2014 IEEE International Conference on Systems, Man, and Cybernetics
October 5-8, 2014, San Diego, CA, USA

Audio Feature Reduction and Analysis for Automatic Music Genre Classification

Babu Kaji Baniya, Joonwhoan Lee
Department of Computer Science and Engineering
Chonbuk National University
South Korea
everwith_7,chlee@jbnu.ac.kr

Ze-Nian Li
School of Computer Science
Simon Fraser University
Canada
li@cs.sfu.ca

Abstract— **Multimedia database retrieval is growing at a fast rate thereby subsequent increase in the popularity of online retrieval system. The large datasets are major challenges for searching, retrieving, and organizing the music content. Therefore, there is a need of robust automatic music genre classification method for organizing these music data into different classes according to the certain viable information. There are two fundamental components to be considered for genre classification namely audio feature extraction and classifier design. In this paper, diverse audio features set have been proposed to characterize the music contents precisely. The feature sets belong to four different groups, i.e. dynamic, rhythm, spectral, and harmony. From the features, five different statistical parameters are considered as representatives, including up to the 4^{th} order central moments of each feature, and covariance components. Ultimately, significant numbers of representative attributes are controlled by MRMR algorithm. The algorithm calculates the score level of all feature attributes and orders them. The high score feature attributes are only considered for genre classification. Moreover, we can visualize that which audio features and which of the different statistical parameters derived from them are important for genre classification. Among them, mel frequency cepstral coefficients (MFCCs) have higher scored level than other feature attributes. Furthermore, MRMR does not transform the feature value like as principal component analysis (PCA). Besides these, the comparison has been made based on classification accuracy between two-dimensionality reduction methodologies using support vector machine (SVM). The classification accuracy of MRMR feature reduction set outperforms than PCA. The overall classification is also higher than other existing state-of-the-art of frame base methods.**

Keywords—music genre; dimensionality; feature reduction; statistical parameters; MRMR

I. INTRODUCTION

Music genre can be viewed as categorical labels created by musicians or composers in order to find the content (style) of the music. It is a consistently growing research area in the field of information retrieval task since it can be applied for practical purposes such as efficient organization and categorization of online data collection. It is essential to design automatic grouping tools which provide meaningful and efficient way of describing the music genre classification. There have already been several well-known approaches in this area. The efficient and accurate automatic music information processing remains as the key issue, so it has been consistently attracting the attention of researchers and musicologists. Based on these facts, there are still enough room for improvement in automatic genre classification. The main concern lies on describing, organizing, and categorizing music content in the Internet. This can be achieved by finding the important characteristics of audio signal. The audio signals of music belonging to the same genre mean they share certain common characteristics, because they are composed of similar types of instruments, having similar rhythmic patterns, and similar pitch distributions [1]. Moreover, the extracted features must be comprehensive, compact, and effective.

Many different features have been introduced for music genre classification. The primary aim of feature extraction is to acquire meaningful information from songs in the reduced form. The acoustic features include tonality, pitch, beat, symbolic features extracted from the scores, and text based features extracted from the song lyrics. The content based acoustic features are divided into timbre texture, rhythmic content, and pitch content features [2]. Timbre features are often calculated for every short-time frame (10ms~100ms) of sound based on the Short Time Fourier Transform (STFT) [3]. Timbral texture features contain Mel-Frequency Cepestral Coefficients (MFCCs), spectral centroid, spectral flatness, spectral flux, spectral rolloff, zero crossing, energy, and Linear Prediction Coefficients (LPCs) [2, 3]. These features are widely used in different area based on the requirement of applications. MFCCs have been extensively used in speech recognition [4]. Later, MFCC features are also used for discriminating the music. Rhythmic content features possess information about continuity of rhythm, beat and tempo. Furthermore, tempo and beat tracking are extensively used in music search and retrieval systems. The tempo value is a number which represents the speed of song or music measured by beats per minute (bpm) [5, 6]. Pitch content feature deals with frequency information of music.

Tzanetakis and Cook [2] proposed a comprehensive set of features for direct modelling of music signals and explore the different application of those features for musical genre classification using *k*-Nearest Neighbor (*k*-NN) and Gaussian Mixture Model (GMM). Other researchers like Lambrou et al. [7] used statistical features in the temporal and the three different wavelets transform domains to classify music into rock, piano, and jazz. Soltau et al. [8] proposed an approach of

978-1-4799-3840-7/14/$31.00 ©2014 IEEE

representing temporal structures of an input signal. He showed that this new set of abstract features can be learned via artificial neural networks (ANNs) [9] and can be used for music genre identification. Deshpande et al. [10] used Gaussian Mixtures, Support Vector Machines and *k*-Nearest Neighbor to classify the music into rock, piano, and jazz based on timbral texture features.

The music genre classification with reduced feature sets using support vector machine classifier is shown in the Figure 1. It represents the overview of genre classification in the paper. Basically, there are two problems need to be addressed in music genre classification i.e. audio feature extraction and classifier design. Besides these, feature analysis also plays a significant role for improving the classification accuracy in genre classification. The feature analysis means finding out the most discriminative feature or set of features among the extracted total features. In this scheme, minimum redundancy maximum relevance approach implements for feature reduction. It gives the maximum relevance value as a score of each feature statistics. Based on the requirement, we can select how many feature statistics are needed to achieve maximum performance. Furthermore, the unique aspect of MRMR [11] is to keep the original features as it is. On the contrary, PCA [12] transforms the original features and order them based on variance difference. Later, we select the desired number of the feature statistics to achieve the maximum classification accuracy.

Temporal feature integration is an alternative approach to combine information. It uses a sequence of short-time feature vectors to create a single new feature vector at a larger time scale. Temporal feature integration is a popular common technique. The most basic statistics like mean and variance of short-time features have been used [13], [14]. In this paper, we have extracted four different audio features namely dynamic, rhythmic, spectral, and harmony. These features are integrated not only mean and variance but also higher order moments like skewness and kurtosis. In addition, covariance components of pairwise features are also included within the same group of features. Consequently, the feature dimension increases sharply and some of them are not meaningful for further consideration. MRMR is used to control unnecessary feature attributes from whole attributes. It gives the score value of each attribute in a sequence. The higher score attributes are considered for classification. Moreover, it also gives us which audio features and their respective statistical parameters are important of genre classification is shown in Table 3. Later, we compared the classification accuracy of reduced feature using two feature reduction methodologies MRMR and PCA. The classification accuracy of GTZAN dataset through MRMR reduced features is comparatively higher than PCA.

The outline of the paper is as follows. Feature extraction which is the critical portion of genre classification is described in section II. Section III deals with the feature reduction using two different approaches i.e. MRMR and PCA, similarly, section IV briefly explains the experimental setup and data preparation, and section V thoroughly explains the classification result and analysis. Finally, section VI describes the conclusion of the proposed method and future work for the genre classification.

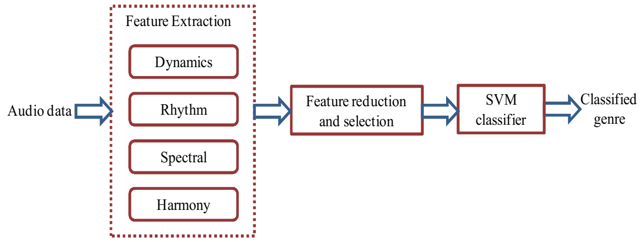

Figure 1: Block diagram of proposed method for Music genre classification

II. FEATURE EXTRACTION

Feature extraction encompasses the analysis and extraction of meaningful information from audio in order to obtain a compact and concise description that is machine-readable. Features are usually selected in the context of a specific task and domain. The features used in our research are divided into four categories: dynamic, rhythmic, spectral and tonal. They belong to low-level and high-level according to the frame size. The frame length for the low-level and high-level features in the experiment were 46ms and 2s, respectively, with both having 50% overlap. For feature extraction, we utilized the MIR Toolbox [15].

From each of the frame-based feature, we have extracted up to the 4^{th} order central moments for a music, such as mean, variance, skewness, and kurtosis. Because it has recently been reported that the higher order moments (skewness and kurtosis) are also useful for the genre classification of music [16], we include them for the genre classification. The high-order moments increase the classification accuracy when it is combined with other low order ones. It generally provides the supplementary statistical information for the audio signal.

In addition, covariance components of pairwise frame-based features also included [17], because it has proved to be meaningful for the genre classification. . The statistics like covariance, however, are calculated by each group of features. In the spectral category, covariance component like MFCC (Mel frequency cepstral coefficient), DMFCC (Delta mel frequency cepstral coefficient) [18], and DDMFCC (Delta-delta mel frequency cepstral coefficient) are calculated separately.

Different frame-based features listed in Table 1 are computed. One can consider following statistics for the frame-based feature: The mean and standard deviation can be calculated by eq. (1) and (2) respectively.

$$Mean(\mu) = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{1}$$

$$Std(\sigma) = \sqrt{\frac{\sum_{n=1}^{N}(X_n - \mu)^2}{N}} \tag{2}$$

The skewness [19] is a measure of asymmetry of the distribution, which represents the relative disposition of the tonal and non-tonal components of each band. If the tonal components occur frequently in a band, the distribution of its spectrum will be left-skewed otherwise it will be rightskewed. Mathematically, the skewness in a music piece can be defined as

TABLE I. EXTRACTED AUDIO FEATURE SETS
(M = MEAN, SD = STANDARD DEVIATION, *S* = SKEWNESS, K = KURTOSIS, PC = PAIRWISE COVARIANCE, HCDF = HARMONIC CHANGE DETECTION FUNCTION)

No.	Category	Feature	Acronyms
1	Dynamic	RMS energy	M, S_{td}, S_k K_t, Cov
2		Slope	,,
3		Attack	,,
4	Rhythm	Tempo	M, S_{td}, S_k K_t
5	Spectral	Spectral centroid	M, S_{td}, S_k K_t, Cov
6		Brightness	,,
7		Spread	,,
8		Rolloff85	,,
9		Rolloff95	,,
10		Spectral entropy	,,
11		Flatness	,,
12		Irregularity	,,
13		Roughness	,,
14		Zerocrossing	,,
15		Spectral flux	,,
16		MFCC(1~13)	,,
17		DMFCC (1~13)	,,
18		DDMFCC (1~13)	,,
19	Harmony	Chromagram peak	,,
20		Chromagram centroid	,,
21		Key clarity	,,
22		Key mode	,,
23		HCDF	,,

$$Skewness = \frac{\sum_{n=1}^{N} (X_n - \mu)^3}{(N-1)\sigma^3} \quad . \tag{3}$$

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean. The kurtosis measure can sketch the effective dynamic range of the audio spectrum. Mathematically it can be defined as

$$Kurtosis = \frac{\sum_{n=1}^{N} (X_n - \mu)^4}{(N-1)\sigma^4} - 3 \tag{4}$$

Therefore, we have considered 4*n* components for the *n* frame-based audio features.

Covariance is measured between two features. The aim of considering the covariance is usually to see if there is any mutual relationship between the features. It is useful to measure the polarity and the degree of the correlation between two features. The covariance of two features X_n and Y_n in a music piece is given as

$$Cov(X_n, Y_n) = \frac{1}{N} \sum_{n=1}^{N} (X_n - \mu_X)(Y_n - \mu_Y) \ , \tag{5}$$

where μ_X and μ_Y are corresponding means of X_n and Y_n , respectively. For *n* timbral texture features we acquired $n(n-1)/2$ mutual covariance values.

Therefore, the feature dimension for a piece of music increases sharply. However, it is not known whether all the statistics derived frame-based features are equally significant for genre classification or not, that could be evaluated by MRMR.

III. FEATURE REDUCTION

A. Minimum Redundancy Maximum Relevance (MRMR):

The MRMR criterion has been proposed in [11], [19] in combination with forward selection search strategy. Given a set X_s of selected variables, the criterion consists of updating X_s with the variable $X_i \in X_t$ (it is difference between the original set of input *t* and set of variables X_s selected so far) that maximizes u_i-z_i, where u_i is a relevance term and z_i is a redundancy term. Moreover, *I* is the mutual information of two variables, u_i is the relevance of X_i to the output *Y* alone, and z_i is the average redundancy of X_i to the selected variables $X_j \in X_s$.

$$u_i = I(X_i; Y) \tag{6}$$

$$z_i = \frac{1}{d} \sum_{X_j \in X_s} I(X_i; Y) \tag{7}$$

$$X_i^{MRMR} = \arg \max_{X_i \in X_t} \{u_i - z_i\} \tag{8}$$

At each step, this method selects the variable which has the best trade-off between relevance and redundancy. The selection criterion is fast and efficient. At step *d* of the forward search, the search algorithm computes *n*-*d* evaluations, such evaluation requires the estimation of (*d*+1) bivariate densities (one for each already selected variables plus one with the output). It has been shown in [11] that the MRMR criterion is an optimal first order approximation of the mix-dependency if a feature is selected at the time. Furthermore, MRMR avoids the estimation of multivariate densities by using multiple bivariate densities. Notice that, although the method addresses the bivariate redundancy issue through the term z_i, it is not able to take into account the complementarities between variables.

B. Principal Component Analysis(PCA):

It is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the principal component has largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components.

IV. EXPERIMENTAL SETUP AND DATA PREPARATION

The well-known dataset GTZAN widely used for music genre classification is employed for performance evaluation. The dataset collected by George Tzanetakis [20] consists of 1000 music pieces divided into ten different genres: Classical, Blues, Hiphop, Pop, Rock, Jazz, Reggae, Metal, Disco, and Country. Each class consists of 100 music pieces having durations of 30s. It has ten different classes and each has 100 songs. Each song in the database was stored as a 22050Hz, 16bits, and mono audio file. A ten-fold cross validation scheme [21] is used to evaluate the performance to obtain every experimental result in the paper.

There are numbers of audio features as shown in Table I, which can be extracted from MIRtoolbox [15]. The extraction has done with the frame-based approach, with the frame duration of 46ms, with 50% overlapping for most of the features. For high-level features that require longer frame (tempo, and harmonic related features), the size of the analysis window is 2s with 50% overlapping.

V. RESULT AND ANALYSIS

The proposed method improves the classification accuracy taking with minimum set of features having maximum discrimination power. Therefore, we analysed the audio features (statistics) into two stages i.e. lower (mean and standard deviation) and higher order statistics. There are 118 features in total while considering the lower order statistics. The motif is to reduce feature statistics by using PCA and MRMR without degrading the classification accuracy and their comparison. The classification accuracy is achieved by 80.75% only using two different statistics of each feature from whole set of features. The feature reduction is performed by using MRMR algorithm. Similarly, 75% of classification accuracy is obtained by using the PCA based feature reduction method. The number of features (statistics) is taken in multiple of 25 and their corresponding classification accuracy is shown below in Figure 2. The classification accuracy increases continuously in both cases (MRMR and PCA) till number of features (statistics) reaches up to 75. Thereafter, there is no notable improvement of classification accuracy in both scenarios. Finally, we concluded that the MRMR has the strong ability to select the better discriminative features (statistics) than PCA (transform features) for genre classification.

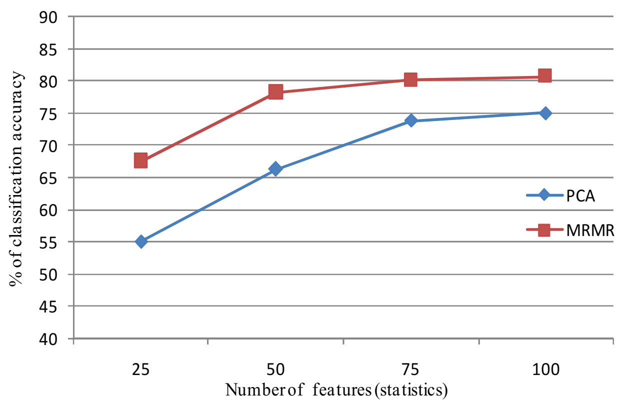

Figure 2: Number of features Vs classification accuracy only low order statistics

Figure 3 shows that classification accuracy including lower and higher order moments and covariance components. The feature dimension increases sharply in this experiment than previous one. Each music piece is expressed as 538-dimensional statistical feature in total. Therefore, the robust dimension reduction method is quite essential before the classification. Based on reduced feature sets, again same number of features (statistics) interval is considered as shown in Figure 2 i.e. multiple of 25. The classification accuracy is continuously increased up to 200 features (statistics) in both cases. Thereafter, there is no improvement of classification accuracy. The accuracy reached up to 87.9% when we considered two hundred different statistics using MRMR approach. At the same time, the PCA approach gives only 78.25% of classification accuracy. Base on experiment, only 37.17% features (statistics) are enough to achieve this amount of classification accuracy. It means that 62.83% features (statistics) are reduced for genre classification. Ultimately, computational complexity of a classifier is also reduced.

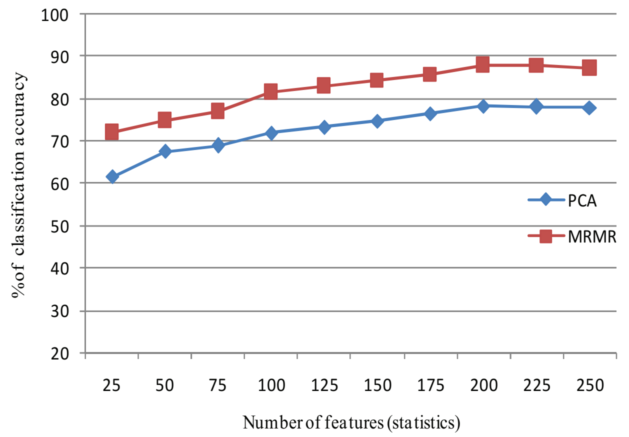

Figure 3: Number of features Vs classification accuracy including all features statistics

Table 3 shows important feature statistics that have higher discrimination power for genre classification. They are obtained from MRMR feature reduction algorithm. These statistics are taken only considering the lower order moments. There are only first seventy-five feature statistics out of one hundred eighteen into Table 3 to represent how many are crucial for genre classification. These feature statistics are selected base on score value of MRMR algorithm and they arrange into descending order. Similarly, one can find other important feature statistics from the combination of all the higher order statistics (lower, higher order moments, and covariance components).

To get a better picture of the classification accuracy of the individual music genre, the confusion matrix is given in Table 4 for GTZAN dataset. The confusion matrix is *n* x *n* matrix, at which each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The diagonal entries of the confusion matrix are the rates of music genre classification that are correctly classified, while the off-diagonal entries correspond to misclassification rates. The genre are arranged in the order of Blues (Bl), Classical (Cl), Country (Co), Disco (Di), Hiphop (Hi), Jazz (Ja), Metal (Me), Pop (Po), Reggae (Re), and Rock (Ro)respectively.

TABLE 3: IMPORTANT FEATURE STATISTICS OBTAINED FROM MRMR ALGORITHM FOR LOWER ORDER MOMENTS

No.	Selected Feature	No.	Selected Feature
1	Standard deviation of 3^{rd} DMFCC component	39	Standard deviation of spectral flux
2	Mean of flatness	40	Standard deviation of 13^{th} MFCC component
3	Mean of spread	41	Standard deviation of 8^{th} DDMFCC component
4	Mean of key mode	42	Standard deviation of 12^{th} MFCC component
5	Mean of 2^{nd} DMFCC component	43	Standard deviation of 7^{th} DDMFCC component
6	Standard deviation of spectral energy	44	Mean of chromagram centroid
7	Standard deviation of 5^{th} DDMFCC component	45	Standard deviation of RMS energy
8	Mean of rolloff95	46	Standard deviation of 4^{th} MFCC component
9	Standard deviation of 1^{st} DDMFCC component	47	Standard deviation of 8^{th} MFCC component
10	Standard deviation of 4^{th} DMFCC component	48	Mean of zero crossing
11	Mean of spectral flux	49	Standard deviation of harmonic change detection function
12	Standard deviation of flatness	50	Standard deviation of 9^{th} MFCC component
13	Standard deviation of 3^{rd} DDMFCC component	51	Standard deviation of 8^{th} DMFCC component
14	Standard deviation of spread	52	Mean of roughness
15	Standard deviation of 2^{nd} DDMFCC component	53	Standard deviation of key clarity
16	Standard deviation of attack	54	Standard deviation of 7^{th} MFCC component
17	Mean of spectral energy	55	Standard deviation of 1^{st} MFCC component
18	Standard deviation of 5^{th} DMFCC component	56	Standard deviation of 5^{th} MFCC component
19	Standard deviation of spectral centorid	57	Mean of 9^{th} MFCC component
20	Standard deviation of 1^{st} DMFCC component	58	Standard deviation of key mode
21	Standard deviation of 4^{th} DDMFCC component	59	Mean of RMS energy
22	Mean of harmonic change detection function	60	Standard deviation of 2^{nd} MFCC component
23	Mean of 1^{st} MFCC component	61	Mean of key clarity
24	Mean of chromagram peak	62	Standard deviation of 9^{th} DDMFCC component
25	Standard deviation of 6^{th} DDMFCC component	63	Standard deviation of roughness
26	Standard deviation of tempo	64	Mean of 3^{rd} MFCC component
27	Standard deviation of spread	65	Mean of 6^{th} MFCC component
28	Mean of brightness	66	Standard deviation of 9^{th} DMFCC component
29	Standard deviation of 3^{rd} MFCC component	67	Mean of 4^{th} MFCC component
30	Standard deviation of 10^{th} MFCC component	68	Mean of 2^{nd} MFCC component
31	Standard deviation of zero crossing	69	Standard deviation of 6^{th} MFCC component
32	Standard deviation of 6^{th} DMFCC component	70	Standard deviation of chromagram centroid
33	Standard deviation of rolloff95	71	Mean of slope
34	Mean of attack	72	Standard deviation of 10^{th} DDMFCC component
35	Mean of spectral centorid	73	Mean of 8^{th} MFCC component
36	Standard deviation of brightness	74	Standard deviation of 11^{th} DDMFCC component
37	Standard deviation of 11^{th} MFCC component	75	Variance of 13^{th} DMFCC component
38	Standard deviation of 7^{th} DMFCC component		

TABLE 4: CONFUSION MATRIX OF GTZAN DATASET CLASSIFICATION ACCURACY USING SUPPORT VECTOR MACHINE

	Bl	Cl	Co	Di	Hi	Ja	Me	Po	Re	Ro
Bl	90.0	3.0	2.5	0.0	0.0	2.5	2.0	0.0	0.0	0.0
Cl	0.0	98.5	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0
Co	3.0	0.0	85.0	0.0	0.0	0.0	0.0	5.0	0.0	7.0
Di	0.0	0.0	0.0	89.5	0.0	0.0	0.0	6.5	4.0	0.0
Hi	0.0	0.0	0.0	2.0	97.0	0.0	0.0	1.0	0.0	0.0
Ja	1.0	0.0	0.0	0.0	0.0	92.0	0.0	0.0	3.0	4.0
Me	0.0	0.0	0.0	0.0	3.5	0.0	88.5	0.0	0.0	8.0
Po	0.0	0.0	0.0	8.5	0.0	0.0	0.0	86.5	3.0	2.0
Re	0.0	0.0	3.0	4.0	6.0	0.0	0.0	5.0	82.0	0.0
Ro	10.0	0.0	5.0	4.0	0.0	2.0	3.0	5.0	0.0	71.0

TABLE 5: COMPARISON OF CLASSIFICATION ACCURACY WITH OTHER APPROACH OF GTZAN DATASET

Reference	CA
Our approach	87.90%
Jin S. Seo [22]	84.09%
Bergstra et al [23]	82.50%
Li et al.[24]	78.50%
Lidy et al. [25]	76.80%
Benetos et al [26]	75.00%
Tzanetakis [2]	61.00%

From the confusion matrix, one can notice that some music genres are classified with significant accuracies like Blues, Classical, Disco, Hiphop, Jazz, and Metal. Except for Rock, other music genres also show the competitive classification performance. Rock music has a minimum accuracy of classification rate. It confuses with Blues, Pop, and Country. In our conjecture Rock music is diverse in nature as compared to other genres so that it might share the characteristic with other genres. Similarly, Pop music mainly confused with Disco and Reggae. It also somehow confused with Rock music respectively. Table 5 shows the classification accuracy of our approach compared with other reported ones for the same data set.

VI. CONCLUSION

In this paper, we adopted four groups of diverse audio features: dynamics, rhythm, spectral, and harmony for the overall impact in genre classification. In the next stage, these features are integrated using low (mean and standard deviation) and high order moments (skewness and kurtosis) and also considered the covariance matrix. The experiment has been performed into two stages: at first considering low order moments, and then including high order moment and covariance components. Feature dimension increases sharply. Therefore, they are controlled by two feature reduction methods i.e. MRMR and PCA. Almost 63% features from whole set of features are selected to be insignificant for genre classification. From the selected features, the overall classification accuracy is 87.9% and 78.2% using support vector machine for MRMR and PCA feature reduction method, respectively. The MRMR based feature reduction method using SVM outperforms other contemporary methodologies. Moreover, MRMR algorithm finds the set of maximum relevance features among the whole set of features shown in Table 3. From the table, one can clearly visualize which feature and their corresponding statistics put the significant impact for genre classification with MRMR. However, it is seen that PCA doesn't keep the original features as it is. It transforms the original features that makes it impossible to visualize the original ones.

ACKNOWLEDGMENT

This work was partially supported by a National Research Foundation of Korea grant funded by the Korean government (2011-0022152) and BK21PLUS.

REFERENCES

[1] Xu, N. C. Maddage, and X. Shao, "Automatic music classification and summarization," *IEEE Trans.Speech Aud. Processing*, vol. 13,no. 3, pp. 441–450, May 2005.

[2] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process*, vol. 10, no. 3, pp. 293–302, Jul. 2002

[3] Babu Kaji Baniya, Deepak Ghimire, and Joonwhoan Lee, "A Novel Approach of Automatic Music Genre Classification Based on Timbral Texture and Rhythmic Content Features," *Int. Conference on Advance Communication Technology (ICACT), pp.*96-102, 2014

[4] B. Logan. Mel frequency cepstral coefficients for music modeling. *In Proc. Int. Symposium on Music Information Retrieval ISMIR,* 2000

[5] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle , P. Cano, An Experimental Comparison of Audio Tempo Induction Algorithms, *IEEE Transactions on Speech and Audio Processing*, vol.14, page 1832-44, 2006.

[6] E. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1), 1998.

[7] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney. Classification of audio signals using statistical features on time and wavelet transform domains. In Proc. *Int. Conf. Acoustic, Speech, and Signal Processing* (ICASSP-98), volume 6, pages 3621–3624, 1998.

[8] H. Soltau, T. Schultz, and M. Westphal. Recognition of music types. In Proceedings of the 1998 IEEE *International Conference on Acoustics, Speech and Signal Processing*, 1998

[9] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 525-536, 1998.

[10] H. Deshpande, R. Singh, and U. Nam. Classification of music signals in the visual domain. In Proceedings of the COST-G6 *Conference on Digital Audio Effects*, 2001.

[11] H. Peng, F. Long, and C. Ding, "Feature selection based on mutualinformation: Criteria of max-dependency, max-relevance and min-redundancey," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 27, no. 8, pp.1226–1238, Aug. 2005.

[12] L. Smith, A Tutorial on Principal Components Analysis, www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf, 2002

[13] S. H. Srinivasan and M. Kankanhalli, "Harmonicity and dynamicsbased features for audio," in *Proc. ICASSP*, 2004, pp. 321–324

[14] Y. Zhang and J. Zhou, "Audio segmentation based on multi-scale audio classification," in *IEEE Proc. ICASSP*, May 2004, pp. 349–352

[15] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): A toolbox for musical feature extraction from audio," in *Proc. Int. Conf. Music Inf. Retrieval*, 2007, pp. 127–130 [Online]. Available: http://users.jyu.fi/lartillo/mirtoolbox/

[16] Babu Kaji Baniya, Deepak Ghimire, and Joonwhoan Lee, "Evaluation of different audio features for musical genre classification" *in proc. IEEE workshop on Signal Processing Systems*, Taipei, Taiwan, 2013

[17] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multi-feature speech/music discriminator," *in Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, 1997.

[18] R. A. Groeneveld and G. Meeden, "Measuring Skewness and Kurtosis," *Journal of the Royal Statistical Society*, Series D (The Statistician), 33, 391-399, 1984

[19] H. Peng and F. Long, "An efficient max-dependency algorithm for gene selection," in 36th Symp. Interface: Computational Biology and Bioinformatics, May 2004

[20] Marasys, "Data sets" http://marsysas.info/download/data

[21] Corrina Cortes and V Vapnik, "Support Vector Networks," *Journal of Machine Learning*, 1995

[22] Jin S. Seo, Seungjae Lee, Higher-order moments for musical genre classification, *Signal Processing*, vol. 91, Issue 8, pp. 2154-57, 2011

[23] Bergstra, J., Casagrande, N., Erhan, D., Eck, D. and Kegl B. "Aggregate features and AdaBoost for music classification", *Machine Learning*, Vol. 65, No. 2-3, pp. 473-484, 2006

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

[25] T. Lidy, A. Rauber, A. Pertusa, and J. Inesta, "Combining audio and symbolic descriptors for music classification from audio", *Music Information Retrieval Information Exchange* (MIREX), 2007

[26] E. Benetos, and Kotropoulos C. "A tensor-based approach for automatic music genre classification", *Proceedings of the European Signal Processing Conference, Lausanne*, Switzerland, 2008.