A Comparative Study on Content-Based Music Genre Classfication

Tao Li, Mitsunori Ogihara, and Qi Li,

Proceedings of the 26th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR 2003)

Speaker: Ming Jing Tsai

Introduction

- Growth of on-line music information
- Genre structures is specified by experts
  - Time consuming
  - expensive
- Audio signals of music belonging to same genre share certain characteristics.
- Automatic music genre classification
  - Feature extraction
  - Multi-class classification

Feature extraction

- Extract from the music signals information representing the music.
  - Timbral Textural Feature
  - Rhythmic Content Features
  - Pitch Content Features

Timbral Textural Feature

- To differentiate mixture of sounds with the same or similar rhythmic and pitch contents.
- Sound signals are first divided into frames that are statistically stationary, usually by applying a window function at fixed intervals.
- Timbral textural feature are then computed for each frame and the statistical values (mean, variance) of those features are calculated.

Timbral Textural Feature cont.

- Mel-Frequency Cesptral Coefficients(MFCCs)
- Spectral Centroid
  - Measure of brightness
- Spectral Rolloff
  - Measure the shape
- Spectral Flux
  - Measure the amount of local change
- Zero Crossings
  - Measure noisiness
- Low Energy
  - Measure amplitude distribution of the signal

Rhythmic Content Features

- Characterize the movements of music signals over time (rhythmic, beat, tempo)
- Detecting the most salient periodicities of the signal and it is usually extracted from beat histogram.

Pitch Content Features

- The melody and harmony information about music signals
- The dominant peaks of the autocorrelation function are accumulated into pitch histograms and Pitch Content Features are extracted from the pitch histograms

- Timbral Textural Feature are calculated for every short-time frame of sound while rhythmic and pitch content features are computed over the whole file.
- Timbral Textural Feature capture the statistics of local information of music signals from a global perspective, but not enough in representing the global information.
- Rhythmic and pitch content features don’t seem to capture enough information content for classification purpose

DWCHS

- Based on wavelet histogram
  - local and global information of music signals
- Sound file is a kind of waveform in time-domain and can be considered as a two dimensional entity of the amplitude over time.
- The distinguishing characteristics are contained in the amplitude variation, and in consequence, identifying the amplitude variation would be essential for music categorization.

DWCHS

- The raw signal in time domain is not good representation particularly for content-based categorization since the most distinguished characteristics are hidden in frequency domain.
- Decomposition of audio signal using wavelets produces a set of subband signals at different frequencies corresponding different characteristics.
- The wavelet coefficients are distributed in various frequency bands at different resolutions.

DWCHS cont.

- Daubechies wavelet filter Db8 with seven levels of decomposition.
- Construct the histogram of the wavelet coefficients at each subband.
- Compute the first three moments of all histograms
- Compute the subband energy for each subband

- Each music file in datasets is 30s signal
- DWCHs feature can be extracted on a small slice of an input music signal based on an intuition of “self-similarity”

Wavelet Basics

- Wavelet coefficients histogram is the histogram of the wavelet coefficients obtained by convolving a wavelet filter with an input music signal
- Compact support
  - Localization of wavelet
- Vanishing moment
  - Focusing on most important information
- Decorrelated coefficients
  - Reduce temporal correlation

Wavelet Basics cont.

- The complex signal in the time domain can be reduced into a much simpler process in the wavelet domain
- By computing the histograms of wavelet coefficients, we then get a good estimation of the probability distribution over time
- The good probability estimation thus leads to a good feature representation.

Examples

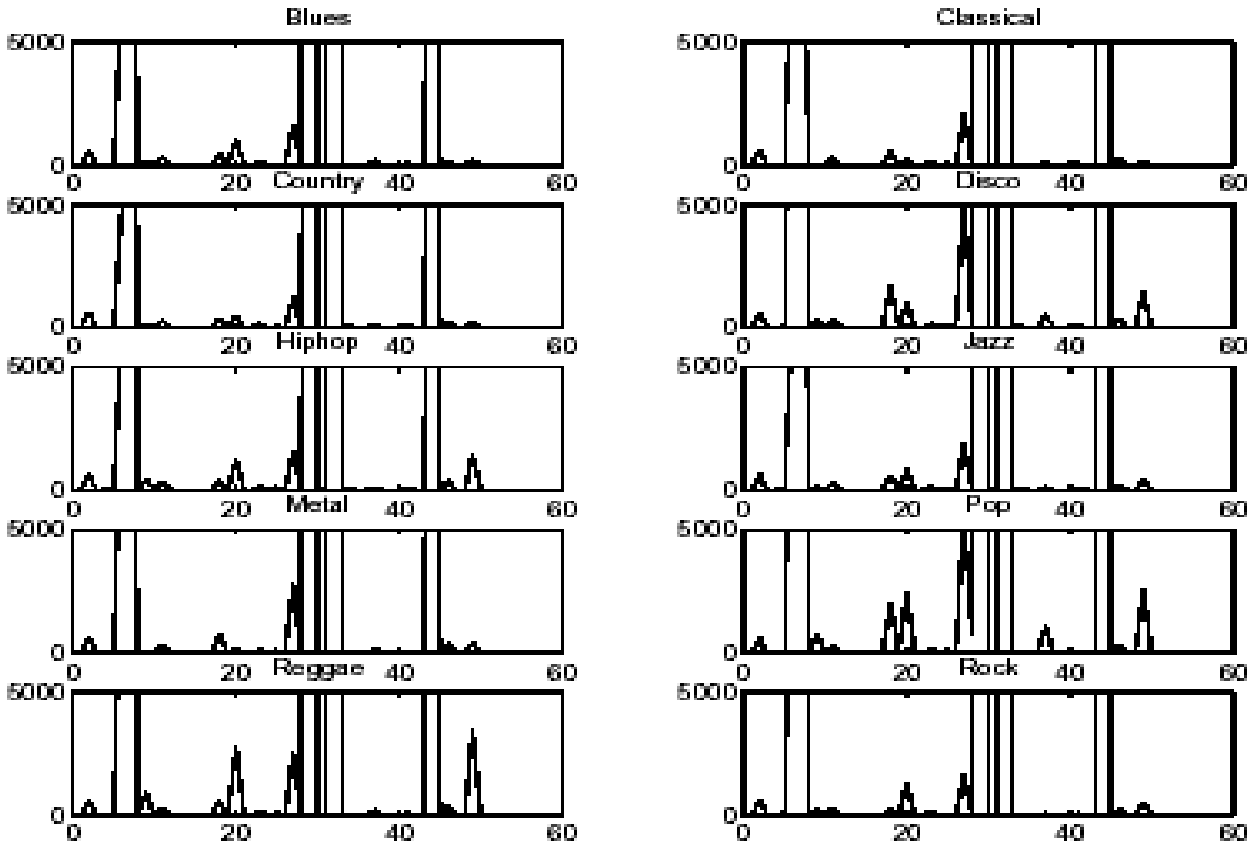

Figure 1: *DWCHs* of 10 music signals in different genre. The feature representations of different genre are mostly different to each other.

Examples cont.

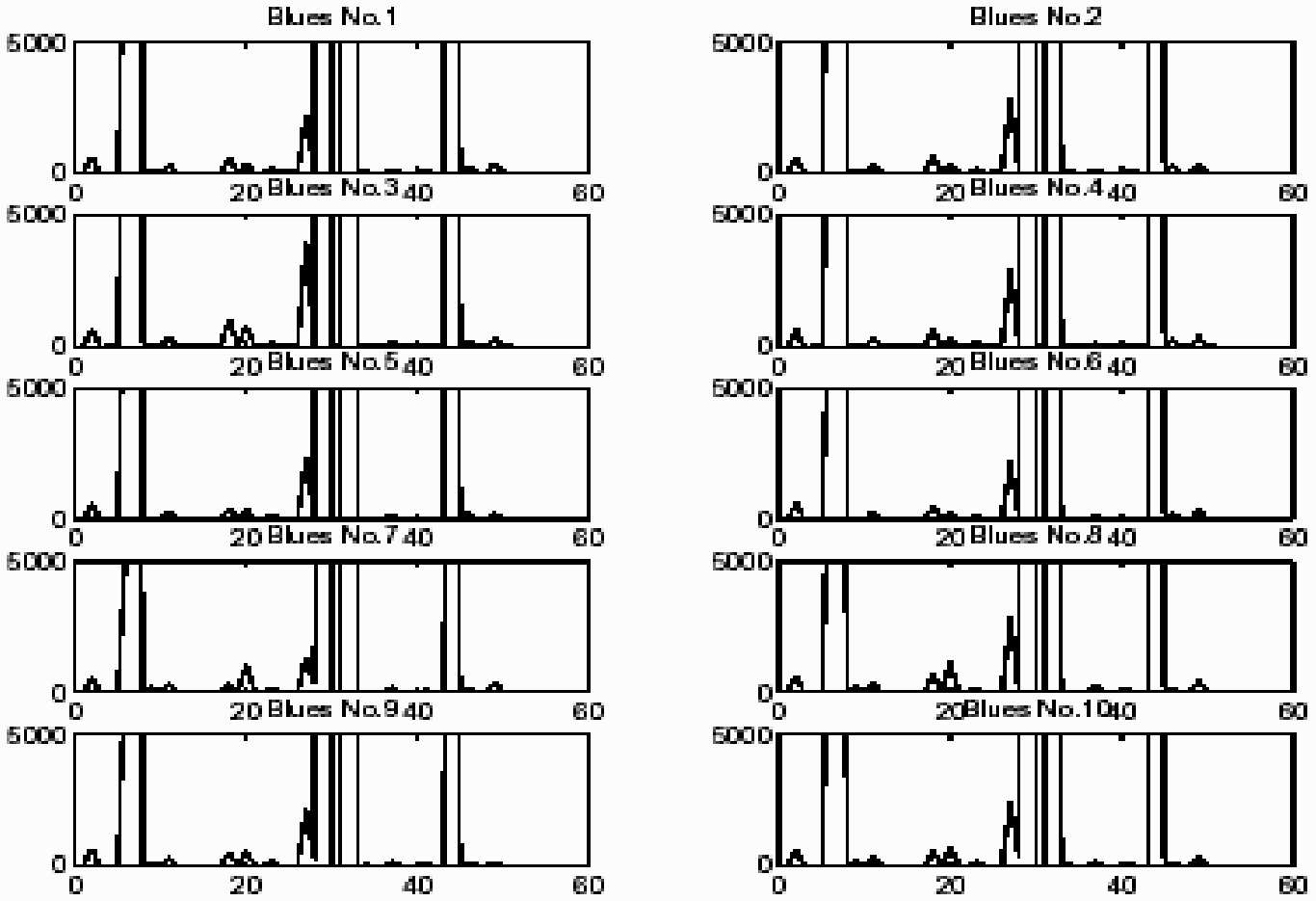

Figure 2: *DWCHs* of ten blues songs. The feature representations are similar.

Examples cont.

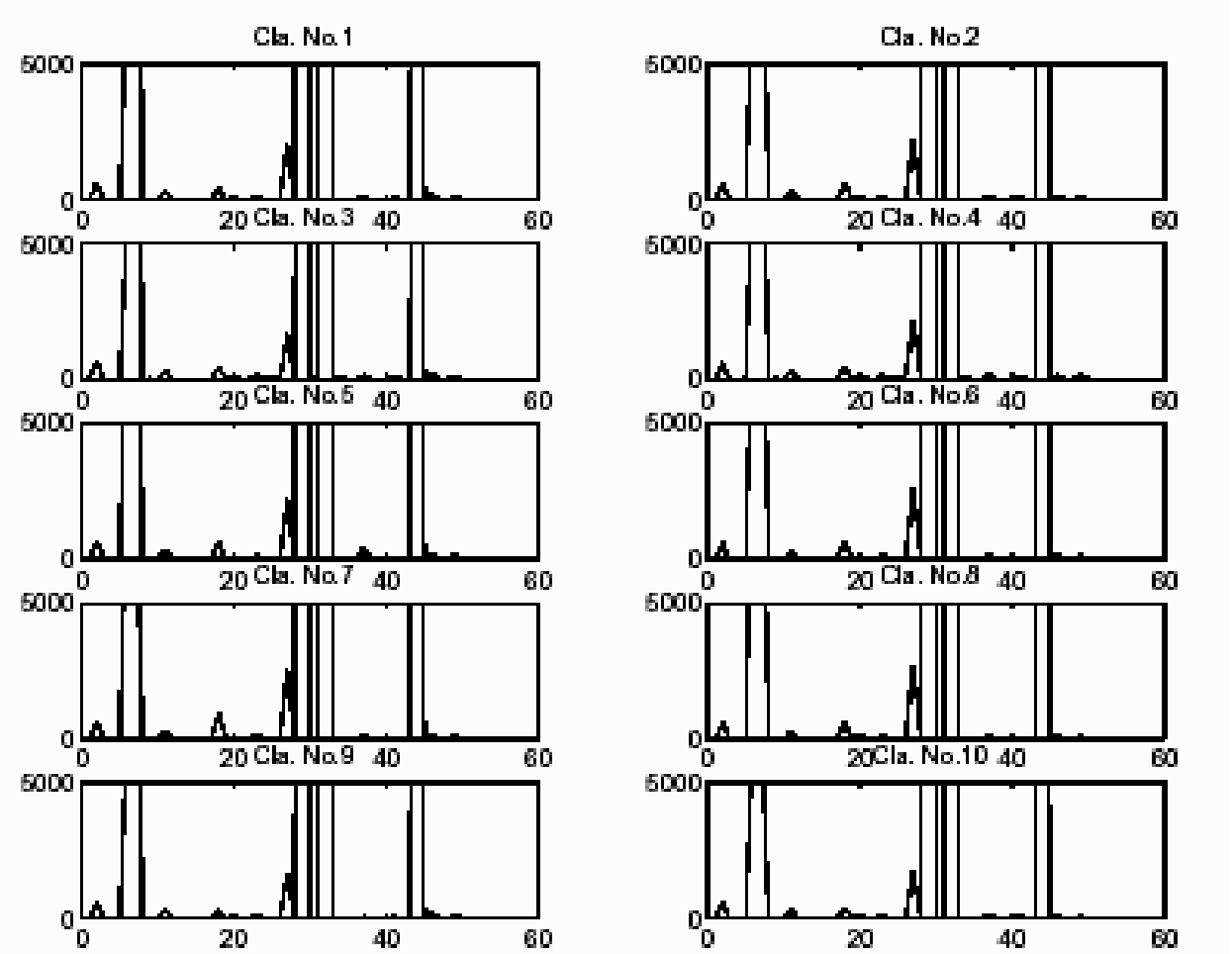

Figure 3: *DWCHs* of ten classical songs.

Multi-Class Learning methods

- Binary classification algorithms extended to handle multi-class.
- Decomposition of multi-class classification problems to a collection binary ones
  - One-versus-the-rest
    - To separate one class from the rest and then the multi-class classification is carried out according to the maximal output of the binary classifiers.
  - Pair-wise comparison
    - A classifier is trained for each possible pair of classes.
  - Multi-class objective functions
    - Modify the objective function of binary SVM in such a way that it simultaneously allows the computation of a multi-class classifier.

Multi-Class Learning methods cont.

- Support Vector Machines
  - Searching for a hyperplane that separates the positive data points and the negative data points with maximum margin.
- K-Nearest Neighbor (KNN)
  - Allow a small number of neighbors to influence the decision on a point.
- Gaussian Mixture Models (GMM)
- Linear Discriminant Analysis (LDA)
  - Find a linear transformation that best discriminates among classes and perform classification in the transformed space based on some metric such as Euclidean distances.

Experimental

- Two dataset
  - 1000 songs over ten genres (Blues, Classical, Country, Disco, Hip pop, Jazz, Metal, Pop, Reggae, and Rock).
  - 756 songs over five genres : Ambient, classical, Fusion, Jazz, and Rock
- MARSYAS for extracting the features
  - MFCCs, FFT, Beat and Pitch
- Classification:
  - SVM: pairwise, one-versus-the-rest, multi-class objective functions

Experiment

Features	Methods					
	SVM1	SVM2	MPSVM	GMM	LDA	KNN
DWCHs	74.9(4.97)	78.5(4.07)	68.3(4.34)	63.5(4.72)	71.3(6.10)	62.1(4.54)
Beat+FFT+MFCC+Pitch	70.8(5.39)	71.9(5.09)	66.2(5.23)	61.4(3.87)	69.4(6.93)	61.3(4.85)
Beat+FFT+MFCC	71.2(4.98)	72.1(4.68)	64.6(4.16)	60.8(3.25)	70.2(6.61)	62.3(4.03)
Beat+FFT+Pitch	65.1(4.27)	67.2(3.79)	56.0(4.67)	53.3(3.82)	61.1(6.53)	51.8(2.94)
Beat+MFCC+Pitch	64.3(4.24)	63.7(4.27)	57.8(3.82)	50.4(2.22)	61.7(5.23)	54.0(3.30)
FFT+MFCC+Pitch	70.9(6.22)	72.2(3.90)	64.9(5.06)	59.6(3.22)	69.9(6.76)	61.0(5.40)
Beat+FFT	61.7(5.12)	62.6(4.83)	50.8(5.16)	48.3(3.82)	56.0(6.73)	48.8(5.07)
Beat+MFCC	60.4(3.19)	60.2(4.84)	53.5(4.45)	47.7(2.24)	59.6(4.03)	50.5(4.53)
Beat+Pitch	42.7(5.37)	41.1(4.68)	35.6(4.27)	34.0(2.69)	36.9(4.38)	35.7(3.59)
FFT+MFCC	70.5(5.98)	71.8(4.83)	63.6(4.71)	59.1(3.20)	66.8(6.77)	61.2(7.12)
FFT+Pitch	64.0(5.16)	68.2(3.79)	55.1(5.82)	53.7(3.15)	60.0(6.68)	53.8(4.73)
MFCC+Pitch	60.6(4.54)	64.4(4.37)	53.3(2.95)	48.2(2.71)	59.4(4.50)	54.7(3.50)
Beat	26.5(3.30)	21.5(2.71)	22.1(3.04)	22.1(1.91)	24.9(2.99)	22.8(5.12)
FFT	61.2(6.74)	61.8(3.39)	50.6(5.76)	47.9(4.91)	56.5(6.90)	52.6(3.81)
MFCC	58.4(3.31)	58.1(4.72)	49.4(2.27)	46.4(3.09)	55.5(3.57)	53.7(4.11)
Pitch	36.6(2.95)	33.6(3.23)	29.9(3.76)	25.8(3.02)	30.7(2.79)	33.3(3.20)

Table 1: Classification accuracy of the learning methods tested on Dataset A using various combinations of features. The accuracy values are calculated via ten-fold cross validation. The numbers within parentheses are standard deviations. SVM1 and SVM2 respectively denote the pairwise SVM and the one-versus-the-rest SVM.

Experiment cont.

Number	Genre	Accuracy
1	Blues	95.49 (1.27)
2	Classical	98.89 (1.10)
3	Country	94.29 (2.49)
4	Disco	92.69 (2.54)
5	Jazz	97.90(0.99)
6	Metal	95.29 (2.18)
7	Pop	95.80 (1.69)
8	Hiphop	96.49 (1.28)
9	Reggae	92.30 (2.49)
10	Rock	91.29 (2.96)

Table 2: Genre specific accuracy of SVM1 on DWCHs. The results are calculated via ten fold cross validation and each entry in the table is in the form of accuracy(standard deviation).

Experiment cont.

Classes	Methods					
	SVM1	SVM2	MPSVM	GMM	LDA	KNN
1 & 2	98.00(3.50)	98.00(2.58)	99.00(2.11)	98.00(3.22)	99.00(2.11)	97.5(2.64)
1, 2 & 3	92.33(5.46)	92.67(4.92)	93.33(3.51)	91.33(3.91)	94.00(4.10)	87.00(5.54)
1 through 4	90.5(4.53)	90.00(4.25)	89.75(3.99)	85.25(5.20)	89.25(3.92)	83.75(5.92)
1 through 5	88.00(3.89)	86.80(4.54)	83.40(5.42)	81.2(4.92)	86.2(5.03)	78.00(5.89)
1 through 6	84.83(4.81)	86.67(5.27)	81.0(6.05)	73.83(5.78)	82.83(6.37)	73.5(6.01)
1 through 7	83.86(4.26)	84.43(3.53)	78.85(3.67)	74.29(6.90)	81.00(5.87)	73.29(5.88)
1 through 8	81.5(4.56)	83.00(3.64)	75.13(4.84)	72.38(6.22)	79.13(6.07)	69.38(5.47)
1 through 9	78.11(4.83)	79.78(2.76)	70.55(4.30)	68.22(7.26)	74.47(6.22)	65.56(4.66)

Table 3: Accuracy on various subsets of Dataset A using *DWCHs*. The class numbers correspond to those of Table 2. The accuracy values are calculated via ten-fold cross validation. The numbers in the parentheses are the standard deviations.

Experiment cont.

Features	Methods					
	SVM1	SVM2	MPSVM	GMM	LDA	KNN
DWCHs	71.48(6.84)	74.21(4.65)	67.16(5.60)	64.77(6.15)	65.74(6.03)	61.84(4.88)
Beat+FFT+MFCC+Pitch	68.65(3.90)	69.19(4.32)	65.21(3.63)	63.08(5.89)	66.00(5.57)	60.59(5.43)
FFT+MFCC	66.67(4.40)	70.63(4.13)	64.29(4.54)	61.24(6.29)	65.35(4.86)	60.78(4.30)
Beat	43.37(3.88)	44.52(4.14)	41.01(4.46)	37.95(5.21)	40.87(4.50)	41.27(2.96)
FFT	61.65(5.57)	62.19(5.26)	54.76(2.94)	50.80(4.89)	57.94(5.11)	57.42(5.64)
MFCC	60.45(5.12)	67.46(3.57)	57.42(4.67)	53.43(5.64)	59.26(4.77)	59.93(3.49)
Pitch	37.56(4.63)	39.37(3.88)	36.49(5.12)	29.62(5.89)	37.82(4.67)	38.89(5.04)

Table 4: Classification accuracy of the learning methods tested on Dataset B using various combinations of features calculated via via ten-fold cross validation. The numbers within parentheses are standard deviations.