May 11-13, 2009

An Adaptive Classifier Fusion Method for Analysis of Knee-Joint Vibroarthrographic Signals

Yunfeng Wu* and Sridhar Krishnan
Department of Electrical and Computer Engineering, Ryerson University
350 Victoria Street, Toronto, ON, M5B 2K3, Canada
Email: *y.wu@ieee.org

***Abstract*—Externally recorded knee-joint vibroarthrographic (VAG) signals bear diagnostic information related to degenerative conditions of cartilage disorders in a knee. In this paper, the number of atoms derived from wavelet matching pursuit (MP) decomposition and the parameter of turns count with the fixed threshold that characterizes the waveform variability of VAG signals were extracted for computer-aided analysis. A novel multiple classifier system (MCS) based on the adaptive weighted fusion (AWF) method is proposed for the classification of VAG signals. The experimental results shows that the proposed AWF-based MCS is able to provide the classification accuracy of 80.9%, and the area of 0.8674 under the receiver operating characteristic curve over the data set of 89 VAG signals. Such results are superior to those obtained with best component classifier in the form of least-squares support vector machine, and the popular Bagging ensemble method.**

I. Introduction

Knee-joint vibroarthrographic (VAG) signals, also referred to as knee-joint vibration sounds, can be recorded using an accelerometer at the mid-patella position of a knee during normal movement of the leg. Normal joint surfaces are smooth and produce little or no sound, whereas joints affected by osteoarthritis and other degenerative diseases may have suffered cartilage loss and produce grinding sounds [1]. Previous studies [2–5] have shown that the applications of advanced signal processing and pattern recognition techniques in VAG signal analysis can could reduce the diagnostic use of arthroscopy, and also help develop effective clinical tools to detect pathological conditions of the knee joint with degenerative arthritis.

Due to the fact that the quality of the knee-joint surfaces coming in contract may not be the same from one angular position to another during articulation of the joint, the VAG signal presents nonstationary characteristics in both of the time and frequency domains. This paper studied the time-frequency feature derived with the matching pursuit decomposition, and analyzed the signal variability in the time domain. Then we applied the multiple classifier system with a novel classifier fusion method for the analysis of VAG signals.

II. Data and Feature Descriptions

A. Data Acquisition

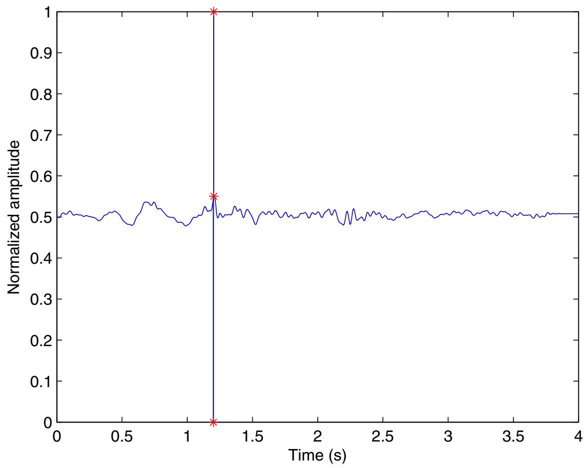
In the present study, we used the same database as that investigated in a few recent studies [3–5]. The VAG signals were collected by the research group of Rangayyan, University of Calgary, Calgary, AB, Canada. During the data acquisition procedure, each subject sat on a rigid table in a relaxed position with the leg being tested freely suspended in air. Each VAG signal was recorded by placing an accelerometer (model 3115a, Dytran, Chatsworth, CA) at the mid-patella position of the knee as the subject swung the leg over an approximate angle range of 135º to 0º and back to 135º in 4 s. The VAG signal was conditioned using a bandpass filter with a bandwidth of 10 Hz to 1 kHz, and amplified before digitizing at a sampling rate of 2 kHz. There are a total of 89 VAG signals included in the database, in which 51 from normal health volunteers and 38 from subjects with knee-joint pathology. The database available permits screening only, that is, to distinguish between the normal and abnormal signals.

B. Features

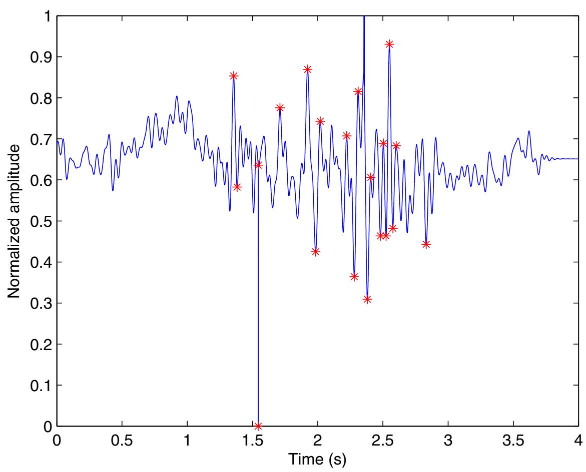
The matching pursuit (MP) method proposed by Mallat and Zhang [6] is suited for analysis of inherent nonstationary VAG signals. To represent time-frequency characteristics of a signal, the MP algorithm decomposes the signal using basis functions with good time-frequency properties which are referred to as atoms. The basis functions are used to build up dictionaries for the orthogonal projections in signal decompositions based on the MP algrithm. In the present study, we considered the dictionary of wavelet packet bases calculated with a Daubechies 8 (db8) filter for the decomposition of each VAG signal, because the Daubechies wavelets are a family of orthogonal wavelets that have a support of minimum size for any give number of vanishing moments, and such wavelets can be used for signal decomposition with excellent time and scale properties [6]. To determine the iterations of the wavelet MP decomposition in our experiments, we utilized the indicator in terms of signal-to-noise ratio (SNR), rather than the commonly used decay parameter that characterizes the convergence of residual energy. With a given SNR, the MP decomposition of an abnormal VAG signal normally requires relatively smaller number of iterations than that of a normal signal, because the abnormal signal is more noisy and contaminated by a larger amount of artifacts such as muscle contraction interference. Thus, the number of MP atoms can be considered as a potential feature for classification applications. After testing the SNR at different levels, we found that the SNR of 15 dB is an excellent indicator to determine the wavelet MP iterations. The p value of the number of atoms (Natom) obtained with the Student's t-test is 0.0002, which implies that the Natom presents an

978-1-4244-3820-4/09/$25.00 ©2009 IEEE

significance of separation between the normal and abnormal signals.



(a)



(b)

Fig. 1. Illustration of significant turns in the filtered (a) normal and (b) abnormal VAG signals, in which the significant turns detected have been marked with asterisks. The delay of 0.05 s caused by the lowpass Butterworth filter has been calibrated.

Besides the number of the wavelet MP atoms, the signal variability analysis in the time domain may be useful for VAG signal classification as well. The turns count method [7], the procedure of which involves the detection of the number of spikes, swings, or changes in amplitude larger than a certain threshold, is worthy of consideration. In the present study, we first normalized each VAG signal to the amplitude range from zero to unity, the same as in the recent studies [3–5]. In each signal, the amplitude of all samples was amplified with the same scale so that the variability information of the signal can be preserved. Before applying the turns count method in the signal, we implemented a filtering procedure using a 10th-order lowpass Butterworth filter (-3 dB cutoff at 50 Hz) with unit gain at direct current (DC). This lowpass Buttworth filter causes a delay of 0.05 s (100 samples), which was calibrated after the filtering procedure in our experiments. The reason why it is better to use the lowpass Butterworth filter instead of the signal reconstructed with the MP atoms is that the MP method is unable to eliminate the artifacts present in the VAG signal such as the interference caused by muscle contractions or 50 or 60 Hz power-supply lines. Then, we fixed the amplitude threshold at 0.2 to compute the turns over the normalized and filtered VAG signals.

Fig. 1 shows the results of the turns count with the fixed threshold (TCFT) method for the signals after the procedures of normalization and filtering. It can be observed that more significant turns (asterisks) have been detected in the abnormal signal, shown in Fig. 1 (b), than those of the normal signal in Fig. 1 (a). The p value of the TCFT obtained with the Student's t-test is 0.0013, which indicates a statistical significance of the differences between the normal and abnormal signals.

Fig. 2 draws in the two-dimensional space the features (Natom and TCFT) associated with the normal and abnormal VAG signals using scatter markers of circles and crosses, respectively. It is noted that the Natoms associated with many abnormal signals are smaller than 800, and the locations of the normal signals congregate in the ranges from 600 to 2400 Natoms and from 0 to 30 TCFTs.

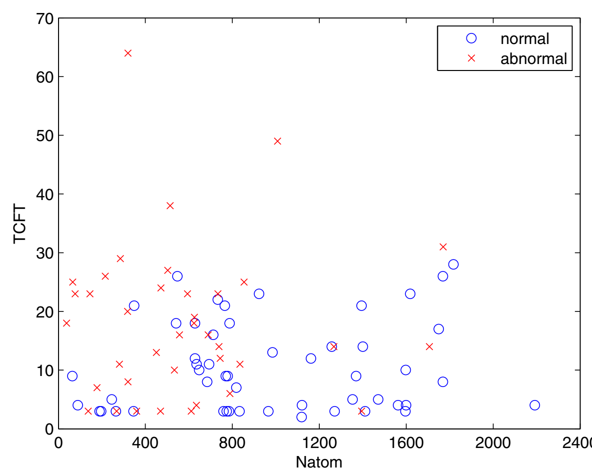


Fig. 2. Scatter plot of the features derived from the wavelet matching pursuit decomposition and the time-domain signal variability analysis. Natom: number of the wavelet matching pursuit atoms; TCFT: turns count with the the fixed threshold.

C. Classifier Fusion Method

Classifier fusion methods are emerging machine learning techniques that follow the principle of "divide and conquer" [8]. The classifier fusion methods are very useful in the design of a multiple classifier system (MCS) with the aim to ameliorate classification performance. By combing a finite number of component classifiers (CLs) with a combination rule [9] or fusion strategy [10, 11], the MCS is expected to provide an informative overall decision that is supposedly superior to that attained by any one of the CLs acting solely [12]. In the present study, we propose the adaptive weighted fusion (AWF) method that can be effectively applied in the MCS for VAG signal analysis.

Suppose that the MCS combines a total of K CLs with the linear fusion strategy. The local decision generated by the kth CL is denoted as $f_k(\mathbf{x}^i)$, with regard to the feature vector of the ith VAG signal, $\mathbf{x}^i$. The ensemble system then provides the overall classification decision, $f_{AWF}(\mathbf{x}^i)$, by linearly combining the CLs with the weights $w_k(\mathbf{x}^i)$ that are varied from one signal to another. Thus the output of the AWF-based MCS can be formulated as

$$f_{AWF}(\mathbf{x}^i) = \sum_{k=1}^{K} w_k(\mathbf{x}^i) f_k(\mathbf{x}^i). \quad (1)$$

And the nonnegative and normalization constraints on the weights, as widely accepted in the literature [13–15], can be written as

$$\sum_{k=1}^{K} w_k(\mathbf{x}^i) = 1,\ w_k(\mathbf{x}^i) \geq 0. \quad (2)$$

The task of the AWF is to determine the fusion weights that help the ensemble system provide an overall classification decision with higher accuracy. To achieve this goal, let us study the error term of the CLs and the AWF-based ensemble. Concerning the kth CL, the squared error that characterizes the difference between the local decision and the desired class label, $d(\mathbf{x}^i)$, in relation to the ith VAG signal is

$$e_k(\mathbf{x}^i) = \left[d(\mathbf{x}^i) - f_k(\mathbf{x}^i)\right]^2. \quad (3)$$

The cost function of the ensemble, $C_{AWF}(\mathbf{x}^i)$ is defined as follows:

$$C_{AWF}(\mathbf{x}^i) = \sum_{k=1}^{K} w_k^2(\mathbf{x}^i) e_k(\mathbf{x}^i). \quad (4)$$

Considering Equations (2) and (4), the minimization of the squared error of the ensemble is equivalent to the constrained quadratic programming (CQP) problem specified below:

$$\begin{cases} \text{minimize} & C_{AWF}(\mathbf{x}^i) = \sum_{k=1}^{K} w_k^2(\mathbf{x}^i) e_k(\mathbf{x}^i), \\ \text{subject to} & \sum_{k=1}^{K} w_k(\mathbf{x}^i) = 1,\ w_k(\mathbf{x}^i) \geq 0, \end{cases} \quad (5)$$

Note that both of $\mathbf{w}(\mathbf{x}^i)$ and $\mathbf{e}(\mathbf{x}^i)$ are nonnegative, the cost function of the AWF-based ensemble $C_{AWF}(\mathbf{x}^i)$ is therefore convex, so that a solution to the global optimization exists. In order to solve the CQP problem presented in Equation (5), we applied the Lagrange multiplier method [16] and define the corresponding loss function as

$$L(w_k(\mathbf{x}^i), \lambda(\mathbf{x}^i)) = \sum_{k=1}^{K} w_k^2(\mathbf{x}^i) e_k(\mathbf{x}^i) - \lambda(\mathbf{x}^i)\left[\sum_{k=1}^{K} w_k(\mathbf{x}^i) - 1\right], \quad (6)$$

where the nonnegative coefficient $\lambda(\mathbf{x}^i)$ is the Lagrange multiplier, which is also varied from one signal to another.

According to the weak Lagrangian principle [16], the optimum solution to the CQP problem, $\{\mathbf{w}^*(\mathbf{x}^i), \lambda^*(\mathbf{x}^i)\}$, is the stationary point of the loss function presented in Equation (5), and satisfies the following unique equations:

$$\begin{cases} \frac{1}{\partial w_k(\mathbf{x}^i)} \partial L(w_k(\mathbf{x}^i), \lambda(\mathbf{x}^i)) = 2 e_k(\mathbf{x}^i) w_k(\mathbf{x}^i) - \lambda(\mathbf{x}^i) = 0, \\ \frac{1}{\partial \lambda(\mathbf{x}^i)} \partial L(w_k(\mathbf{x}^i), \lambda(\mathbf{x}^i)) = \sum_{k=1}^{K} w_k(\mathbf{x}^i) - 1 = 0. \end{cases} \quad (7)$$

Thus the optimum solution $\lambda^*(\mathbf{x}^i)$ can be obtained by solving Equation (7), i.e.,

$$\lambda^*(\mathbf{x}^i) = \frac{2}{\sum_{k=1}^{K} \left[e_k(\mathbf{x}^i)\right]^{-1}}. \quad (8)$$

The optimal weights $w_j^*(\mathbf{f}^i)$ of the AWF that minimize the squared error of the ensemble system can be derived as

$$w_j^*(\mathbf{x}^i) = \frac{\left[e_j(\mathbf{x}^i)\right]^{-1}}{\sum_{k=1}^{K} \left[e_k(\mathbf{x}^i)\right]^{-1}},\ j = 1, \ldots, K, \quad (9)$$

Because the error term of the CL can be estimated when the ith VAG class label $d(\mathbf{x}^i)$ is given and the CL parameters are specified, the optimal fusion weights can be directly computed according to Equation (9).

III. Experiments and Results

In our experiments, five least-squares support vector machines (LS-SVMs) [17] were used to work as the component classifiers in the MCS. The LS-SVM is a reformulation to the standard support vector machine (SVM), and the moderate complexity of the LS-SVM makes the learning more efficient than that of a standard SVM. We chose the Gaussian kernel function with the spread parameter equal to 2, and the number of Gaussian kernel functions, the network weights, and the bias will be automatically determined with the LS-SVM learning algorithm.

To compare the classification performance of the proposed AWF method, we also implemented the most popular Bagging ensemble that trained the 5 component LS-SVMs (CSVMs) using the training data resampled with the bootstrap procedure. In the Bagging ensemble, the outputs of the 5 CSVMs were combined by averaging, whereas in the proposed MCS, the overall decision was produced with the AWF method.

Table I lists the classification accuracy in percentage provided by the best CSVM, the Bagging ensemble, and the AWF-based MCS. In addition, the present study used the receiver operating characteristic (ROC) curve method which presents the test of true positive rate (TPR) and false positive rate (FPR) evaluated with several different diagnostic thresholds. The area (*Az*) under the ROC curve was derived to serve as a measure of the overall diagnostic performance in each classification experiment. From Table I, the AWF-based

MCS provides the accurate rate of 80.90%, better than that obtained with either the Bagging ensemble (68.54%) or the best CSVM (74.16%). Such an accurate rate of the AWF-based MCS is much better than the previous studies using the logistic regression analysis with the energy, energy spread, frequency, and frequency spread features derived from the Gabor MP method (accuracy: 68.9%) [2].

TABLE I
CLASSIFICATION RESULTS OBTAINED WITH THE BEST COMPONENT LS-SVM CLASSIFIER AND THE CLASSIFIER ENSEMBLE METHODS.

Classifier	Accuracy (%)	*Az*	Standard error (SE)
Best CSVM	74.16	0.7812	0.0498
Bagging	68.54	0.7828	0.0490
AWF	80.90	0.8674	0.0378

The ROC curves derived in the experiments are illustrated in Fig. 3, from which we can observe that the ROC curve provided by the AWF-based MCS is consistently over that obtained with the Bagging ensemble. The AWF-based MCS also presents the superiority of overall diagnostic performance, with the the best *Az*: 0.8674, and the lowest standard error (SE): 0.0378, as depicted in Table I. Moreover, the ROC curve result of the AWF-based ensemble is better or comparable to that of the RBFN classifier with the features of form factors, skewness, kurtosis, entropy (*Az*: 0.8172) [3] or with the features of variance of mean-squared values and turns count with adaptive threshold (*Az*: 0.9174) [5].

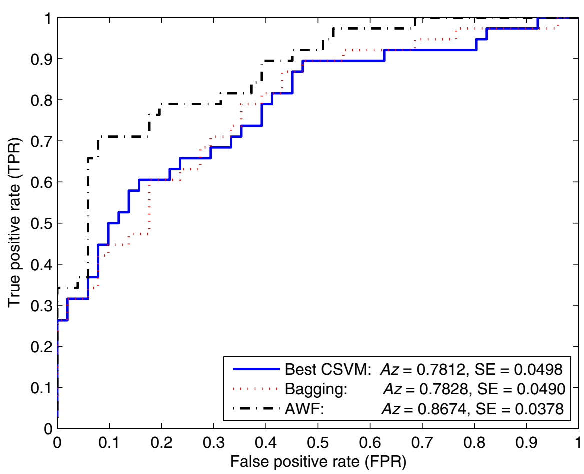


Fig. 3. Receiver operating characteristic (ROC) curves obtained with the best component LS-SVM classifier, the Bagging, and the adaptive weighted fusion (AWF) method. Az: Area under the ROC curve. SE: standard error.

IV. CONCLUSION

Computer-aided analysis of VAG signals can be used in the noninvasive detection of roughness, softening, breakdown, or the state of lubrication of the articular cartilage surfaces, which leads to better clinical tools for early detection, localization, and quantitative analysis of knee-joint disorders. In the present study, the Natom and TCFT features were derived from the time-frequency wavelet MP decomposition and time-domain signal variability analysis, respectively. The adaptive weighted fusion method was effectively utilized to distinguish between the normal and abnormal VAG signals. Compared with the prevailing Bagging ensemble, the multiple classifier system using the adaptive weighted fusion method can effectively provide a more accurate classification rate and the ROC curve with higher *Az* and lower SE, over the data set of 89 VAG signals.

REFERENCES

[1] C. B. Frank, R. M. Rangayyan, and G. D. Bell, "Analysis of knee sound signals for non-invasive diagnosis of cartilage pathology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, no. 1, pp. 65–68, 1990.

[2] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank, "Adaptive time-frequency analysis of knee joint vibroarthrographic signals for non-invasive screening of articular cartilage pathology," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 6, pp. 773–783, 2000.

[3] R. M. Rangayyan and Y. F. Wu, "Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions," *Medical & Biological Engineering & Computing*, vol. 46, no. 3, pp. 223–232, 2008.

[4] R. M. Rangayyan and Y. F. Wu, "Modeling and classification of knee-joint vibroarthrographic signals using probability density functions estimated with Parzen windows," in *Proceedings of the 30th Annual International Conference of IEEE EMBS (EMBC'08)*, Vancouver, BC, Canada, 2008, pp. 2099–2102.

[5] R. M. Rangayyan and Y. F. Wu, "Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions," *Annals of Biomedical Engineering*, vol. 37, no. 1, pp. 156–163, 2009.

[6] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[7] R. M. Rangayyan, *Biomedical Signal Analysis: A Case-Study Approach*. New York, NY: IEEE and Wiley, 2002.

[8] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall PTR, 1998.

[9] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matus, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[10] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 281–286, 2002.

[11] A. Sinha, H. Chen, D. G. Danu, T. Kirubarajan, and M. Farooq, "Estimation and decision fusion: A survey," *Neurocomputing*, vol. 71, no. 13–15, pp. 2650–2656, 2008.

[12] Y. F. Wu and J. I. Arribas, "Fusing output information in neural networks: Ensemble performs better," in *Proceedings of the 25th Annual International Conference of IEEE EMBS (EMBC'03)*, Cancun, Mexico, 2003, pp. 2265–2268.

[13] Y. F. Wu, J. J. He, Y. Man, and J. I. Arribas, "Neural network fusion strategies for identifying breast masses," in *Proceedings of the 2004 International Joint Conference on Neural Networks (IJCNN'04)*, Budapest, Hungary, 2004, pp. 2437–2442.

[14] Y. F. Wu and S. C. Ng, "Unbiased linear neural-based fusion with normalized weighted average algorithm for regression," in *Proceedings of the 4th International Symposium on Neural Networks (ISNN'07), LNCS 4493*, Nanjing, China, 2007, pp. 664–670.

[15] Y. F. Wu and S. C. Ng, "Breast tissue classification based on unbiased linear fusion of neural networks with normalized weighted average algorithm," in *Proceedings of the 2007 International Joint Conf. Neural Networks (IJCNN'07)*, Orlando, FL, 2007, pp. 2846–2850.

[16] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*. Columbus, OH: McGraw-Hill, 1995.

[17] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific Publishing, 2002.