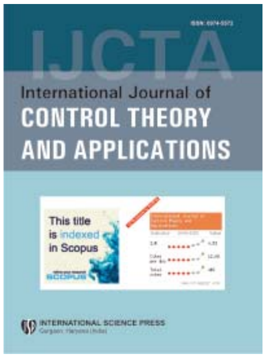

International Journal of Control Theory and Applications

ISSN : 0974–5572

© International Science Press

Volume 9 • Number 40 • 2016

A Supervised Learning of a Singer in a Video Song using Linear Prediction Code Coefficients

Metilda Florence.S[a] and Mohan.S[b]

[a] *Assistant Professor, SRM University*

[b] *CCIS, Al Yamamah University, Riyadh 11512, Saudi Arabia*

Abstract: Automatic Video Annotation System refers to the extraction of information about the video contents automatically. Extracted information can serve as an initial step for various data access techniques such as browsing, searching, comparison, and categorization. In past research papers, annotating music information in a video was not covered much. The massive amount of video songs reachable to the general public calls for developing tools to efficiently retrieve and manipulate the music of interest to the end users. Thus, the proposed system would enable them to search for their favourite singer's video song in a large unstructured dataset. The proposed methodology performs the search in a video store by comparing the content of the video and not the user's textual query and tags associated with the videos. We make an effort for identifying underlying Singer in the video songs by mining their feature called Linear Prediction Code Coefficients (LPCC). Five classification algorithms are used for Statistical Analysis, namely to mention a few Naïve Bayes algorithm, Sequential Minimal Optimization (SMO) algorithm, etc. The proposed system gives a maximum of 95% accuracy in identifying a Singer in a video song using SMO classifier. Experimental outcomes show that users can retrieve the songs of their choice from a large dataset.

Keywords: *Video Annotation, Classification, Artist Identification, Content Based Search, Linear Prediction Code Coefficients, Signal Processing.*

1. INTRODUCTION

A recent statistics states that [1] YouTube has more than 1 billion users. Each day people watch hundreds of millions of hours of videos on You Tube. In Short period watching online videos will increase in high amount. There is a requirement to annotate the available content to reuse the material available in a large video store. In several video production companies, this task is still executed manually, and it is a tough and tedious job, which besides, has led to depend on the human. We have proposed a novel technique that annotates the video automatically from audio information. The main contribution of this work is the use of music to annotate video, which is a much less explored problem.

In prior works, the videos are annotated in following categories: Motion and Gesture recognition [2,3,4], Genre Classification [5] ,Objects in the video[6] and Semantic Level [7, 8, 9]. In these works, music in the video is not focused much. We have fully concentrated on music and categorize the video songs 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 this paper, the implementation and results of Artist Identification module are discussed in detail. Large amount of different feature sets, mainly originating from the area of speech recognition are available to characterize audio signals [10]. From these features LPCC is selected for Artist Identification system. LPCC is selected mainly for differentiating the biological structure of human vocal tract[11]. Five efficient classifiers are applied to perform statistical analysis of generated dataset. This system will enable the music lovers to locate quickly the video song which contains the music they are interested in.

2. PROPOSED METHODOLOGY

In the proposed Artist(Singer) Identification system, three Singers namely Yesudas, S.Janaki and Sujatha are selected for analysis. For each Singer 100 video songs of length 10 seconds duration are taken. From each video song the vocal track alone is extracted. Mathematical functions are applied to calculate the LPCC features from the extracted signal. These features are used to standard classifiers for classification. The proposed system is depicted in Fig. 1 and the details are given.

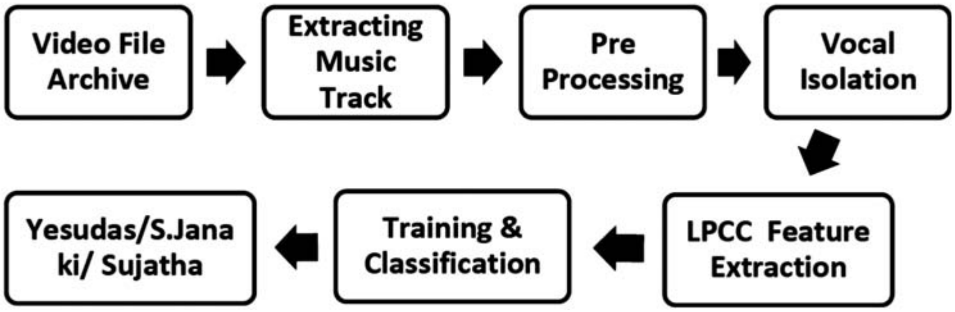

Figure 1: Block diagram of Proposed System

2.1. Video file archive

A novel database is prepared from South Indian movie songs. Video songs are collected in 15 to 20 years duration for getting different varieties of songs. More contributing video clips of duration 10 seconds are collected from Video CDs and Internet. Compiled video files are stored in this Video block for further process. Video files can be of any type like MPEG, WMV and AVI.

2.2. Extracting music track

To proceed further, we need to extract the music signal from video file. Extraction of the music track is achieved by Matlab code. Using MultimediaFileReader() method, audio frames are read one by one from video and stored separately by using MultimediaFileWriter() method. Extracted audio tracks are ready for further processing.

2.3. Preprocessing

Wiener Filter is used to filtering noise and unwanted signals from extracted source signal. It separates signals based on their frequency spectra [12]. The concept behind this filter is, it will allow only signal frequencies and block typically noise frequencies.

$$W[frq] = \frac{Sig[frq]^2}{Sig[frq]^2 + N[frq]^2} \tag{1}$$

W[frq] is determined by the frequency spectra of the noise N[frq]and signal Sig [frq].

2.4. Vocal isolation

In this phase vocal from background music is isolated.The majority of energy in the singing voice falls between 200Hz and 2000 Hz (There may be deviation depending on the Singer). Our motive is to perceive frequency range of the song.A simple method is to filter the audio signal with a band-pass filter which allows the vocal range to pass through while weakening other frequency areas. Chebychev Infinite Impulse Response (IIR), a digital filter of order 12, is used to achieve this. This filter will suppress other Instruments that fall outside of this frequency region. But the voice is not the only component having energy in this region. Other Instruments may scatter energy in this range, for example, Drum. Another measure is needed to separate the voice from the other sources.

The Singing voice is highly harmonic [13] than other high energy sounds in this region, particularly Drum is not harmonious as vocal. Inverse comb filter bank is used to detect the large amounts of harmonic energy. By passing the previously filtered signal (F) through a set of inverse comb filters with varying delays, we can extract only the harmonic components of the signal. By taking the ratio of the total signal energy of the maximally harmonic attenuated signal, harmonicity can be measured:

$$\text{Harmonicity} = \frac{\mathrm{F}_{\text{original}}}{\min_i(\mathrm{F}_{\text{filtered}, i})} \tag{2}$$

By thresholding the harmonicity against a fixed value, we have a detector for harmonic sounds. Most of these signals correspond to regions of singing.

2.5. Feature extraction

Various numerical values are extracted from the signal describes the Music. These numerical values are called as features of the signal. A lot of different feature sets, mainly originating from the area of speech recognition, have been proposed to characterize audio signals [14]. Many features can be extracted from a signal. From prior art we selected LPCC is a most contributing feature for Artist Identification system, which is defined below.

2.6. Linear Prediction Code Coefficients (LPCC)

To decrease the number of parameters needed to represent a signal, LPCCs are used [15]. The value of the signal at time *t* is approximated with the linear combination of real signal values in previous moments as shown in (3) where a_i are LPCCs. LPCCs are calculated so that they minimize the error between an actual signal and the one calculated using LPCs over the interval of interest as defined in (3). We assume that the signal of interest is zero outside this range.

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

$$\widehat{S}_t = \textstyle\sum_{i=1}^{n} a_i s_{t-i}$$

$$\Rightarrow \qquad \hat{s} = \mathrm{L}a \tag{3}$$

Where

$$\mathrm{L} = \begin{bmatrix} s_0 & s_{-1} & s_{-2} & s_{-3} & \cdots & s_{-n+1} \\ s_1 & s_0 & s_{-1} & s_{-2} & \cdots & s_{-n+2} \\ s_{\mathrm{T}-1} & s_{\mathrm{T}-2} & s_{\mathrm{T}-3} & s_{\mathrm{T}-4} & \cdots & s_{-n+\mathrm{T}} \end{bmatrix} \tag{4}$$

And $$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

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

Auto-correlation algorithm (Levinson-Durbin recursion) and Covariance algorithm are used to compute Linear Prediction Coefficients. Covariance algorithm is more common and can be utilized with no conditions. But the drawback is that the result will not be stable. In autocorrelation method, it is guaranteed to be stable. Covariance algorithm is more accurate for periodic speech sounds than autocorrelation. Autocorrelation performs better for fricative sounds than covariance algorithm [16].

Based on this for our proposed system Levinson-Durbin recursion algorithm is chosen.The equations of the Levinson-Durbin recursion, which are used to compute the corresponding reflection coefficients and LPC Cofficients, are given in (4).

$$\begin{aligned} \mathrm{LD} &= \frac{1}{\mathrm{T}} \Sigma_{t=1}^{\mathrm{T}} (\widehat{s_t} - s_t)^2 \\ &= \frac{1}{\mathrm{T}} (\hat{s} - s)' \, (\hat{s} - s) \end{aligned} \tag{4}$$

2.7. Classifiers and Classification

It is necessary to use more than one classifier to get the average accuracy. Because a single classifier will not always gives the best result for different applications. In this module, five efficient classifiers are used to train and test the dataset. They are given as follows: Naive Bayes, Sequential Minimal Optimization (SMO), Multiclass Classifier, J48 and Random Tree.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Dataset

From the Internet and Video CDs, 100 video songs of each Singer are collected. Totally 300 songs are received and the duration of each song is trimmed to 10 seconds.

3.2. Implementation in Mat lab

The proposed 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 LPCs are calculated. These values are fed into standard classifiers.

Five efficient classifiers are used to train and test the feature set. They are Naive Bayes, Sequential Minimal Optimization (SMO), Multiclass Classifier, J48, and Random Tree. For classification, ten cross validation technique was used in WEKA tool. Cross-validation is a model validation technique for assessing how the results of the statistical analysis will generalize to an independent dataset.

In Table 1, the overall accuracy of all these classifiers is listed. This table shows that SMO classifier has given the highest accuracy of 95 % and J48 classifier given the least accuracy of 83%. To conclude, Artist identification system provides a maximum of 95% accuracy in identifying a Singer in a video song using SMO classifier. Fig. 2 gives the pictorial representation of the accuracy percentages.

Table 1
Analysis Report

Classifier	*Accuracy %* *Yesudas*	 *S. Janaki*	 *Sujatha*	*Average Accuracy %*
Naïve Bayes	87	87	82	85
SMO	96	97	92	95
Multiclass	95	93	90	93
Random Forest	94	94	88	92
J48	82	87	80	83

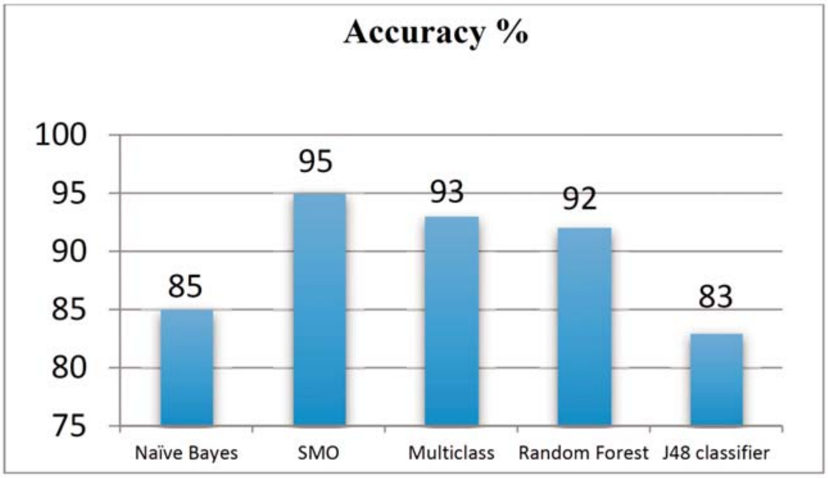

Figure 2: Classification Accuracy for different classifiers

4. CONCLUSION

A novel and efficient approach for identifying a Singer in the video store are presented. This methodology enables the music lovers to choose their favourite singer's video song. In this Artist Identification system, three Singers namely Yesudas, S.Janaki and Sujatha are selected for analysis. 100 video songs of length 10 seconds duration are taken for each Singer. From these video songs the vocal track alone are extracted by using IIR digital filter and inverse comb filter. Mathematical functions are applied to calculate the LPCC feature from the extracted signal. These features are used to five classifiers for classification. This Artist Identification system gives a maximum of 95% accuracy in identifying a Singer in a video song using SMO classifier. Existing search engines will search the video by their tags not by content. But the proposed system will identify the video songs by their content. This Framework restricts the identification for 10 seconds period for experimental purpose. This duration can be extended for the entire song. In future, this work can be extended to cover more Singers. Size of the dataset can also be increased by including more contributing features from the audio track. The increase in features may increase the accuracy percentage.

REFERENCES

[1] https://www.youtube.com/yt/press/statistics.html.

[2] Yixin Gao - Johns Hopkins University," A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery" in *IEEE Transactions on Biomedical Engineering*, Volume.PP, issue 99, 2017.

[3] Ra'na Sadeghi Chegani, Carlo Menon,"Tracking hand movements and detecting grasp" in *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*,2016.

[4] J. Pons, J. Prades-Nebot, A. Albiol, J. Molina," Fast motion detection in compressed domain for video surveillance" in *IET Journals & Magazines*,2002.

[5] Publications - Tianzhu Zhang's Homepage, https://sites.google.com/site/zhangtianzhu2012/publications (accessed April 01, 2017)

[6] 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 .

[7] Cencen Zhong and Zhenjiang Miao,"A two-view concept correlation based video annotation refinement", *IEEE Signal Processing Letters*, VOL. 19, NO. 5, MAY 2012 .

[8] 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.

[9] 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.

[10] George Tzanetakis, Perry Cook ,"Musical genre classification of audio signals", *IEEE Transactions On Speech And Audio Processing*, Vol. 10, No. 5, July 2002.

[11] Genevieve I. Sapijaszko, Wasfy B. Mikael, "An overview of recent window based feature extraction algorithms for speaker recognition", *IEEE 55th International Midwest Symposium on Circuits and Systems* , pp 880-883,2012.

[12] Steven W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, book, chapter – 17.

[13] P. R. Cook, *Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing*. Ph.D. Thesis. Stanford University, Stanford, CA, 1990.

[14] George Tzanetakis, Perry Cook, "Musical genre classification of audio signals", *IEEE Transactions On Speech And Audio Processing*, Vol. 10, No. 5, July 2002.

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

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