REALTIME RECOGNITION OF ORCHESTRAL INSTRUMENTS

Ichiro Fujinaga and Karl MacMillan
{ich, karlmac}@peabody.jhu.edu
Peabody Conservatory of Music
Johns Hopkins University
Baltimore, MD USA 21202

ABSTRACT

This paper describes the culmination of a realtime timbre recognition project based on two previous experiments and Miller Puckette's `fiddle` program. For the recognition task, all experiments use an exemplar-based learning system based on a k-nearest neighbor (k-NN) classifier and a genetic algorithm to seek the optimal set of weights for the features to improve the recognition performance. Although a considerable amount of time is needed for the genetic algorithm to determine the set of weights, the calculation time of the actual k-NN classifier is insignificant and can be performed as soon as the required number of audio samples has been processed, thus making it feasible for realtime applications.

INTRODUCTION

Sophisticated analysis of audio in realtime has been made possible by the recent increase in computing power on personal computers. One excellent example is Miller Puckette's `fiddle` program, which is a very robust and efficient pitch-detector (Puckette et al. 1998). A realtime orchestral instrument recognition system was developed based on `fiddle`. The system is currently running on a 366Mhz Pentium II with GNU/Linux 2.2.x.

OVERALL ARCHITECTURE

The overall architecture of the system is shown in Figure 1. The `fiddle` program is used for the acquisition of live input or soundfile input. The sound is analyzed using a moving 2048-points window with hop size of 1024 points. To estimate the pitch, a spectral envelope, based on strong peaks, is generated for each window. This data including the pitch is used to calculate various features and stored as multi-dimensional feature vectors in a knowledge base used for the classifier.

In k-nearest neighbor (k-NN) classifiers (Cover and Hart 1967) the distances between the feature vector of an unknown sample and the vectors of all classified samples are calculated. The class, in this case a timbre, represented by the majority of the k-nearest neighbors is assigned to the unknown sample. The value of k is typically a small integer.

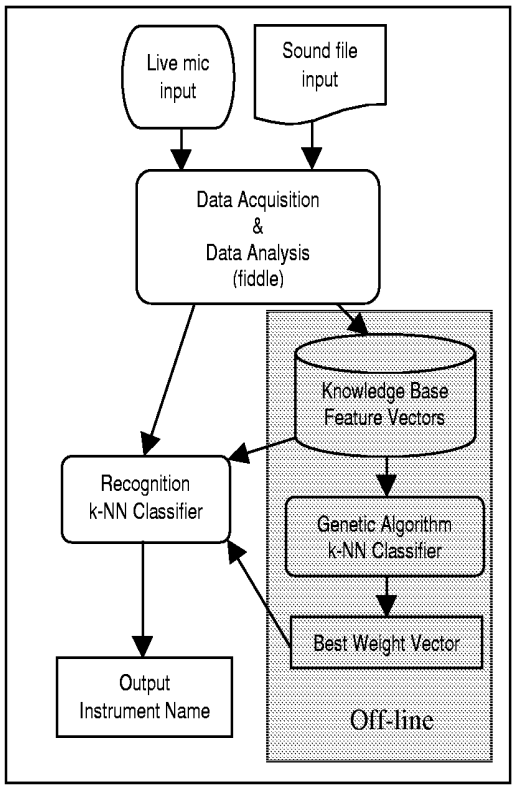

Figure 1. The overall architecture

This simple classifier can be turned into a learning system by continually adding samples into the knowledge base (Aha 1997). Furthermore, the performance of the classifier can be dramatically

increased by using weighted feature vectors. Finding a good set of weights, however, is extremely time-consuming, thus a genetic algorithm (Holland 1975) is used to find a solution (Wettschereck et al. 1997). Note that the genetic algorithm can be run off-line without affecting the speed of the realtime recognition process.

Until recently exemplar-based classifier systems, such as k-NN classifiers, have often been avoided in many applications because of their high memory requirements and computational costs. The current experiment shows that an exemplar-based system can indeed be used in a demanding realtime application.

THE EXPERIMENTS

The current experiment is the culmination of two previous experiments. In all cases, the training data comprised of over 1300 notes from 39 different timbre (23 orchestral instruments, some with different articulations) taken from the McGill Master Samples CD library.

In the first experiment (Fujinaga 1998), only the spectral shapes from manually selected steady-state portion of the sounds were used as data (Sandell 1994) to the recognition process. The features calculated from the spectral data included centroid and other higher order moments, such as skewness and kurtosis. The recognition rate for the 39-timbre group was 50% and for a 3-instrument group (clarinet, trumpet, and bowed violin) was 81%. In all cases, the standard leave-one-out cross-validation is used on the training set to calculate the recognition rate.

In the second experiment (Fraser and Fujinaga 1999), two improvements were made. First, spectral envelope data generated by the `fiddle` program was used as the basis of analysis. Second, the features of dynamically changing spectrum envelopes, such as the velocity of the centroid and its variance were added. The recognition rate increased to 64% for the 39-timbre group and 98% for the 3-timbre group.

In the current experiment, the system was made to work in realtime. The additional enhancements include: the reduction of analysis time from 500 ms to 250 ms, more precise location of attack points, and the addition of spectral irregularity and tristimulus as features describing the spectral envelopes.

In most analyses of timbre, the beginning of an attack is defined by an increase in the amplitude of the time-domain signal above a certain threshold. To allow for noise and different attack profiles of different instruments, a more consistent definition of attack was found to be the point at which `fiddle` was able to determine a pitch. Thus, when `fiddle` reports an attack (using the amplitude threshold method) the recognition system traces back in time to redefine the attack point if a pitch has already been detected.

Spectral irregularity aims to measure the jaggedness of the spectral envelope. For example, the clarinet has relatively low power at even partials. Two kinds of formulae were used: one originally suggested by Krimphoff et al (1994),

$$\text{irregularity} = \sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right|$$

and a modified version by Jensen (1999),

$$\text{irregularity} = \frac{\sum_{k=1}^{N} (a_k - a_{k+1})^2}{\sum_{k=1}^{N} a_k^2}$$

where a_k is the amplitude of the k th partial.

Tristimulus introduced by Pollard and Jansson (1982) also quantifies the spectral envelope. Two values were used to describe the low-order and high-order partials:

$$(a_2 + a_3 + a_4) / \sum_{k=1}^{N} a_k$$

and

$$\sum_{k=5}^{N} a_k / \sum_{k=1}^{N} a_k$$

These modifications have greatly enhanced the recognition rates while reducing the number of samples analyzed for the recognition. Most three to ten instrument groups were recognized in the 95–100% range, while the recognition for the 39-timbre group increased by 4% to 68%. Although direct comparison is difficult, these results indicate a comparable if not a superior performance to experienced musicians. For a detailed comparison of human and computer timbre recognition, see Martin (1999).

CONCLUSIONS

A realtime system for recognizing orchestral instruments was successfully implemented. In addition, the experiments demonstrate the feasibility of deploying exemplar-based classifiers in time-critical environments.

REFERENCES

Aha, D. W. 1997. Lazy learning. *Artificial Intelligence Review* 11 (1–5): 7–10.

Cover, T., and P. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13 (1): 21–7.

Fraser, A., and I. Fujinaga. 1999. Toward real-time recognition of acoustic musical instruments. *Proceedings of the International Computer Music Conference*, 207–10. San Francisco: ICMA.

Fujinaga, I. 1998. Machine recognition of timbre using steady-state tone of acoustic musical instruments. *Proceedings of the International Computer Music Conference*, 207–10. San Francisco: ICMA.

Holland, J. H. 1975. *Adaptation in natural and artificial systems*. Ann Arbor: U. of Michigan Press.

Jensen, K. 1999. *Timbre models of musical sounds.* Ph.D. dissertation. University of Copenhagen.

Krimphoff, J., S. McAdams, and S. Winsberg. 1994. Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. *Journal de Physique* 4 (C5): 625-628.

Martin, K. D. 1999. *Sound-source recognition: A theory and computational model.* Ph.D. dissertation. MIT.

Pollard, H. F., and E. V. Jansson. A tristimulus method for the specification of musical timbre. *Acustica* 51 (3): 162-71.

Puckette, M. S., T. Apel, and D. D. Zicarelli. 1998. Real-time audio analysis tools for Pd and MSP. *Proceedings of the International Computer Music Conference*, 109–12. San Francisco: ICMA.

Sandell, G. J. 1994. SHARC timbre database. `http://www.parmly.luc.edu/sharc.`

Wettschereck, D., D. W. Aha, and T. Mohri. 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* 11 (1–5): 272–314.