneous effects on polarization. The goal of the present research is to examine data from this new device and to develop a method for classifying scans.

Feature extraction is the first and most important step aside from classification itself in a classification task. Original measured signals or data are defined as *attributes* or *patterns* that are seldom directly used as features to feed to a classifier because of their large dimensionality, high correlations and poor performance. *Feature extraction* abstracts high level information about individual patterns to facilitate classification. In general, a good *feature vector* (a set of feature variants), converted from a pattern vector of attributes, contains all of the essential information of the pattern with a possible lower dimensionality. All features should be distinguishable, reliable, and independent [8]. Consequently, some transformation techniques are anticipated to maximize (or enlarge) the distance between classes and probably reduce dimensionality in the feature space. Two common transformations, *fast Fourier transform* (FFT) and *discrete wavelet transform* (DWT), are used most often [3]. For example, in [4] a DWT-based classification technique was employed to predict the breakage of small drill bits, and Fourier analysis was applied to glaucoma detection in [5]. As far as which transform is better for a particular problem, it depends on the characteristics of the transform and the classified signals. In fact, FFT is good for frequency analysis, but DWT is suitable in analyzing both frequency information and temporal (or spatial) information [6]. Furthermore, DWT will reduce the dimensionality of the feature vector because of its multiresolution properties.

Ideally, the set of features used in a classification decision should be statistically independent, in that none of the features can be determined by a function of other features in the set because of correlations [7]. *Principle component analysis* (PCA) is an effective way to eliminate a substantial amount of redundancy caused by such interdependencies, and to achieve dimensionality reduction. Furthermore, PCA is easy to implement and parameterize.

Linear discriminant function (LDF) and *artificial neural networks* (ANN) approaches are most frequently employed for biomedical *classification* problems. Some reports showed ANN had a better classification performance (using cSLO data) than LDF did [2]. However, ANN requires a large sample size and a good training process. In contrast, the LDF is a linear combination of feature vectors such as Fisher's LDF, which maximizes the ratio of its *between-class variance* and *within-class variance*. It is efficient to obtain a LDF based on a small sample size. In addition, the linear function given by *linear discriminant analysis* (LDA, the process of inferring the LDF) can be physically explained in feature space.

Although still fairly new, many researchers are pursuing RNFL imaging for glaucoma detection. Recently, a Fourier analysis (FFTA) method was developed in [1,5], which resulted in a better classification performance than the *The-Number* (a standard metric developed by the manufacturer for classifying prior similar polarimetry data) in terms of sensitivity/specificity and ROC (receiver's operating curve) area [1]. But some spatial information was lost during the Fourier analysis in frequency domain.

To further improve the classification performance, a wavelet-Fourier analysis (WFA) method is proposed that is based on analyzing the characteristics of measured RNFL data and a DWT transformation.

2. Discrete Wavelet Transform

Wavelet-based analysis of signals is an interesting, and relatively recent, tool. Similar to Fourier series analysis, where sinusoids are chosen as the basis function, wavelet analysis is also based on a decomposition of a signal using an *orthonormal* (typically, although not necessarily) family of basis functions. Unlike a sine wave, a wavelet has its energy concentrated in time, or, as in the present application, space. Sinusoids are useful in analyzing periodic and time-invariant phenomena, while wavelets are well suited for the analysis of transient, time-varying signals. Furthermore, in spatial domain, DWT analysis also gives the best performance in detecting discontinuities or abrupt changes in signals.

Suppose $f(x) \in L^2(\mathbf{R})$ (where $\mathbf{R}$ is the set of real numbers, $L^2(\mathbf{R})$ denotes the set of measurable, square-integrable one-dimensional functions) relative to *wavelet function* $\psi(x)$ and *scaling function* $\varphi(x)$. A wavelet series expansion is similar in form to the well-known Fourier series expansion and maps a function of a continuous variable into a sequence of coefficients. If the function being expanded is a sequence of numbers, like samples of a continuous function $f(x)$, the resulting coefficients are called the *discrete wavelet transform* (DWT) of $f(x)$. The DWT transform pair is defined as

$$W_\varphi(j_0,k) = \frac{1}{\sqrt{M}}\sum_x f(x)\varphi_{j_0,k}(x), \quad (1)$$

$$W_\psi(j,k) = \frac{1}{\sqrt{M}}\sum_x f(x)\psi_{j,k}(x), \text{ for } j \geq j_0; \quad (2)$$

and

$$f(x) = \frac{1}{\sqrt{M}}\sum_k W_\varphi(j_0,k)\varphi_{j_0,k}(x) + \frac{1}{\sqrt{M}}\sum_{j=j_0}^{J-1}\sum_k W_\psi(j,k)\psi_{j,k}(x) \quad (3)$$

where $f(x)$, $\varphi_{j_0,k}(x)$ and $\psi_{j,k}(x)$ are functions of the discrete variable $x = 0, 1, 2, ..., M - 1$. Normally, we let $j_0 = 0$, and select M (the length of the discrete samples of $f(x)$) to be a power of 2 (i.e., $M = 2^J$) so that the summations are performed over $x = 0, 1, 2, ..., M - 1$, $j = 0, 1, 2, ..., J - 1$, and $k = 0, 1, 2, ..., 2^j - 1$. The transform itself is composed of M coefficients, the minimum scale is 0, and the maximum scale is $J - 1$. The coefficients defined in Eqs. (1) and (2) are usually called *approximation* and *detail coefficients*, respectively. The process of computing these coefficients is referred as DWT *analysis*. On the other hand, DWT *synthesis* (or inverse DWT) is defined by Eq. (3) to reconstruct $f(x)$ with these coefficients. Finally, it should be remembered that Eqs. (1) through (3) are valid for orthonormal bases and tight frames alone [6].

In practice we select a wavelet from ready-made wavelets for a particular problem. Different wavelets $\psi(x)$ have different effects, for instance, Harr wavelets are suitable for representing a piecewise signal, and Daubechies wavelets are better in compressing data. Here we chose a wavelet called "*Symmlets*".

3. Wavelet-Fourier Analysis

The procedure classifies eyes as "normal" or "glaucomatous" based on RNFL thickness measurements which provided the mean thickness of 64 sectors located on a ring with a diameter equivalent to 1.75 disc diameters and centered on the optic disc. There are three major steps involved as follows.

3.1 Feature extraction

(A) *Fourier analysis method*: A Fourier analysis method [1,5] was recently developed in our lab that can be used for a glaucoma classification. Fourier analysis is a process that decomposes a signal into the summation of a series of sine and cosine waves with different amplitudes, frequencies and phases (harmonics). Each Fourier transform coefficient has a physical meaning that can be interpreted in frequency domain corresponding to the original signal in the spatial domain. With these FFT coefficients directly, or after some kind of process such as filtering, the spatial signal can be reconstructed with an inverse FFT. Therefore, FFT is a standard method in feature representation [3]. In consideration of the shape of our analyzing signal that possesses a "double-hump" pattern (Fig. 1 (a)), similar to a sine wave, (which may lead to a shorter FFT representation with fewer harmonics), it is reasonable to use the signal's FFT coefficients as primary features.

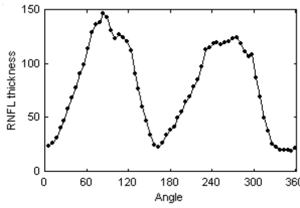
However, it is difficult to reflect abrupt changes, such as tend to occur in a GDx-VCC signal, using FFT coefficients in the frequency domain alone. In other words, some small changes of waveform structure may be lost in the primary feature representation with the FFTA method. Accordingly, a new method is desired that can consider this spatial information contained in GDx-VCC data with a combination of the FFTA.

(B) *Wavelet-Fourier analysis method*: In our current research, a wavelet-Fourier analysis (WFA) method is proposed that was shown to improve the glaucoma classification performance significantly. Here we select DWT as our *first-step* transformation, which is also a common method used in feature extraction [3]. In Section 2, it was shown that DWT is highly suitable for analyzing discontinuities and abrupt changes contained in signals. These small changes are encoded into wavelet transformed coefficients: approximation coefficients, containing down-sampled spatial information; and detail coefficients, containing detailed difference information due to the approximation (Fig. 1 (b)). In addition, DWT is a *multiscale analysis* method that means analysis can be based on various space and frequency resolution scales [6]. At each transformation scale, there is a different resolution ability (*multi-resolution*) of space and frequency, corresponding to the approximation and detail coefficients respectively. A second scale (j=4) of DWT analysis was selected (according to the dimension of original signals and the experimental results) in order to preserve high enough spatial resolution ability. On the other hand, this results in a relatively low frequency resolution overall. For this reason, a further step, drawing from our previous work, applies a Fourier transform to the DWT detail part, considered as the *second-step* transformation, and computes frequency amplitudes so as to achieve high frequency resolution. The final step is to normalize the DWT approximation coefficients and FFT amplitudes separately, and to join them together as the final feature vector.

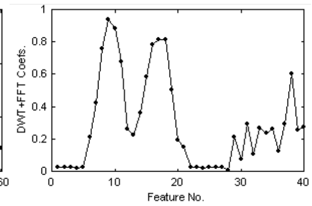
3.2 Feature optimization

Principal component analysis (PCA) is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called *principal components*. The objective of PCA is to reduce the dimensionality of the dataset while retaining most of the original variability in the data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

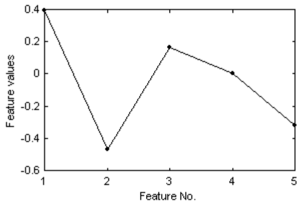
A transformation matrix **A**, used for principal components calculation, can be computed by solving



(a) Original signal



(b) Transformed & normalized



(c) Dimension reduced

Fig. 1: Feature extraction steps (a)-(c) (a normal sample)

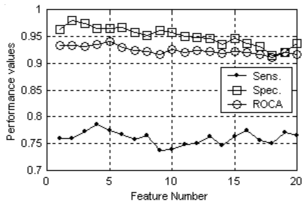


Fig. 2: Classification performance on the test set varying with the feature number k

the eigenvectors and eigenvalues of the covariance of feature vectors [6]. Then the k ($\leq n$, the dimension of a feature vector) eigenvectors corresponding to the k largest eigenvalues to form matrix $\mathbf{A}_k$ are chosen. Using the transformation matrix $\mathbf{A}_k$, a compressed feature vector of dimension k can be derived though linear combinations of original ones (Fig. 1 (c)). The reduced dimensionality of the features can make the classifier more efficient and more stable.

3.3 Classification

The main purpose of a *linear discriminant analysis* is to predict group membership based on a linear combination of a set of *predictor variables* (i.e. a feature vector). The procedure begins with a set of observations where both group membership and the values of the predictor variables are known. The end result of the procedure is a model (i.e. LDF) that allows prediction of group membership when only the predictor variables are known.

We've chosen Fisher's LDF as a classifier to improve the classification robustness [10]. We first randomly and uniformly separate the entire dataset into two subsets. Then one subset is used for training and for obtaining a LDF, and the other subset is used for testing to assess performance of the classification method.

4. Experimental Results and Discussion

Our present sample consists of scans of 59 glaucomatous eyes and 84 normal eyes, obtained from 30 glaucoma patients (there is an invalid scan of one eye) and 42 healthy people. In prior research [9], we documented a correlation of the pattern of RNFL thickness between fellow eyes. With this in mind, we analyzed the results for data from just one eye (selected at random) as well as from the full two-eye data. To minimize the possibility of sampling bias, classification results from WFA and FFTA were obtained for 20 runs and means reported (i.e. a random train- and test-set division, known as *k-fold variation* cross validation [11], performed 20 times).

For the WFA, an 8th-order wavelet named "*Symmlets*" was chosen. Then we applied a two-scale DWT to the 64-region GDx-VCC thickness data (i.e. $M = 64$, $J = 6$ and $j = 4, 5$ in Eq. (2)) by using coefficients of $j = 4$ while discarding that of $j = 5$ (as they are similar to that of j=4, and would be removed by PCA because of their correlations even if they were included), and retained the approximation coefficients but applied an FFT to the DWT detail coefficients. The DWT coefficients and FFT coefficients were normalized to the range [0,1], and joined together as preliminary feature vectors (Fig. 1 (b)). It is difficult to decide how to specify the parameter k, (the reduced feature number) used in $\mathbf{A}_k$. Hence, while varying k from 1 to 20 with a step increment 1, we observed the classification performance (averaged over 20 runs) on the test set as shown in Fig. 2. Clearly, the best value of k was 5 for performing the succeeding PCA. Finally we fed the optimized feature vectors (5 elements in each vector) by PCA to the Fisher's LDF for training and testing.

For the FFTA classification, the amplitudes of the FFT transform of GDx-VCC data were computed and treated as features. The manufacturer's suggested discrimination metric presently is the *Nerve Fiber Indicator* (NFI, a normalized score yielded by a trained *support vector classifier*) when an image was acquired. The *NFI* ranges from 0 to 100: a small number suggests normal; and a large number tends to be glaucomatous. To make a judgment, a threshold of 30 was suggested by the company.

Performance of the three methods for discriminating normal from glaucoma cases is shown in Table 1. On the most comprehensive measure, area under the ROC curve, the WFA performance was best (0.941) compared to 0.918 for both *NFI* and FFTA methods for the two-eye data. Similarly, performance was best for the WFA method (0.927) for the one-eye data. Performance was also better for this method in terms of sensitivity. A second point is that the better performance with the two-eye data set is consistent with the suggestion that there is a relevant relationship between the two eyes' data [9]. In contrast, the results obtained by using only the DWT coefficients (j=4) as features are Sens./Spec./ROC area = 0.771/0.957/0.925 (averaged over 20 runs on the 2-eye data).

5. Conclusion

A new feature extraction method, wavelet-Fourier analysis (WFA), was presented and applied to a glaucoma detection problem. Both DWT and FFT were employed to present the spatial and frequency information contained in the source GDx-VCC data jointly. This is followed by dimension reduction by the PCA technique and use of Fisher's LDF as a classifier. The k-fold variation was used for cross validation ensuring reliable results. We conclude that the WFA method offers an improvement over the prior FFT-based method of classifying glaucomatous RNFL scans that is due to the present method's incorporation of the benefits from DWT instead of relying only on FFT. The WFA method can also be applied to other signal analysis and classification problems in addition to GDx-VCC eye-scan data.

Table 1: Cross validation results averaging of 20 runs on test subset using GDx VCC data (N = normal, G = Glaucoma)

Dataset	Method	Sensitivity	Specificity	ROC area
1-eye data (42N×30G)	WFA	0.783	0.962	0.927
	FFTA	0.767	0.938	0.901
	NFI	0.671	1.0	0.918
2-eye data (84N×59G)	WFA	0.775	0.965	0.941
	FFTA	0.755	0.958	0.918
	NFI	0.671	1.0	0.918

Acknowledgments

We thank Dr. Michael J Sinai of Laser Diagnostic Technologies for providing the eye scan data. This research was funded by a grant from the Glaucoma Research Foundation.

References

[1] E. A. Essock, M. J. Sinai, et al., Fourier analysis of OCT and GDx RNFL measurements in the diagnosis of glaucoma, Archives of Ophthalmology (accepted for publ.).

[2] Christopher Bowd, Kworkleung Chan, et. al., Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthamoscopy of the optic disc, Investigative Ophthalmology & Visual Science 43 (11) (2002) 3444-3454.

[3] Abhijit S. Pandya, Robert B. Macy, Pattern Recognition with Neural Networks in C++, CRC Press LLC, Florida, 1996.

[4] K. Mori, N. Kasashima, et al., Prediction of small drill bit breakage by wavelet transforms and linear discriminant function, International Journal of Machine Tools and Manufacture 39 (1999) 1471-1484.

[5] Edward A. Essock, Michael J. Sinai, et al., Fourier analysis of nerve fiber layer measurements from scanning laser polarimeter in glaucoma: emphasizing shape characteristics of the 'double-hump' pattern, Jounal of Glaucoma 9 (6) (2000) 444-452.

[6] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing (Second Edition), Prentice Hall, Upper Saddle River, NJ, 2002.

[7] Carl G. Looney, Pattern Recognition Using Neural Networks — Theory and Algorithms for Engineers and Scientists, Oxford University Press, New York, 1997.

[8] Kenneth R. Castleman, Digital Image Processing, Prentice Hall, NJ, 1996.

[9] Sinai, M. J., Essock, E. A., Fechtner, R. D., and Srinivasan, N., Diffuse and Localized Nerve Fiber Layer Loss Measured with a Scanning Laser Polarimeter: Sensitivity and Specificity of Detecting Glaucoma, Journal of Glaucoma 9 (2000) 154-162.

[10] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[11] Sarunas Raudys, Statistical and neural classifiers — An integrated approach to design, Springer, London, 2001.