International Journal of Latest Trends in Engineering and Technology
Vol.(9)Issue(2), pp.128-131
DOI: http://dx.doi.org/10.21172/1.92.22
e-ISSN:2278-621X

A HYBRID APPROACH TO IDENTIFY A SINGER IN A VIDEO SONG

S.Metilda Florence[1] & Dr. S. Mohan[2]

Abstract- Extraction of data concerning the video contents automatically refers to Automatic Video Annotation System. The extracted information will function as the initial step for numerous information access strategies like searching, comparison, and classification. The vast number of video songs accessible to the public requires tools to efficiently retrieve and manage the music of interest to the users. Thus, the projected system would enable them to look for his or her favourite singer's video song. Underlying Singer within the video songs is distinguished by mining their Linear Prediction Coefficients (LPC) and Spectral features. A unique methodology is applied to boost the performance of the classifiers. A new hybrid system by combining the outputs of two classifiers' posterior possibilities using product and Mean rule are proposed. The combined classification result shows significant improvement in the outcome. Naïve Bayes and K–Nearest Neighbor (K-NN) algorithms are used for Statistical Analysis. The projected system gives 93% accuracy in identifying a Singer in a video song. Experimental outcomes show that users will retrieve the songs of their selection.
Keywords – Video annotation, Classification, Combined Classifiers, Artist Identification, Hybrid system

I. INTRODUCTION

YouTube has more than one billion users. Every day people watch many numerous hours of videos on You Tube. In close to future looking at online videos can increase in enormous amount. There is a demand to utilize the available material in a video store. In many video production corporations, this task remains manual. It is often a tedious job that relies on the human. We have proposed a unique technique that annotates the video automatically from audio data. The most contribution of this work is that the use of music to annotate the video.

Commonly in Video Annotation, the videos are annotated in following categories: Genre Classification [1] (cartoon, commercial, sports, movie, news, music), Objects within the video [2] (car, mountain, sky, road, man, animal, birds etc.) and semantics Level [3,4 and 5] (desert, indoor, outdoor, shore etc.). In these works, music within the video not targeted a lot. We have entirely concentrated on music and categorize the video based on music as follows: Artist Identification (Singer identification), Instrument Recognition (Piano, Violin, Guitar, etc.), Mood Classification (Happy, Sad, Angry, etc.) and Genre Classification (Rock, Pop, Classical, etc.). In Genre Classification, five Genres namely Blues, Classical, Folk, Raphi, and Rock are taken and classified [6]. Similarly, in Mood Classification three Moods are considered namely angry, joy and sad. Principal Component Analysis technique is applied to reduce the size of the feature set and then classified [7]. In Instrument Recognition, three Instruments are taken (Guitar, Flute, and Drum), by using Sparse Non-Negative Matrix Factorization and onset values these Instruments are identified [8]. In this paper, the implementation and results of Artist Identification module using Hybrid Classification System are discussed in detail. This Hybrid System will enable the music lovers to locate quickly the video file which contains their favorite singer's song. The rest of the paper is organized as follows. Proposed research method is explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

2. RESEARCH METHODOLOGY

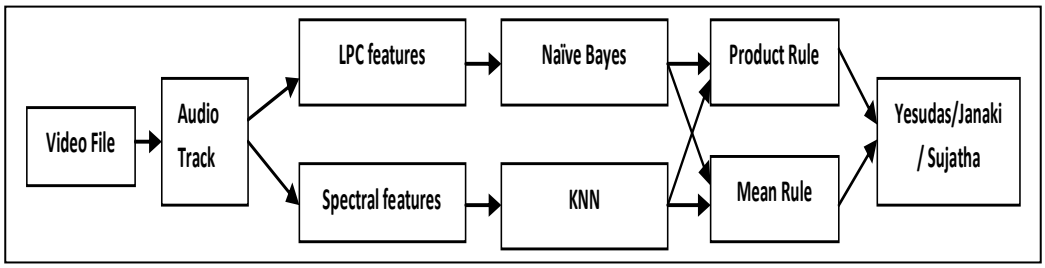
The proposed Hybrid system used three Singers video songs namely K.J. Yesudas, S. Janaki, and Sujatha. 100 video songs of 10 seconds duration are taken for each Singer. From each video song the vocal track alone is extracted. Mathematical functions are applied to calculate the LPC and spectral features from the extracted signal. These features are applied to standard classifiers for classification. The proposed system is depicted in Figure. 1 and the details are given below.

2.1 Video File Archive

300 video songs are collected from Video CDs and Internet. Collected video files are stored in this Video block for further process. Video files can be of any type like MPEG, WMV, and AVI.

Fig.1: Block diagram of Hybrid Classifier System

[1] Research Scholar, Research and Development Centre,Bharathiar University,Coimbatore,India
[2] CCIS, Al Yamamah University, Kingdom of Saudi Arabia



2.2 Extracting Audio track

In the initial stage, audio track from video files is needed to be extracted. This task is implemented by using Multimedia functions in Mat lab. The Extracted audio tracks are ready for further processing.

2.3 Feature Extraction

In this phase vocal from background music is isolated. We intend to identify frequency range of singing. A band-pass filter is used to allow the vocal range to pass through while weakening other frequency areas. Chebychev Infinite Impulse Response (IIR) digital filter of order 12 is used to achieve this. Other Instruments may scatter energy in this range, for example, drum. Inverse comb filter bank is used to separate the voice from the other sources.

The features are used to signify timbre texture based on typical features proposed for music-speech separation [9]. From the available features, most relevant features LPC and Spectral elements are defined below.

2.4 Linear Prediction Coefficients

Levinson-Durbin recursion[10,11] is used to compute the corresponding reflection coefficients and LPC parameters, are given in (1).

$$LD = \frac{1}{T}\sum_{t=1}^{T}(\hat{s_t} - s_t)^2 = \frac{1}{T}(\hat{s} - s)'(\hat{s} - s) \quad (1)$$

Given the signals, s = [s_1, s_2… s_T], a linear predictor of order n predicts the sample at time t as a weighted linear interpolation of its n preceding samples (2):

$$\widehat{s_t} = \sum_{i=1}^{n} a_i\, s_{t-i} \Rightarrow \quad \hat{s} = L\,a \quad (2)$$

where

$$L = \begin{bmatrix} s_0 & s_{-1} & s_{-2} & s_{-3} & \cdots & s_{-n+1} \\ s_1 & s_0 & s_{-1} & s_{-2} & \cdots & s_{-n+2} \\ & & \vdots & & & \\ s_{T-1} & s_{T-2} & s_{T-3} & s_{T-4} & \cdots & s_{-n+T} \end{bmatrix} \quad \text{and} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

where { a_i, $1 \le i \le n$ } are known as the Linear Prediction Coefficients. L is known as a Toeplitz matrix.

2.5 Spectral features

1) The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights:

$$\text{centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (3)$$

where x(n) represents the weighted frequency value, or magnitude, of bin number n, and f(n) represents the center frequency of that bin.

2) The Spectral Rolloff is the frequency Rt under which 95% of the power distribution is concentrated.

$$\sum_{n=1}^{Rt} Mt[n] = 0.85 * \sum_{n=1}^{N} Mt[n] \quad (4)$$

Where n is time index ranges from 0 <= n <= N-1, N is duration of file and t is current time frame.

3)The Spectral flux is defined as the squared difference between the normalized magnitudes of successive spectral distributions.

$$Ft = \sum (Nt[n] - Nt-1[n])^2 \quad (5)$$

Where $N_t[n]$ and $N_{t-1}[n]$ are the normalized magnitudes of the Fourier transform at the current time frame t, and the previous time frame t-1, respectively.

2.6 Hybrid Classifiers

In this paper, Naive Bayes and K-NN classifiers are used for classification. LPC features are applied to Naïve Bayes and Spectral features are applied to K- NN classifier. The combing of information from different classifiers is produced by

building new predictions for the posterior probabilities from the individual classifiers' predictions. The combined prediction for class wi is discussed below.

2.6.1 Product Combination Rule

It has been shown in the literature [12] that when using uncorrelated and independent feature sets, a product combination rule is a good choice. Let R be the number of independent classifiers and w (w= $w_1, w_2, ..w_n$) the known n classes. When every classifier produces conditional output probabilities $P(w_k| x_i)$, k = 1 … n, according to the feature vector x_i, the product combination rule to assign an input sample to the class w_c is presented as follows,

$$w_c = \arg\max_{k=1}^{n} \left[\prod_{i=1}^{R} P(w_k|x_i) \right] \quad (6)$$

where the final decision is made according to the maximum of combined values.

2.6.2 Mean Combination Rule

The averaging strategy works well when correlated outputs are used [12]. In our case, Rhythmic and Spectral features extracted from the same person are assumed to correlate very closely, so that the Mean combination rule can be applied to it correspondingly,

$$w_c = \frac{1}{n} \left[\sum_{i=1}^{R} P(w_k|x_i) \right] \quad (7)$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1 Dataset

Totally 300 songs are collected from the Internet and Video CDs. The duration is restricted to 10 seconds for experimental purpose.

3.2 Implementation in Mat lab

The Hybrid System is implemented in Mat lab version 7.11.0 (2010b). Using Matlab code the audio track is extracted from video, and the vocal track is isolated by using the IIR and inverse comb filter. By applying mathematical equations LPC and Spectral features are calculated. These values are fed into standard classifiers for classification. Two efficient classifiers namely Naïve Bayes and K-NN are used to train and test the feature set. LPC feature set is classified by Naïve Bayes and Spectral feature set by KNN classifiers. The outputs of two classifiers' posterior probabilities are combined by using Product (6) and Mean (7) rule. The results of Hybrid classifier system is given in figure 2.

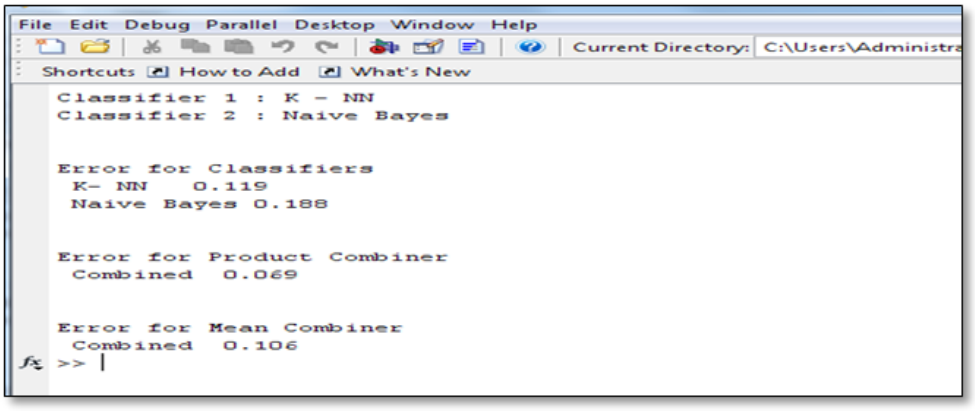


Fig.2: Output of Hybrid Classifier System

Table 1.Analysis Report for Hybrid Classifier System

Classifiers	KNN	Bayes	Combined (Product)	Combined (Mean)
Error rate	0.119	0.188	0.069	0.106
Accuracy	88%	81%	93%	89%

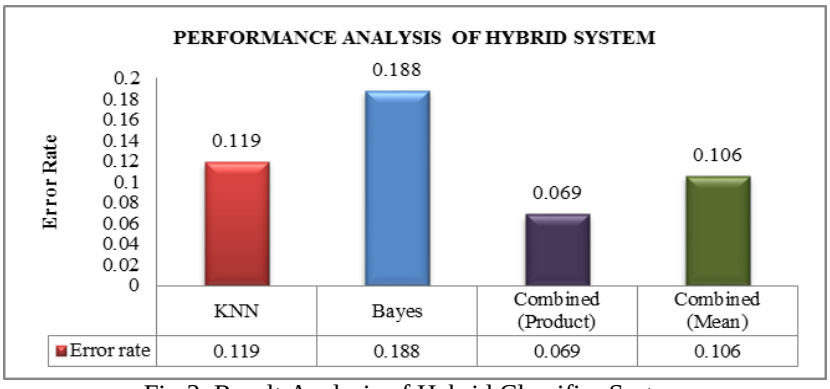


Fig.3: Result Analysis of Hybrid Classifier System

The error rate for the individual classifiers is high when compared with the hybrid classifiers. For our system product combiner gives more accuracy than Mean combiner. Result analysis is presented in Table1, and the same is depicted in Figure 3. The result shows that the Hybrid classifiers are performing well when compared with individual classifiers.

4.CONCLUSION

A novel and efficient approach for identifying a Singer in the video store is presented. The Hybrid Classifier System will enable the music lovers to choose their favourite video song. In this Hybrid system, three Singers namely K.J. Yesudas, S. Janaki, and Sujatha are selected for analysis. 100 video songs of 10 seconds duration are taken for each Singer. From each video song, the vocal track alone is extracted by using IIR digital filter and inverse comb filter. Mathematical functions are applied to calculate the LPC and Spectral features from the extracted signal. These features are applied to Naïve Bayes and K-NN classifiers respectively. The outputs of two classifiers' posterior probabilities are combined by using Product and Mean rule. This Hybrid Classifier System gives a maximum of 93% accuracy in identifying a Singer in a video song. This Framework performs the analysis for 10 seconds duration for testing. The period can be extended to the whole song. This work can be continued to increase the number of Singers for analysis. Size of the dataset can also be improved by including more contributing features from the audio track.

5. REFERENCES

[1] Tianzhu Zhang, Changsheng Xu, Guangyu Zhu, Si Liu and Hanqing Lu, A Generic Framework for Video Annotation via Semi-Supervised Learning, IEEE Transactions On Multimedia, Vol. 14, No. 4, August 2012

[2] Jinhui Tang, Xian-Sheng Hua, Meng Wang, Zhiwei Gu, Guo-Jun Qi, and Xiuqing Wu, Correlative Linear Neighborhood Propagation for Video Annotation,IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 39, No. 2, April 2009

[3] Cencen Zhong and Zhenjiang Miao, A Two-View Concept Correlation Based Video Annotation Refinement, IEEE Signal Processing Letters, Vol. 19, No. 5, May 2012

[4] Yu-Gang Jiang, Qi Dai, Jun Wang,Chong-Wah Ngo, Xiangyang Xue and Shih-Fu Chang, Fast Semantic Diffusion for Large-Scale Context-Based Image and Video Annotation, IEEE Transactions On Image Processing, Vol. 21, No. 6, June 2012

[5] Ming-Fang Weng and Yung-Yu Chuang,Cross-Domain Multicue Fusion for Concept-Based Video Indexing, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 34, No. 10, October 2012

[6] S.Metilda Florence,Dr.S.Mohan, Automatic Video Annotation for Music Genre Based on Spectral and Cepstral Features,in ELSEVIER Proc. Int. Conf. on Applied Information and Communications Technology(ICAICT 2014),Oman, pp. 27 – 32.

[7] S.Metilda Florence,Dr.S.Mohan, Automatic Video Annotation for Music Mood using PCA with Rhythm and Cepstral Features, ELSEVIER Proc. Int. Conf.on Emerging Research in Computing, Information, Communication and Applications, ERCICA 2014, Bangalore, pp. 355 - 360.

[8] S.Metilda Florence,Dr.S.Mohan, A Novel Search Engine for Identifying Musical Instruments in a Video File,International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34144-34148

[9] E. Scheirer and M. Slaney, Construction and evaluation of a robust multifeature speech/music discriminator, in Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1997, pp. 1331–1334.

[10] https://www.comp.nus.edu.sg/~simkc/slides/lecture04.pdf

[11] A. Krishnamurthy and D.Childers,Two channel speech analysis and signal processing in IEEE Transactions on signal processing vol.4 issue 34, 1986.

[12] [12] D.M.J. Tax, M. van Breukelen, R.P.W. Duin, and J. Kittler, "Combining classifiers by averaging or by multiplying?, " Pattern Recognition, vol. 33, pp. 1475–1485, 2000.