Automatic Music Genre Classification Using Ensemble of Classifiers

Carlos N. Silla Jr., Celso A. A. Kaestner, Alessandro L. Koerich

Abstract— **This paper presents a novel approach to the task of automatic music genre classification which is based on multiple feature vectors and ensemble of classifiers. Multiple feature vectors are extracted from a single music piece. First, three 30-second music segments, one from the beginning, one from the middle and one from end part of a music piece are selected and feature vectors are extracted from each segment. Individual classifiers are trained to account for each feature vector extracted from each music segment. At the classification, the outputs provided by each individual classifier are combined through simple combination rules such as majority vote, max, sum and product rules, with the aim of improving music genre classification accuracy. Experiments carried out on a large dataset containing more than 3,000 music samples from ten different Latin music genres have shown that for the task of automatic music genre classification, the features extracted from the middle part of the music provide better results than using the segments from the beginning or end part of the music. Furthermore, the proposed ensemble approach, which combines the multiple feature vectors, provides better accuracy than using single classifiers and any individual music segment.**

I. Introduction

With the continuous expansion of the Internet, a huge quantity of data from different sources have been become available on-line. An study from the UC Berkeley shows that in 2002 there were about five million terabytes of new information produced in films, printed media or magnetic/optic storage media [1]. In the Web alone, more than 170 terabytes of information is available. However it is very difficult to use in an efficient manner such a huge amount of information. Many important problems such as search for information sources, retrieval/extraction of information, automatic summarization of information, etc. have been the subject of intensive research in the last years.

In this context, a research area that has been growing in the past few years is the multimedia information retrieval which aims at building tools to effectively organize and manage the great quantity of multimedia information available [2], [3]. The current practice for indexing multimedia data is based on textual meta-data information, which is the case of the ID3 tags in MP3 music files. Although ID3 tags are very useful for indexing, searching, and retrieval, usually, such tags are manually generated and associated with the multimedia data.

One of the most important types of multimedia data distributed over the Web is the digital music in MP3 format. There are many studies and methods related to the analysis of the music audio signal [3], [4], [5], [6], [7]. One important component for a content-based music information retrieval system is a module for the automatic music genre classification [8]. Music genres are categorical labels created by humans in order to determine the style of music. These labels are related to the instrumentalization, rhythmic structure and harmonic content of the music. Even if the music genre is a somewhat ambiguous descriptor, it has been used to categorize and organize large collections of digital music [3], [7], [9].

The issue of automatic music genre classification as a pattern recognition problem has been brought in the work of Tzanetakis & Cook [9]. They proposed a comprehensive set of features to represent music signals. These features were used to train three different types of classifiers: Gaussian Classifier, Gaussian Mixture Models and *k* nearest neighbors (k-NN). The feature set proposed is composed by timbral texture features, beat-related features and pitch-related features. The experiments were evaluated on a dataset containing 1.000 songs from 10 distinct genres (100 songs per genre). The initial accuracy achieved on this dataset was about 60% using a hundred iterations of a ten-fold cross-validation evaluation model. It is important to notice that the experiments were performed considering only the 30-first seconds of each music piece. Another interesting aspect of this work is that the feature set is available as part of the MARSYAS Framework[1], a free software framework for development and evaluation of computer audio applications [10]. The work of Tzanetakis and Cook has motivated the research and development of novel approaches to the task of automatic music genre recognition.

Kosina [11] developed MUGRAT[2] – a prototype system for musical genre recognition, using a subset of the features proposed by Tzanetakis. The evaluation of the MUGRAT was done on 189 music pieces from three genres: Metal (63), Dance (65) and Classical (61). For each music piece the feature vectors were obtained from three seconds long segments extracted randomly. A 3-NN classifier achieved the accuracy of 88.35% using a stratified ten-fold cross-validation approach. When building the dataset for this experiment Kosina has confirmed that manually-made genre classification is really inconsistent: MP3 files of the same song gathered from three different sources have presented different ID3 genre tag information. This fact confirms that

C. N. Silla Jr. and A. L. Koerich are with Postgraduate Program in Computer Science (PPGIa), Pontifical Catholic University of Paraná (PUCPR), R. Imaculada Conceição, 1155, Curitiba, PR, 80215-901, Brazil. silla@ppgia.pucpr.br, alekoe@ppgia.pucpr.br

C. A. A. Kaestner is with the Department of Informatics (DAINF), Technological Federal University of Paraná (UTFPR), R. Sete de Setembro, 3165, Curitiba, PR, 80230-901, Brazil. kaestner@dainf.cefetpr.br

[1] Music Analysis, Retrieval and Synthesis for Audio Signals, available at: http://marsyas.sourceforge.net/

[2] MUsic Genre Recognition by Analysis of Texture, available at: http://kyrah.net/mugrat/

1-4244-0991-8/07/$25.00/©2007 IEEE

the use of the ID3 tags is not suitable for music genre classification. Li et al. [6] proposed a novel method for feature extraction based on the Daubechies Wavelet Coefficients Histogram (DWCH) and compared it with the feature set proposed in [9]. In this work the classifiers evaluated were support vector machines (SVM), k-NN, GMM and linear discriminant analysis. The best results were achieved using the SVM classifier. An unsupervised approach using hidden Markov models (HMMs) was proposed in the work of Shao et. al [12]. The idea of decomposing and ensembling specialized classifiers have also been used for music genre classification in the work of Grimaldi et al. [13], [14]. In their work they have carried out experiments using different ensemble strategies and feature selection techniques. They have evaluated the performance of OAA, Pairwise Comparison (also referred as Round Robin) and Random Subspace method [15] (also referred as Feature Subspace), with some feature ranking approaches for feature selection, namely Principal Component Analysis (PCA), Information Gain and Gain Ratio. They have performed these experiments on a dataset of 200 music pieces of five classes (Jazz, Classical, Rock, Heavy Metal and Techno) employing a 5-fold cross-validation procedure. All experiments were carried out using only the k-NN classifier. To extract features from the music signal they have used a Discrete Package Wavelet Transform (DPWT), which was applied to the entire music piece.

One common aspect of most works in the area is that they often use only one feature vector extracted from a music segment (usually thirty seconds). One of the few exceptions is the work of Costa et al. [16] which introduced the idea of segmenting the music audio signal into three 30-second segments, training a classifier for each music segment, and combining the classifiers decision in order to improve the final prediction about the music genre. In this work the segmentation method was evaluated employing a k-NN and a multilayer perceptron neural network (MLP) classifier.

The main motivation of this work is to analyze Latin music audio signals, which present a great variation in time. To account for such a variation one of the possible hypothesis, which is also investigated in this paper, is that feature vectors generated from the whole music signal provide better results relative to feature vectors generated only from short segments even if it is known that this is time consuming and computational expensive. In order to overcome this problem, the strategy that is often adopted is extracting features only from parts of the music. However, this approach is not reliable since the classification of different parts of a music piece can lead to different classification outputs and different error rates. For this reason, in this work we present an extension of the approach proposed by Costa et al. [16] with other learning algorithm (decision trees, SVM and Naïve Bayes), different feature set and ensemble of classifiers with the aim of improving the accuracy in the classification of music genres, in special, Latin music genres.

The experiments are carried on a large dataset which is composed of more than 3,000 music samples from ten different Latin music genres. The reason for considering Latin music is because we believe that the development of tools for different music styles is as important as the development of tools for other languages than English. For music, the main reason is that different music genres have different influences and instrumentalization.

This paper is organized as follows. Section II presents an overview of the proposed approach for music genre classification which considers several feature vectors extracted from the same music piece. Section III gives a brief description about the features that are extracted from the music signal. The problem of music genre classification is presented in Section IV while the ensemble strategies that are used to combine classifier outputs are presented in Section V. Section VI reports the experiments on a large database of Latin music as well as an analysis of the results. Conclusions are stated in the last section and some perspective about future work.

II. System Overview

The Latin music genre classification system proposed in this paper is composed of three main phases (Fig.1): feature extraction, classification and decision based on an ensemble of individual classifiers. First, features are extracted from three 30-second music segments taken from the audio signal. These segments are chosen from the beginning, middle and end part of the music since for many music pieces, the audio signal has a great variability in time. In this way each music segment is represented by a feature vector.

Since this is a system that employs supervised learning algorithms, it operates in two modes: training and classification. In the training mode the feature vectors are used together their respective labels by the learning algorithms. The labels consist in the textual information that represents the musical genre assigned to the music by human experts. In the classification mode, a music piece whose genre is unknown is provided to the system. Similarly to the training mode, three 30-second music segments are selected and for each of such music segments, feature vectors are generated. Each feature vector feeds a individual classifier which, at the end, will assign a genre to the feature vector (music piece). The output of the classifiers are then fused through some combination rules such as the majority vote, max, sum, and product rule. Based on the results of such a combination, a musical genre is assigned to the music piece. Fig.1 illustrates such a process.

In the next sections the most important components of the proposed approach are described, such as the feature extraction process and the feature set, the supervised learning algorithms that are further used as classifiers, and the ensemble method that combines the output of the classifiers.

III. Feature Extraction

In this work the problem of automatic music genre recognition is viewed as a pattern recognition problem where a music sample is represented in terms of feature vectors. The aim of feature extraction is to represent a music piece into

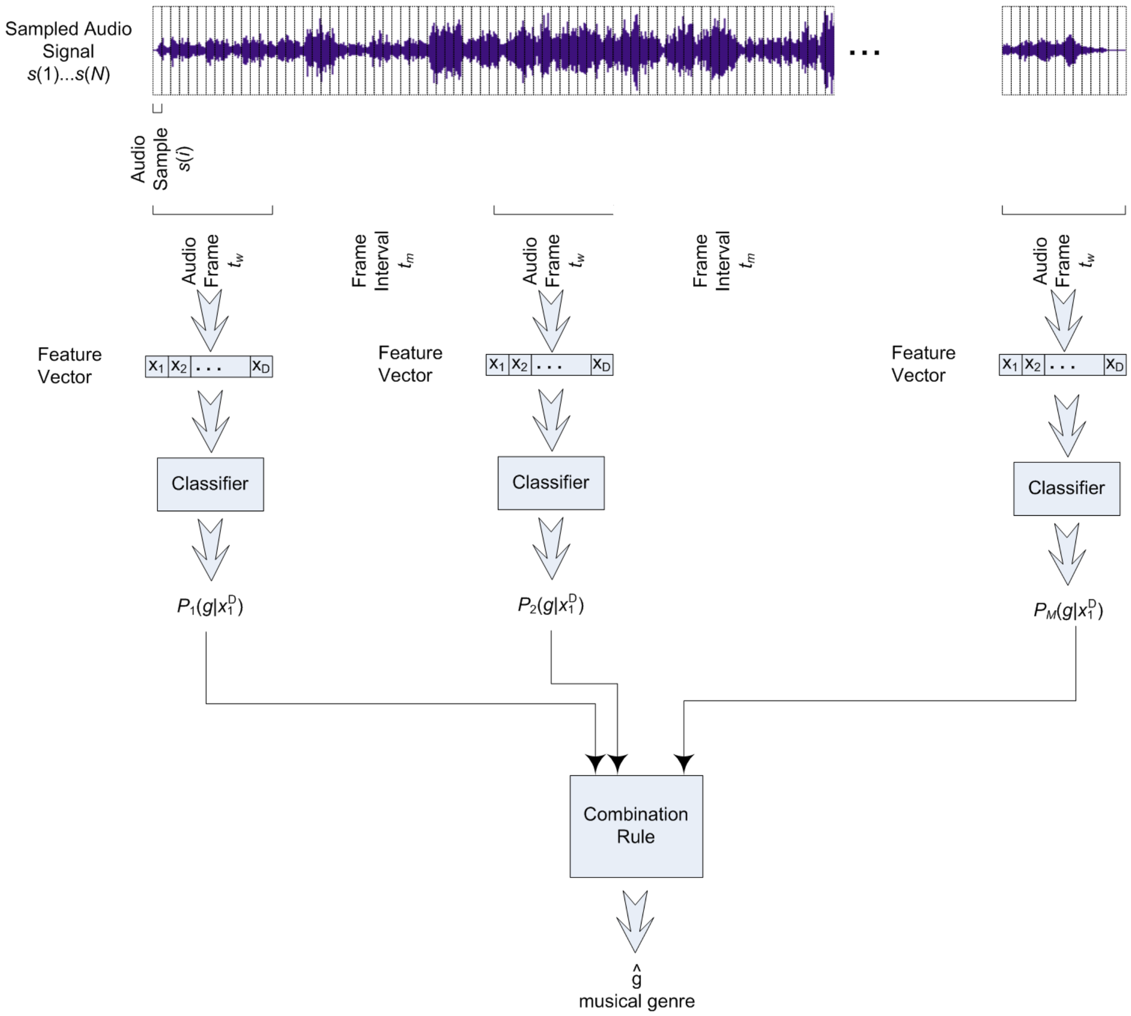

Fig. 1. An overview of proposed approach for music genre classification: feature extraction from several segments of the music signal, classification of each feature vector by an individual classifier, and combination of classifier outputs.

a compact and descriptive way and that is suitable to deal with learning algorithms.

Since digital music of good quality has about 1MB per minute, the extraction of features from the whole music can be prohibitive due to the required processing time. For that reason features are extracted from three 30-second music segments. The music segments, denoted as audio frame (t_w) in Figure 1 have the same duration, which is equivalent to 1,153 audio samples, or simply $t_w = 1,153$, in MP3 format. It is important to notice that regardless of the bitrate of the file, when dealing with MP3 files, the number of audio samples (which denotes the duration of the music) is always the same [17]. For this reason we use the following strategy to extract features from three music segments of a music sample:

- The first segment is extracted from the beginning of the music, from audio sample $s(0)$ to audio sample $s(1153)$;
- Let N denotes the total number of audio samples of a music, the second segment is extracted from the middle of the music, from audio sample $s(N/3+500)$ to audio sample $s(N/3+1653)$;
- The third segment is extracted from the end part of the music but a particular strategy is adopted to avoid getting noisy or silenced endings that are common in some MP3 files. Then, the third segment is extracted from audio sample $s(N-1453)$ to audio sample $s(N-300)$.

For the extraction of features from the music segments, the MARSYAS [10] framework was employed. The MARSYAS framework implements the original feature set proposed by Tzanetakis and Cook [9]. The features used can be divided into three groups: Timbral Texture, Beat Related and Pitch Related. The features based on the Timbral Texture are extracted based on the means and variance of the spectral centroid, rolloff, flux, the time zero domain crossings, the first five MFCCs and low energy. Features that are beat-related include the relative amplitudes and the beat per minute. Pitch related features include de maximum periods of the pitch peak in the pitch histograms. The final feature vector concatenates all these features into a 30-dimensional feature vector (timbral texture: nine FFT and ten MFCC; beat: six; pitch: five) [9].

IV. Classification

Formally we can define a digital audio signal as a sequence $S =< s(1), s(2), \ldots, s(N) >= s_1^N$ where $s(i)$ represents the signal sampled at the instant i, and N is the total number of samples that form the digital audio stream.

The problem of music genre classification can now be defined. In order to apply a pattern recognition approach, we extract several features from the digital audio signal S. If we consider D features, the digital audio signal S can be represented by a D-dimensional feature vectors. We denote a sequence of M feature vectors of the digital music signal as

$$X_t =< \bar{x}_D(1), \bar{x}_D(2), \ldots, \bar{x}_D(m), \ldots, \bar{x}_D(M) > \quad (1)$$

where each component $\bar{x}_D(m)$ represents an appropriate feature vector related to the segment m, where $m = 1, 2, \ldots, M$.

In the classification problem we wish to assign a class (i.e. a musical genre) $g \in \mathcal{G}$ which better represents the music given by the digital audio signal S. $\mathcal{G}$ denotes the set of all possible music genres. This problem can be framed from a statistical perspective where the goal is to find the musical genre g that is most likely, given the feature vector $\bar{x}_D(.)$.

$$\hat{g} = \arg\max_{g \in \mathcal{G}} P(g|\bar{x}_D(.)) \quad (2)$$

where $P(g|\bar{x}_D(.))$ is the *a posteriori* probability of a music genre g given a feature vector $\bar{x}_D(.)$ and it can be rewritten using Bayes' rule:

$$P(g|\bar{x}_D(.)) = \frac{P(\bar{x}_D(.)|g)P(g)}{P(\bar{x}_D(.))} \quad (3)$$

where $P(g)$ is the *a priori* probability of the musical genre, which is estimated from frequency counts in the training data set. The probability of data occurring $P(\bar{x}_D(.))$ is unknown, but assuming that the genre g is in $\mathcal{G}$ and that the classifier computes the likelihoods of the entire set of possible hypotheses (all musical genres in $\mathcal{G}$), then the probabilities must sum to one:

$$\sum_{g \in \mathcal{G}} P(g|\bar{x}_D(.)) = 1 \quad (4)$$

In such a way, estimated *a posteriori* probabilities can be used as confidence estimates [18]. Then, we obtain the posterior $P(g|\bar{x}_D(.))$ for the genre hypothesis as:

$$P(g|\bar{x}_D(.)) = \frac{P(\bar{x}_D(.)|g)P(g)}{\sum_{g \in \mathcal{G}} P(\bar{x}_D(.)|g)P(g)} \quad (5)$$

In this work we have used the following machine learning algorithms as component classifiers for the ensemble methods: Naïve Bayes [19], Support Vector Machines [20] with the pairwise classification decomposition strategy and multilayer perceptron (MLP) neural network trained with the backpropagation momentum algorithm. These machine learning algorithm were chosen because they are in accordance with the probabilistic framework described above, since they provide at the output, a posteriori estimates, given a feature vector as input pattern.

The Naïve Bayes classifier is based on the Bayes Rule but naively assumes independence between the attributes. The Naïve Bayes classifier can also support handle multi-class problems. The MLP neural network is composed of thirty neurons in the input layer (one for each attribute), twenty neurons in the hidden layer, and ten neurons in the output layer (one for each class). The neural network classifiers can work with high dimensional planes and create make shapes of dividing the data. The layout of the network is important to the problems at hand and it can also be customized to work with multi-class problems. The support vector machine (SVM) classifier is an interesting machine learning algorithm for create a maximum hyper-plane that divides two regions in the feature space. It is commonly used in two class problems, and for that reason it is needed to use some decomposing strategy to handle multi-class problems. In this work we have used pairwise classification as the decomposing scheme for a linear support vector machine trained with the sequential minimum optimization algorithm [21].

V. Ensemble Method

The two main reasons for combining classifiers are efficiency and accuracy [22]. Kittler et al. distinguish from two different scenarios for classifier combination. In the first scenario, all the classifiers use the same representation of the input pattern. Although each classifier uses the same feature vector, each classifier will deal with it in different ways. They illustrate this with two examples: the first one would be using a set of k-NN classifiers where each classifier has a different value for the number of k nearest neighbors; the second example would be using a set of neural networks, where each network is trained with a different learning algorithm. In the second scenario, each classifier uses its own representation of the input pattern.

In this work we propose a novel ensemble-based approach that is related to the second scenario, based on the segmentation strategy presented in Section II. We use several representations of the digital audio signal, since each segment generates a different feature vector. When using this segmentation strategy it is possible to train a specific classifier for each one of the segments, and to compute the final decision about the class (assigning a class label in this context the music genre) from the ensemble of the results provided by each classifier.

A sequence of M feature vectors of the digital music signal is denoted in Eq. 1 in which each $\bar{x}_D(m)$ is an appropriate feature vector related to the segment m, where $m = 1, 2, \ldots, M$. Similarly, we denote a set of M component classifiers as:

$$C =< c(1), c(2), \ldots, c(m), \ldots, c(M) > \qquad (6)$$

without loss of generality we assume that this is a set of homogeneous probabilistic classifiers, whose output of each classifier is *a posteriori* probability estimate denoted as $P(g|\bar{x}_D(.))$, where $\sum_{g \in \mathcal{G}} P(g|\bar{x}_D(.)) = 1$. $\mathcal{G}$ denotes the set of all possible music genres. The relationship between $\bar{x}_D(.)$ and C is straightforward, i.e., it is an one-to-one relationship, and the feature vector $\bar{x}_D(m)$ of the sequence of vectors X_t is classified by the component classifier $c(m)$ from C.

In order to find the best ensemble of classifiers, i.e., the most diverse set of classifiers that brings a good generalization, we have used a single objective function, namely, maximization of the recognition rate of the ensemble. To combine the decisions of the component classifiers trained on each music segment of the same music sample, their outputs are taken. The combination of the results is achieved through the majority voting rule, max rule, sum rule and product rule. The majority vote is a simple decision rule where only the class labels are taken into account and the one with more votes wins:

$$\hat{g} = \underset{\substack{g \in \mathcal{G} \\ m \in [1,\ldots,M]}}{maxcount} \; P_m(g|\bar{x}_D(m)) \qquad (7)$$

where $maxcount$ returns the most frequent value of a multiset. In the *max* rule, the class with the highest confidence score is chosen:

$$\hat{g} = \underset{\substack{g \in \mathcal{G} \\ m \in [1,\ldots,M]}}{\arg\max} \; P_m(g|\bar{x}_D(m)) \qquad (8)$$

The *sum* rule is based on the output probabilities for all classes from each classifier; the probabilities are summed up for each class and the class with the highest value is chosen:

$$\hat{g} = \arg\max_{g \in \mathcal{G}} \sum_{m=1}^{M} P_m(g|\bar{x}_D(m)) \qquad (9)$$

The *product* rule is based on the output probabilities for all classes from each classifier; the probabilities are multiplied for each class and the class with the highest value is chosen:

$$\hat{g} = \arg\max_{g \in \mathcal{G}} \prod_{m=1}^{M} P_m(g|\bar{x}_D(m)) \qquad (10)$$

In the next section the results of the ensembles strategy are evaluated relative to the conventional classification approaches that use single feature vectors and classifiers.

VI. Experiments

We have selected 3,000 music samples from ten different Latin musical genres (Tango, Salsa, Forro, Axe, Bachata, Bolero, Merengue, Gaucha, Sertaneja, Pagode) and split them into balanced datasets. The training dataset is composed by 150 samples from each musical genre, summing up to 1,500 samples (50%); the validation dataset is composed by 60 samples from each musical genre, summing up to 600 samples (20%); and test dataset is composed by 90 samples from each musical genre, summing up to 900 samples (30%). The total number of artists represented in this whole dataset is 543. It is important to notice that to avoid any biasing in the experiments, all the available music has been random selected without reposition from the database. Another important aspect of this dataset is that each music sample was labeled by an human expert after manual inspection. Regardless of Pachet's suggestion [7] of using CDs from collections of CDs or theme, in the case of Latin music this approach is inefficient for labeling.

TABLE I
Music genre classification accuracy using single music segments and single classifiers

Classifier	Music Genre Classification Accuracy (%)		
	1st segment	2nd segment	3rd segment
J4.8	39.60	**44.44**	38.80
3-NN	45.83	**56.26**	48.43
MLP	53.96	**56.40**	48.26
Naïve Bayes	44.43	**47.76**	39.13
SVM	57.43	**63.50**	54.60

TABLE II
Music genre classification accuracy using ensemble of classifiers

Classifier Ensemble	Music Genre Classification Accuracy (%)
J48	47.33
3-NN	60.46
MLP	59.43
Naïve Bayes	46.03
SVM	65.06

Table I shows that in the case of Latin rhythms using only the beginning music segment is not a good strategy. In all cases, the best results were achieved on the middle segment, and in all other cases there is no pattern for the second best classification accuracy since it was achieved using sometimes the second or the third segment.

The results achieved using the method of combination and decision based on the majority vote rule are presented in Table II. For the majority of the classifiers used in the ensembles, the correct music genre classification rate is greater relative to the results provided by single classifiers that take into account only single music segments. The only exception is the case of the Naïve Bayes classifier. In the case of the J48 and MLP classifier, the accuracy was improved in more than 3%; more than 4% for the 3-NN classifier, and about 1.5% for the SVM classifier.

As mentioned earlier, this method of ensemble and decision based on three music segments extracted from each music sample was originally proposed in [16] where the method was evaluated using music from the musical genres Rock and Classic. In the previous experiments, the results were not improved significantly using this method. However in this work the music samples used are from different musical genres which seem to benefit from the ensemble strategy adopted. This might be due to the nature of the

TABLE III

MUSIC GENRE CLASSIFICATION ACCURACY USING ENSEMBLE OF CLASSIFIERS AND THE MAX, SUM AND PRODUCT RULES.

Classifier Ensemble	**Music Genre Classification Accuracy (%)**		
	MAX	SUM	PROD
MLP	57.40	61.83	**62.50**
Naïve Bayes	45.96	**46.66**	46.13
SVM	64.13	**65.73**	65.50

genres, since Rock and Classic are usually more constant than Latin rhythms. In the case of Salsa, most of music samples starts slow (sometimes as slow as a Bolero) in the introduction and after a while they "explode"(at the time when all instruments come into play). The results are in accordance with the positioning of Li et al. [8] who states that different strategies are needed for the classification of different music genres when some sort of hierarchical classification is taken into account. This indicates that the strategy of segmenting the music piece into three segments and the combination from the ensemble of the classifiers trained in these segments might be more appropriate to use with specific genres or sub-genres. Unfortunately a direct comparison with the experiments performed earlier in [9], [6] is not possible due to the fact that although the data set used is available it contains only the first thirty seconds of each music sample.

We have also investigated the impact of using other combination rules that make use of the output probabilities provided by each individual classifier. To be consistent with the probabilistic framework described in Section IV only the classifiers that provide *a posteriori* probability estimates were considered in the ensemble. Table III shows the results of using the max, sum and product rule to combine the three individual classifiers. We can observe that these combination rules have further improved the performance of the ensemble approaches relative to the individual classifiers.

VII. CONCLUDING REMARKS

In this paper we have presented an evaluation of different classifiers with an ensemble technique applied to three different segments of the same music piece for the task of automatic music genre classification. The genres considered in the experiments were ten different Latin genres, namely Tango, Salsa, Forro, Axe, Bachata, Bolero, Merengue, Gaucha, Sertaneja, Pagode. The results achieved on large dataset composed by 3,000 music samples have shown that the ensemble approach also provides a more accurate genre classification relative to the individual classifiers. The improvement in accuracy depends on the nature of the individual classifier and ranges from 1% to 7%.

An analysis of the results achieved shows that without the ensemble approach, the music segment from the middle of the music piece is always the one that provides the best classification accuracy. This is an interesting finding since most works in the literature [9] considers segments only from the beginning (the first 30-seconds) of each music sample.

As future work, we plan to use more sophisticated combination rules to weight the output of the classifiers because we have observed that the classifier that takes the middle segment is always more accurate than the classifiers that deal with the beginning or end part of a music piece.

REFERENCES

[1] P. Lyman and H. R. Varian, "How much information," Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on [06/25/2005], 2003.

[2] M. Fingerhut, "The ircam multimedia library: A digital music library," in *IEEE Forum on Research and Technology Advances in Digital Libraries*, 1999, pp. 19–21.

[3] E. Pampalk, A. Rauber, and D. Merkl, "Content–based organization and visualization of music archives," in *ACM Multimedia 2002*, Juan-les-Pins, France, 2002, pp. 570–579.

[4] G. Guo and S. Z. Li, "Content–based audio classification and retrieval by support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 209–215, 2003.

[5] T. Zhang and C. C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 441–457, 2001.

[6] Tao Li, Mitsunori Ogihara, and Qi Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, Canada, 2003, pp. 282–289.

[7] J. J. Aucouturier and F. Pachet, "Representing musical genre: A state of the art," *Journal of New Music Research*, vol. 32, no. 1, pp. 83–93, 2003.

[8] Tao Li and M. Ogihara, "Music genre classification with taxonomy," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, vol. 5, pp. 197–200.

[9] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[10] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis," *Organized Sound*, vol. 4, no. 3, 2000.

[11] Karin Kosina, "Music genre recognition," Tech. Rep., Fachlochschul Hagenberg, 2002.

[12] Xi Shao, Changsheng Xu, and Mohan S. Kankanhalli, "Unsupervised classification of music genre using hidden markov model," in *IEEE International Conference on Multimedia and Expo*, June 2004, vol. 3, pp. 2023–2026.

[13] Marco Grimaldi, Pdraig Cunningham, and Anil Kokaram, "A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques," in *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*. 2003, pp. 102–108, ACM Press.

[14] Marco Grimaldi, Pdraig Cunningham, and Anil Kokaram, "An evaluation of alternative feature selection strategies and ensemble techniques for classifying music," in *Workshop on Multimedia Discovery and Mining at ECML/PKDD-2003*, 2003.

[15] Tim K. Ho, "Nearest neighbors in random subspaces," in *Proc. of the 2nd Intĺ Workshop on Statistical Techniques in Pattern Recognition*, 1998, pp. 640–648.

[16] C. H. L. Costa, J. D. Valle Jr, and A. L. Koerich, "Automatic classification of audio data," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2004, pp. 562–567.

[17] Scot Hacker, *MP3: The Definitive Guide*, O'Reilly, 1st edition, 2000.

[18] A. Stolcke, Y. Konig, and M. Weintraub, "Explicit word error minimization in N-best list rescoring," in *Proc. Eurospeech '97*, Rhodes, Greece, 1997, pp. 163–166.

[19] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[20] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

[21] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, and K.R.K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Computation*, vol. 13, no. 3, pp. 637–649, 2001.

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