Chinese Opera Genre Classification Based on Multi-feature Fusion and Extreme Learning Machine

JianRong Wang*, ChenLiang Wang†, JianGuo Wei‡ and JianWu Dang§
* Tianjin University, Tianjin, China
E-mail: wjr@tju.edu.cn
† Tianjin University, Tianjin, China
E-mail: wangchenliang007@foxmail.com Tel: +86-150-22016525
‡ Tianjin University, Tianjin, China
E-mail: jianguo.fr@gmail.com
§ Tianjin University, Tianjin, China
E-mail: jian_wu_dang@163.com

***Abstract*—Chinese traditional opera plays an important role in Chinese traditional culture, it reflects the customs and value tendency of different areas. Though researchers have already gained some achievements, studies on this field are scarce and the existing achievements still need to be improved. This paper proposes a system based on multi-feature fusion and extreme learning machine (ELM) to classify Chinese traditional opera genre. Inspired by music genre classification, each aria is split into multiple segments. 19 features are then extracted and fused to generate the fusion feature. Finally, we use ELM and majority voting methods to determine the genre of the whole aria. The research data are 800 arias of 8 typical genres collected from Internet. This system achieves a mean classification accuracy of 92% among 8 famous Chinese traditional opera genres. The experimental results demonstrated that multi-feature fusion improves classification accuracy of Chinese traditional opera genres. Feature fusion is more effective than decision fusion in dealing with this problem.**

I. Introduction

Chinese traditional operas originate from folk songs and dances. It is a comprehensive art form consisting of music, dance, fine arts, Chinese martial arts, acrobatics, etc. Almost all kinds of operas are based on the same Chinese history, literature and popular legend. However, differences of geographic location, language and customs cause variety of vocal and pronunciation systems. According to the statistics, China has more than 360 opera genres and tens of thousands traditional theatrical pieces [1]. However, it has caused a huge culture barrier between Chinese traditional operas and modern audience. Some operas are even on the verge of disappearing. Meanwhile, there exists little research on Chinese traditional opera using modern science and technology [2], [3], [4]. Therefore, it is important to promote the development of Chinese Opera in particular by computer technique.

Modern music has been studied for years [5], [6]. Jiang et al. constructed a feature based on spectrum contrast which was suitable for music classification [6]. It considered the distribution of spectrum peaks and valleys in each sub-band and can distinguish harmonic and noise components better. James Bergstra et al., used ADABOOST to select from a set of audio features and then aggregated. They demonstrated that the technique of classifying features aggregated over segments of audio is better than classifying the whole song or individual short-timescale features [7]. Chang-Hsing Lee et al., used modulation analysis to capture the time-varying or rhythmic information of music signals [8]. Yin-Fu Huang et al., used a self-adaptive harmony search algorithm to select the most relative features with lower dimension, their methods improved the classification accuracy effectively [9]. However, there was little research on Chinese traditional operas. In recent years, many operas, such as Peking Opera, have been listed as China's intangible cultural heritage [10]. Chinese traditional operas aroused great attention among researchers again. YI-BIN ZHANG et al., made a study of classification and similarity analysis among 8 typical genres of Chinese traditional operas and their classification accuracy reached 82.4% [4]. Ziqiang Zhang et al., combined the theory of Peking opera with information technology to analyze opera structure, the result fit human's usual practice appropriately [2]. Sundberg et al., studied the acoustic characteristics of different kinds of opera performers (Sheng, Dan, Jing, Chou) in classical opera singing [3]. They found that the voice timbres in these roles differ dramatically from those in Western operas.

This paper aims to build a system to classify 8 typical Chinese traditional opera genres (Jin opera, Peking opera, Qin opera, Henan opera, Shao opera, Cantonese opera, Zhuizi, Kunqu). As presented above, segment level features are better than song level or short-time scale features, we split each aria into multiple segments. 19 features are then extracted to form the original data set. Since single feature is not enough to represent an aria, we adopt decision fusion and feature fusion respectively [11]. Considering the high dimension of fusion feature, ELM is selected as the classifier for its fast speed

[12]. Finally, majority voting methods are used to calculate the genre of each aria. The experimental results demenstrated that multiple-feature fusion improves the classification accuracy in Chinese opera genre classification. Our system achieves a mean accuracy of 92% through feature fusion, which is more effective than that using decision fusion.

The rest of this paper is structured as follows. Section II introduces the system overview. Section III describes feature extraction in details. Section IV deals with the process of information fusion. Section V compares ELM with three other classification algorithms. Section VI gives the experimental results. Section VII discusses the experimental results. Finally, Section VIII is the conclusion.

II. System overview

As presented in Fig. 1, the system architechture consists of five components: aria partition, feature extraction, feature fusion, classification and majority voting. The details are as follows.

A. Aria partition

Each aria is split into multiple segments evenly with half overlapping and each segment corresponds to a texture window [5]. Predicting the genre of an aria is then transformed from the whole aria to each segment.

B. Feature extraction

In order to recognize the genre of a segment, multiple features are extracted from audio segment to analyze the signals. Analysis window and texture window are proposed to describe audio signals [5]. For analysis window, audio signals are split into many short frames (20∼30ms) and each frame is processed respectively. Since the characteristics of audio signals in analysis window are relatively stable, it is necessary to extract short-time feature in analysis window. Texture window captures the variation of multiple consecutive analysis windows. The actual texture window features used in this system are the means and variances of the extracted features over a number of analysis windows. As presented in Table I, this paper extracts 19 texture window features to capture the time-varying information, such as octave-based spectral contrast (OSC) [6], normalized audio spectral envelope (NASE) [13], [14], spectral pitch chroma [15], Mel frequency cepstrum coefficients (MFCCs) [16], etc.

C. Feature fusion

Considering the complementary information and the relationship among different features, 19 kinds of texture window features are normalized firstly and then combined together to form the fusion feature.

D. Classification

The dimension of the fusion feature vector is high. Dimension reduction algorithms usually lost useful information. We select ELM as the classifier of this system because of its extremely fast speed. [12].

TABLE I
Texture Window Feature Set

No.	Descriptor	Dimession	Type
1	Time zero crossing rate	2	Time domain features
2	Time autocorrelation coefficent	2	
3	Time max autocorrelation	2	
4	Time peak envelope	4	
5	Time predictivity ratio	2	
6	Time standard deviation	2	
7	OSC	32	Frequency domain features
8	MFCCs	52	
9	NASE	38	
10	OMSC	32	
11	MSFM-MSCM	16	
12	Spectral pitch chroma	24	
13	Spectral crest	2	
14	Spectral slope	2	
15	Spectral decrease	2	
16	Spectral flatness	2	
17	Spectral centroid	2	
18	Spectrul rolloff	2	
19	Spectral flux	2	

E. Majority voting

As mentioned above, an aria is split into multiple segments and ELM only predicts the genre of each segment. Therefore, majority voting is applied to determine the genre of the whole aria.

III. Feature Extraction

The Chinese opera singing is so flexible that single feature is difficult to describe an opera effectively. In order to represent audio signals more comprehensively from different perspectives, this paper extracts multiple features and combines with information fusion technique. The original feature sets are shown in Table I.

In general, audio signals are studied in two ways: time domain and frequency domain. In time domain, the strength of sampling signals varies via time axis. In frequency domain, each amplitude sample is transformed from time domain to frequency domain by discrete Fourier transform (DFT).

A. Time Domain Features

In this paper, 6 time domain features are extracted.

1) Zero crossing rate: Zero crossing rate describes the ratio of sign-changes in audio signals.

2) Time autocorrelation: Time autocorrelation coefficient tries to find the fundamental frequency and the period of audio signal.

3) Time max autocorrelation: Max autocorrelation describes the max value of audio signals' autocorrelation function.

4) Time peak envelope: Peak envelope estimates 2 kinds of envelope peak value.

5) Time predictivity ratio: Predictivity ratio is obtained by calculating the linear prediction coefficients.

6) Time standard deviation: Standard deviation of audio signals.

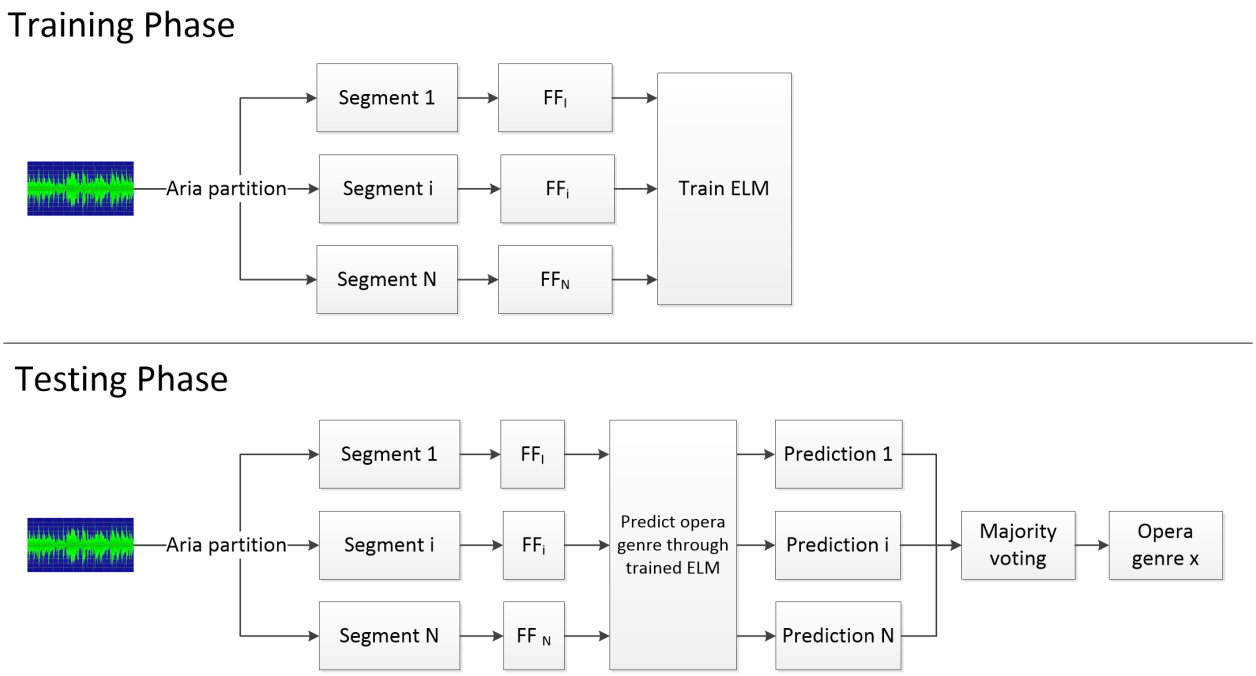

Fig. 1. System overview. Segment i corresponds with the ith texture window of an aria. FF_i is the fusion feature of the ith Segment. Prediction i indicates which genre the ith segment belongs to.

B. Frequency Domain Features

Comparing with time domain, frequency feature is more suitable to human auditory mechanism. Researchers commonly analyse the frequency content through transforming signals from time domain to frequency domain. In this paper, 13 frequency features are selected to describe the frequency characteristics.

1) Octave-based Spectral Contrast: Generally, the strong spectrum peaks roughly correspond with harmonic components, while noise often appears at valleys. Based on this point, Jiang et al., proposed a feature (OSC) to describe the relative distribution of spectrum peaks and valleys in each sub-band [6].

2) Mel frequency cepstrum coefficients: As is shown in Fig. 2, there are some differences between MFCCs and OSC. MFCCs use Mel-scale filters and emphasize on the average spectrum distribution instead of peaks and valleys in each sub-band [6]. Therefore, MFCCs and OSC are complementary in theory.

3) Normalized Audio Spectral Envelope: NASE [13], [14] is proposed by MPEG-7 audio group for obtaining a low-complex description of audio content. It is derived by normalizing each audio spectral envelope coefficient and mainly used for audio classification.

4) Modulation Spectral measures: Modulation spectrum describes the time-varying and rhythm information of music signals [17], [18]. This paper selects four modulation features: Octave-based modulation spectral contrast (OMSC), modulation spectral flatness measure (MSFM) and modulation spectral crest measure (MSCM). They are derived from modulation

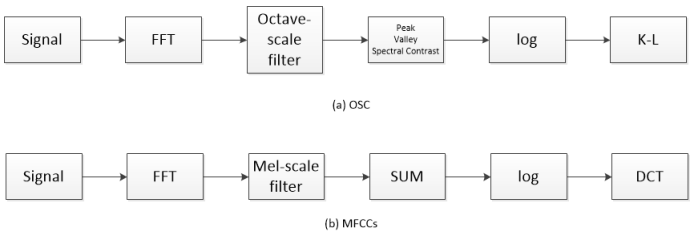

Fig. 2. Comparison of extraction process of OSC and MFCCs. For each sub-band, OSC considers spectrum peaks and valleys, while MFCCs emphasizes on spectrum average distribution.

analysis of OSC, spectral flatness measure (SFM) and spectral crest measure (SCM). Therefore, OMSC, MSFM and MSCM capture the variation of OSC, SFM and SCM respectively [18], [19].

5) Spectral Pitch Chroma: The key or scale information of music segment can provide important clues to music information retrieval [20], and spectral pitch chroma is just a measure describing key information of an aria.

6) Other Spectral measures: Spectral crest computes the ratio of max energy and root mean square of energy, it is usually used to distinguish noise and harmonic sound. Spectral slope computes the amount of decreasing of the spectrum amplitude through linear regression. Similar to spectral slope, spectral decrease is another measure describing the decreasing rate of spectral amplitude. Spectral flatness reflects the flatness properties of the power spectrum through dividing the geometric mean by the arithmetic mean of the power spectrum. Spectrum centroid calculates the ”center of gravity”,

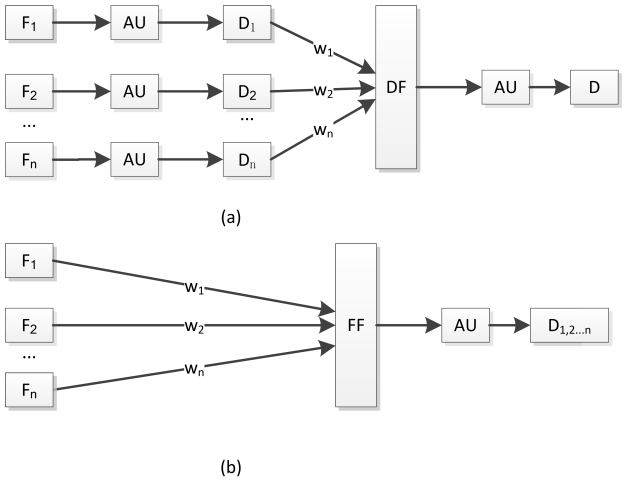

Fig. 3. Information fusion process. F_i is the ith feature, AU is short for analysis unit, FF represents fusion feature, D_i is the ith decision from the classifier and w_i is the weight of the ith information source. **(a)** decision fusion, **(b)** feature fusion.

it represents the "brightness" of sound intuitively. Spectral roll-off refers to a frequency threshold that 95% of signal energy is below this point, it roughly represents the cutting frequency between noise and harmony [21]. Finally, spectral flux estimates the distance between two consecutive frames (analysis window) of audio signals.

19 features consider different aspects and give different classification results. Feature classification accuracy between 2 operas is shown in Table II.

IV. Information Fusion

When there is a variety of information sources, the most common approach is to fuse multiple information [11], [22], [23] through decision fusion and feature fusion (Fig. 3).

A. Decision Fusion

In decision fusion, a decision is the output of an analysis unit. In this paper, a decision is defined as each classifier's output, i.e., the probability vector $P_s = [p_1, \ldots, p_i, \ldots, p_N]$, where i represents a specific opera genre index and P_s corresponds with the sth decsion [24]. Linear weighted fusion methods are used to combine different decisions with different weights by sum operators (Fig. 3 (a)) as follows:

$$P = \sum_{i=1}^{N} w_i \times P_i \tag{1}$$

where P_i is the probability vector obtained from the ith decision and w_i is the corresponding weight. Therefore, the testing sample x will be classified as class C_i if p_i is the maximum in P.

However, how to assign reasonable weight to each decision is the key point of linear weighted fusion. In this paper, harmonic search algorithm, simulating the process of playing music, is used to search an optimum weights vector heuristically. The HS algorithm mimics the behaviors of music players

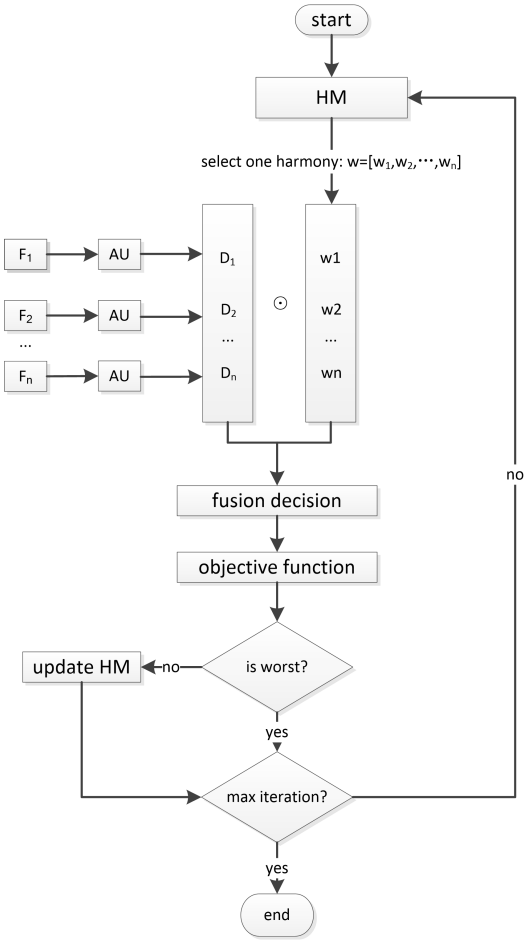

Fig. 4. Procedure of decision fusion with harmony search. HM represents the harmony memory matrix. D_i is the ith decision from the ith feature F_i with the corresponding weight w_i. Fusion decision is derived by linear weighted fusion methods as in (1).

in an improvisation process, where each player generates a pitch with each instrument based on three operations: random selection, memory consideration, and pitch adjustment [25]. The general procedures of an HS are as follows.

- Step 1. Create and randomly initialize an HMS-size harmony memory (HM).
- Step 2. Improvise a new harmony from the HM.
- Step 3. Update the HM. If the new harmony is better than the worst harmony in the HM, include the new harmony in the HM, and exclude the worst harmony from the HM.
- Step 4. Repeat Steps 2 and 3 until the maximum number of iterations is reached.

Therefore, the procedure of weight adjusting combining harmony search in decision fusion is shown in Fig. 4. The raw data set is divided into three parts (training set, validation set and test set) with the proportion of 8:1:1. The error rate of classification is used as the objective function, other parameters of HS are shown in Table III.

As is shown in Fig. 5, some weak features with low dimension usually give bad determination alone, decision fusion with 19 features only gains an average accuracy of

TABLE II
FEATURE CLASSIFICATION ACCURACY BETWEEN 2 OPERAS

Operas \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Jin&Jing	45.5	66.0	73.0	56.5	56.0	58.0	83.5	79.0	89.0	85.5	77.5	89.0	49.5	58.5	50.5	53.0	50.0	50.0	52.0
Jin&Qin	56.0	58.5	63.0	44.5	56.0	49.0	89.5	66.5	89.5	82.0	79.5	85.5	50.5	57.5	45.5	47.5	50.0	50.0	40.0
Jin&Yu	46.5	52.0	83.0	62.0	61.0	60.5	93.0	79.5	89.0	77.0	81.0	95.5	57.5	65.5	49.5	61.5	50.0	44.0	61.0
Jin&Shao	53.0	71.0	78.0	57.0	58.5	57.5	92.0	75.5	91.0	88.0	76.0	87.0	63.0	53.5	50.0	49.0	50.0	50.0	51.5
Jin&Cantonese	54.0	63.0	71.0	50.5	66.0	55.0	85.0	82.0	94.0	91.0	90.5	93.5	50.0	54.0	49.5	57.5	50.0	87.0	62.0
Jin&Zhui	50.5	86.0	70.0	60.5	74.5	64.0	97.5	86.5	96.5	94.0	94.5	94.5	61.5	65.0	49.5	56.5	50.0	72.5	47.0
Jin&Kun	49.5	90.0	70.0	89.5	76.5	86.0	98.0	88.0	99.0	94.0	94.0	95.0	69.5	84.5	51.0	77.5	49.5	88.5	82.5
Jing&Qin	53.5	67.5	68.5	55.0	57.5	57.5	93.5	74.5	94.0	90.0	91.0	85.5	61.0	56.0	49.5	57.5	50.0	50.0	52.5
Jing&Yu	51.0	57.0	61.5	64.5	61.5	66.0	83.5	71.5	85.0	78.0	78.5	80.0	65.5	68.5	49.5	66.5	50.0	50.0	66.0
Jing&Shao	52.5	61.0	60.5	43.5	51.5	52.5	77.5	76.0	84.5	72.0	76.0	79.5	64.5	45.0	50.0	46.5	50.0	50.0	51.0
Jing&Cantonese	54.5	60.5	77.0	48.0	57.5	61.0	91.5	70.0	90.5	85.0	92.0	78.0	55.0	60.0	48.5	59.5	50.0	63.0	58.0
Jing&Zhui	50.5	76.0	63.5	59.5	63.0	56.0	89.0	85.5	92.5	86.5	85.5	83.5	62.5	55.5	45.5	58.0	50.0	50.0	59.0
Jing&Kun	56.5	72.5	88.0	83.5	68.5	79.0	96.5	89.0	93.5	87.0	92.0	87.0	75.0	80.0	49.5	76.5	50.0	50.0	81.0
Qin&Yu	42.0	56.5	75.0	61.0	55.0	60.0	83.0	73.0	80.5	77.5	88.5	90.0	59.5	59.5	49.5	56.5	50.0	47.0	59.0
Qin&Shao	50.5	69.0	77.5	57.0	61.0	52.0	90.5	76.0	88.5	84.5	85.5	88.0	67.0	52.0	50.0	53.5	50.0	45.0	51.0
Qin&Cantonese	53.5	66.5	61.0	52.0	67.5	60.5	90.0	78.5	91.5	80.5	86.0	90.5	56.0	57.5	51.0	62.5	50.0	50.0	62.5
Qin&Zhui	51.0	83.5	68.0	56.0	75.0	62.0	84.5	87.0	93.0	94.0	95.0	91.0	66.0	57.0	45.0	51.0	46.5	48.5	51.0
Qin&Kun	55.0	77.5	84.0	84.5	80.0	83.0	96.5	87.5	95.0	90.5	93.5	94.0	72.0	86.0	53.5	75.0	50.0	50.0	81.5
Yu&Shao	50.0	38.0	74.5	64.0	60.0	61.5	79.5	74.0	84.5	74.0	87.0	85.5	73.0	61.5	48.0	63.5	50.0	50.0	65.0
Yu&Cantonese	49.5	51.5	81.0	64.5	69.5	68.5	91.5	80.0	91.0	85.0	94.0	86.0	65.5	66.0	58.5	72.5	45.0	50.5	74.0
Yu&Zhui	52.5	81.5	77.0	63.5	71.0	62.5	93.0	83.5	94.0	90.0	87.0	86.0	73.0	64.0	49.0	60.0	50.0	58.0	62.5
Yu&Kun	53.0	83.0	94.5	91.5	84.5	91.5	98.5	91.5	95.5	94.5	96.5	92.5	84.0	92.5	50.0	86.0	50.0	52.5	93.5
Shao&Cantonese	48.5	74.5	87.0	52.5	60.0	57.5	88.0	78.5	90.5	81.0	91.0	82.5	68.0	46.5	48.5	61.0	50.0	50.0	57.0
Shao&Zhui	51.5	75.5	71.5	51.0	70.0	50.0	87.5	85.5	90.0	82.0	87.5	87.0	55.5	57.5	51.0	46.5	50.0	49.0	53.5
Shao&Kun	59.0	69.5	85.0	82.0	71.0	79.0	96.0	87.5	90.5	86.5	82.5	84.0	73.5	81.0	50.0	76.0	50.0	50.0	78.5
Cantonese&Zhui	58.0	74.5	79.0	51.5	71.0	65.0	94.5	88.0	94.5	94.0	95.0	86.0	56.5	61.0	50.0	65.5	50.0	50.0	65.5
Cantonese&Kun	55.5	75.5	89.0	86.0	51.0	84.5	94.0	87.0	91.0	88.5	93.0	88.0	68.5	88.0	61.5	72.5	50.0	50.0	81.5
Zhui&Kun	52.5	46.5	89.0	84.5	77.5	82.0	97.5	95.0	88.0	90.5	96.0	88.0	65.5	82.5	50.0	82.5	50.0	50.0	76.5

Note: number i represents the ith feature displayed in Table I.

86.3% (Table IV). Therefore, we only select OSC, MFCCs, NASE, OMSC, MSFM-MSCF and spectral pitch chroma as six information sources. Finally, the classification accuracy of Chinese opera genre reaches 89.3%.

B. Feature Fusion

Besides decision fusion, feature fusion is also performed (Fig. 3 (b)). Compared with decision fusion strategy, a huge advantage of feature fusion is that it considers the correlation among multiple features. Generally, there are two techniques of feature fusion: serial combination and parallel combination [26]. If m and n are the weights of two features α and β, then fusion feature will be transformed into $[m*\alpha; n*\beta]$ by serial combination. In parallel combination, two real features will be transformed into a complex feature and the absolute value of the complex is regraded as the final fusion feature. In most of the case, weights are adjusted manually or all the features are given the same weight [27]. Though some low dimension features give bad determination, they provide supplementary information. In this paper, 19 normalized texture window features are combined through serial fusion manner with equal weights. As is shown in Fig. 5, fusion feature achieves better classification accuracy than single feature apparently.

TABLE III
PARAMETERS OF HARMONIC SEARCH ALGORITHM

Parameters	Values
Harmony memory size	20
Bandwidth	0.1
Harmony memory considering rate	0.9
Pitch adjustment rate	0.1
Iteration times	1000

V. CLASSIFICATION

After extracting multiple texture window features and information fusion, ELM is trained to classify 8 opera genres. It is a supervised single-hidden layer feedforward neural networks (SLFNs) with random hidden neurons and random feature proposed by [12]. Most of the traditional neural networks adopt gradient-based learning algorithms and the parameters are tuned iteratively, therefore, their learning speed is in general far slower than required. However, ELM calculates the parameters without iteration. Therefore, an advantage of ELM lies in its extremely fast speed and high classification accuracy. Compared with traditional backpropagation (BP) neural network, there exists no local minimum, time consuming or other common problems in traditional BP algorithms. The dimension of fusion feature is relatively high, a fast and accurate classifier

TABLE IV
COMPARISION OF FEATURES USED IN DECISION FUSION

Features	Accuracy
19 features	86.3%
OSC, MFCCs, NASE, OMSC, MSFM-MSCF, spectral pitch chroma	89.1%

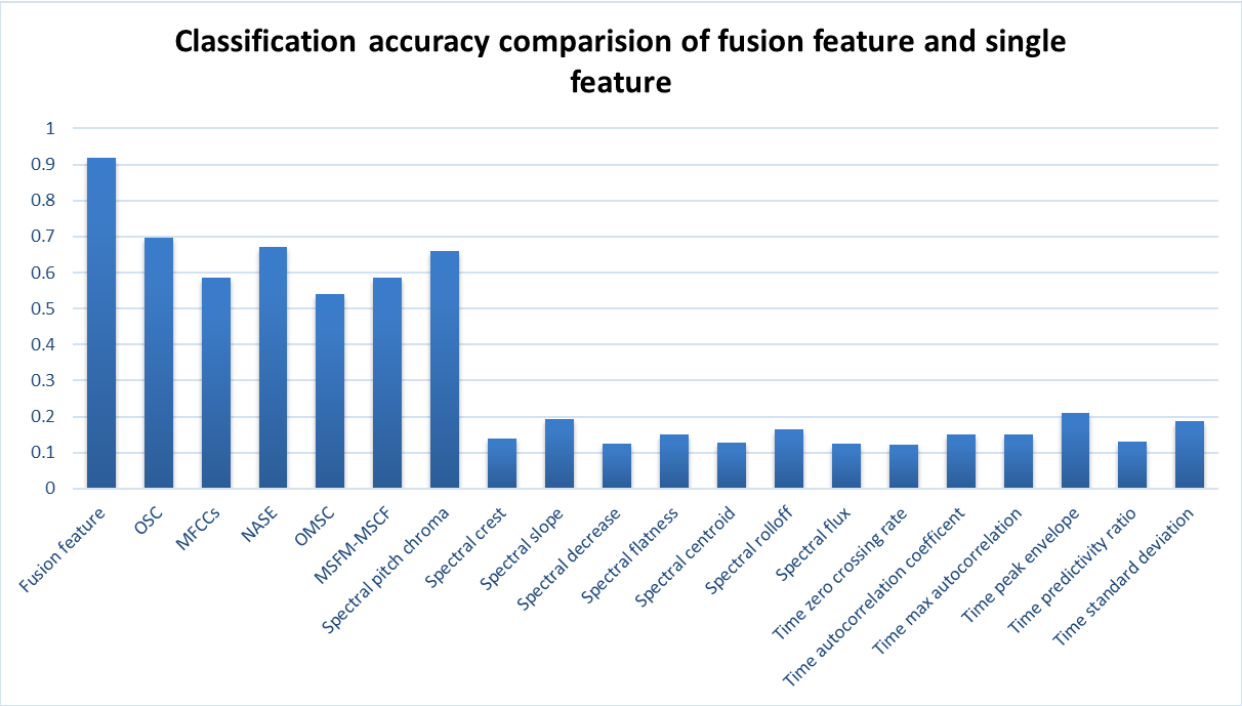

Fig. 5. Classification accuracy comparision of fusion feature and single feature.

is necessary, therefore, ELM is selected as the default classifier in our experiment.

For a testing sample x, the basic ELM only outputs a fuzzy vector $O = [o_1, o_2, \ldots, o_N], o_i \in [-1, 1]$, where o_i indicates the degree of x belonging to class i. The fuzzy vector is transformed into probability vector through the following formulas [24].

$$o_i = o_i - (-1) \tag{2}$$

$$p_i = \frac{o_i}{\sum_{i=1}^{N} o_i} \tag{3}$$

where p_i is the probability of x belonging to class i.

In order to demonstrate the efficiency and accuracy of ELM, three additional classifiers (SVM, random forest and sparse representation classification) are tested. SVM [28] with RBF kernel (gamma=0.25) achieves an accuracy of 91.2%, but its speed is far slower than ELM. Random forest (RF) [29] algorithm constructs a "forest" by generating a number of decision trees (500 trees in our experiment), each of which plays a role in weak classifier and the final decision is made by voting methods of multiple trees in the "forest". Although sparse representation classification (SRC) has gained much attention recently, the time and space complexities are extremely high when solving large-scale problems [30]. As shown in Table V, ELM achieves the highest classification accuracy with the lowest time cost.

TABLE V
THE DETAILS OF FOUR DIFFERENT CLASSIFIERS WITH FUSION FEATURE

Classifier	Time cost (Training + Testing)	Accuracy
Extreme Learning Machine	100s	92.0%
SVM	1800s	91.2%
Random Forset	750s	83.8%
Sparse representation classification	6050s	85.4%

VI. EXPERIMENTS

A. Experimental Data

This paper collects 800 arias of 8 genres (Jin opera, Peking opera, Qin opera, Henan opera, Shao opera, Cantonese opera, Zhuizi, Kunqu), and each opera genre contains 100 pieces of arias. All of the audio samples are represented by extracting the starting 30 seconds, the middle 30 seconds and the ending 30 seconds at sampling rate of 22050Hz. Frame length is 23ms with half overlapping and texture window length is 9 seconds with half overlapping.

B. Evaluation Methodology

The final classification accuracy of this system is obtained by 10-fold cross-validation. The data set is randomly split into 10 subsets, 9 of which are selected as training set and the left one is selected as testing set. Then repeat this process 10 times until each subset is selected as the testing set for one time. Repeat 10-fold cross-validation 10 times, the mean accuracy is regarded as the final classification accuracy.

C. Confusion Matrix

Confusion matrix is a two-dimension matrix used to record whether a test sample is classified correctly. Each row of the matrix indicates the actual category of a testing sample and each column indicates the theory category predicted by a

TABLE VI
CONFUSION MATRIX OF CHINESE OPERA GENRE CLASSIFICATION

	Jin	Jing	Qin	Yu	Shao	Cantonese	Zhui	Kun
Jin	95	0	2	1	0	2	0	0
Jing	2	82	1	1	8	1	3	2
Qin	2	0	96	1	0	1	0	0
Yu	0	2	1	95	0	1	1	0
Shao	1	1	1	3	94	0	0	0
Cantonese	2	3	0	1	2	91	0	1
Zhui	0	1	2	3	4	1	88	1
Kun	0	0	0	0	0	0	0	100

classifier. For each row and column, opera genre is arranged in the order of Jin opera, Peking opera, Qin opera, Henan opera, Shao opera, Cantonese opera, Zhuizi, Kunqu. For example, the 5th row and the first column of Table VI shows that there is only one piece of Shao Opera aria which is misclassified into Jin Opera. The diagonal of the matrix displays the number of arias that are correctly classified.

Therefore, the average classification accuracy of 10-fold cross validation can be calculated as follows.

$$Accuracy = \frac{\sum_{i=1}^{N} data_{i,i}}{\sum_{i=1}^{N}\sum_{j=1}^{N} data_{i,j}} \quad (4)$$

where $data_{i,j}$ lies in the ith row, jth column of the confusion matrix.

VII. DISCUSSION

As presented in Fig. 5, fusion feature improves classification accuracy in Chinese traditional opera genre classification. During decision fusion stage, some weak features with low dimension give very bad decisions (Fig. 5), which even decrease the overall classification accuracy to 86.3% (Table IV). Therefore, we only select 6 representative features (OSC, MFCCs, NASE, OMSC, MSFM-MSCF and spectral pitch chroma) in decision fusion which achieves an accuracy of 89.1%. Experimental results show that if several weak features are removed away from fusion feature, the mean classification accuracy always decreases. We draw the conclusion that weak features provide supplementary information instead of decision information. Comparing with decision fusion, a huge advantage of feature fusion is that it considers the relationship among different features. Therefore, feature fusion achieves better classification accuracy in this system.

VIII. CONCLUSIONS

This paper proposes an effective framework for Chinese traditional opera genre classification based on multi-feature fusion technique and extreme learning machine. Though most features are commonly used in music and audio classification, they are also effective in classifying Chinese opera genre. Finally, our system reached a mean classification accuracy of 92%. Chinese traditional operas have been famous for its unique artistic charm among world culture. We are convinced that it is a valuable artistic attempt to study Chinese operas with computer technology.

For future work, we plan to improve this system from two aspects. In order to gain real-time response, it is necessary to explore lower dimension feature without decreasing the classification accuracy of this system. What's more, there are massive opera resources on the Internet, therefore, it is also suggested to research Chinese operas on large-scale data set.

ACKNOWLEDGMENTS

The research was supported by part of the National Natural Science Foundation (surface project No.61175016, surface project No.61304250).

REFERENCES

[1] H. Zhao-liang, The characteristics of chinese traditional operas geography [j], ECONOMIC GEOGRAPHY 1 (2000) 016.
[2] Z. Zhang, X. Wang, Structure analysis of chinese peking opera, in: Natural Computation (ICNC), 2011 Seventh International Conference on, Vol. 1, IEEE, 2011, pp. 237C241.
[3] J. Sundberg, L. Gu, Q. Huang, P. Huang, Acoustical study of classical peking opera singing, Journal of Voice 26 (2) (2012) 137C143.
[4] Y.-B. Zhang, J. Zhou, X. Wang, A study on chinese traditional opera, in: Machine Learning and Cybernetics, 2008 International Conference on, Vol. 5, IEEE, 2008, pp. 2476C2480.
[5] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, Speech and Audio Processing, IEEE transactions on 10 (5) (2002) 293C302.
[6] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, L.-H. Cai, Music type classification by spectral contrast feature, in: Multimedia and Expo, 2002. ICME02. Proceedings. 2002 IEEE International Conference on, Vol. 1, IEEE, 2002, pp. 113C116.
[7] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, B. Kégl, Aggregate features and adaboost for music classification, Machine learning 65 (2-3) (2006) 473C484.
[8] C.-H. Lee, J.-L. Shih, K.-M. Yu, H.-S. Lin, Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features, Multimedia, IEEE Transactions on 11 (4) (2009) 670C682.
[9] Y.-F. Huang, S.-M. Lin, H.-Y. Wu, Y.-S. Li, Music genre classification based on local feature selection using a self-adaptive harmony search algorithm, Data & Knowledge Engineering 92 (2014) 60C76.
[10] B. Jin-hua, B.-e. LIANG, A literature review on the protection and utilization of chinas intangible cultural heritage [j], in: Tourism Forum, Vol. 6, 2008, p. 026.
[11] P. K. Atrey, M. A. Hossain, A. El Saddik, M. S. Kankanhalli, Multimodal fusion for multimedia analysis: a survey, Multimedia systems 16 (6) (2010) 345C379.
[12] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, Vol. 2, IEEE, 2004, pp. 985C990.
[13] H.-G. Kim, N. Moreau, T. Sikora, Audio classification based on mpeg-7 spectral basis representations, Circuits and Systems for Video Technology, IEEE Transactions on 14 (5) (2004) 716C725.
[14] H.-G. Kim, N. Moreau, T. Sikora, MPEG-7 audio and beyond: Audio content indexing and retrieval, John Wiley & Sons, 2006.
[15] M. A. Bartsch, G. H. Wakefield, Audio thumbnailing of popular music using chroma-based representations, Multimedia, IEEE Transactions on 7 (1) (2005) 96C104.
[16] B. Logan, et al., Mel frequency cepstral coefficients for music modeling., in: ISMIR, 2000.
[17] S. Sukittanon, L. E. Atlas, J. W. Pitton, Modulation-scale analysis for content identification, Signal Processing, IEEE Transactions on 52 (10) (2004) 3023C3035.
[18] C.-H. Lee, J.-L. Shih, K.-M. Yu, J.-M. Su, Automatic music genre classification using modulation spectral contrast feature., in: ICME, 2007, pp. 204C207.
[19] D. Jang, C. D. Yoo, Music information retrieval using novel features and a weighted voting method, in: Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on, IEEE, 2009, pp. 1341C1346.

[20] Y. Zhu, M. S. Kankanhalli, S. Gao, Music key detection for musical audio, in: Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International, IEEE, 2005, pp. 30C37.
[21] G. Peeters, X. Rodet, A large set of audio feature for sound description (similarity and classification) in the cuidado project, Tech. rep., Ircam, Analysis/Synthesis Team, 1 pl. Igor Stravinsky, 75004 Paris, France (2004).
[22] C. Sanderson, K. K. Paliwal, Identity verification using speech and face information, Digital Signal Processing 14 (5) (2004) 449C480.
[23] C. Neti, B. Maison, A. W. Senior, G. Iyengar, P. Decuetos, S. Basu, A. Verma, Joint processing of audio and visual information for multimedia indexing and human-computer interaction., in: RIAO, 2000, pp. 294C301.
[24] S. Wang, Y. Chen, Z. Chen, Recognizing transportation mode on mobile phone using probability fusion of extreme learning machines, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 21 (supp02) (2013) 13C22.
[25] Z. W. Geem, J. H. Kim, G. Loganathan, A new heuristic optimization algorithm: harmony search, Simulation 76 (2) (2001) 60C68.
[26] J. Yang, J.-y. Yang, D. Zhang, J.-f. Lu, Feature fusion: parallel strategy vs. serial strategy, Pattern Recognition 36 (6) (2003) 1369C1381.
[27] S. S. UG Mangai, A survey of decision fusion and feature fusion strategies for pattern classification, IETE Technical Review (Medknow Publications & Media Pvt. Ltd.) 27 (4) (2010) 293C307.
[28] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.
[29] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5C32.
[30] J.-B. Huang, M.-H. Yang, Fast sparse representation with prototypes, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 3618C3625.