International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34144-34148

A Novel Search Engine for Identifying Musical Instruments in a Video File

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

[1]Research Scholar, Research and Development Centre,Bharathiar University,Coimbatore,India
metilda_florence@yahoo.com
[2]Professor, Department of Electrical and Communication Engineering,
Coimbatore Institute of Engineering and Technology, Coimbatore,India,s.mohan77@gmail.com

Abstract
To improve searching performance in a video store, this paper presents anovel methodologybased on content-based searching which enables the searching for a musical instrument in a video file.The main innovation of this methodology is to perform musical instrument search in a video by comparing the content of the video rather than comparing user's textual query and tags associated with the videos. Using Sparse Non-Negative Matrix Factorization (SNMF) algorithm, the extracted polyphonic audio track from video is split into sources which are identified by onset. Onset values for each instrument are calculated using Complex Spectral Difference method (CSD). In this paper we concentrated on three instruments namely Guitar, Flute, and Drums. We present experimental result that shows that our approach is effective in identifying the musical instrument with acceptable level of confidence.

Keywords: Video annotation,Musical Instrument, content based,complex spectral difference

1 Introduction

A recent statistics of YouTube states that it has more than 1 billion users. Each day people watch hundreds of millions of hours of videos on You Tube. In near future watching online videos will increase in huge amount. In the proposed system it focuses on music track in the video file. This will help the music lovers to locate their interested video song from the video store. The proposed novel search engine will perform the search based on the content of the video file rather than the filename or Meta data.

In this paper, we have used three musical instruments namely Guitar, Flute, and Drums. The dataset contains 300 video songs which were collected from the internet and Video CDs. For each instrument 100 songs were collected.Lengths of all video songs are fixed into 10 seconds.Using SNMF algorithm [1], the extracted audio signal from video are separated into different sources. By using Onset value these sources are identified. Several methods are available to calculate the onset value, from that we preferred the CSD method as being the most appropriate for our work [2]. The reason for selecting SNMF and Onset calculation methods are discussed in section 2.

The paper is structured as follows. Literature survey is given in Section 2. In Section 3, we provided a detailed study of the proposed system. In this section SNMFalgorithm, onset calculation, wiener filtering and thresholding are discussed. Implementation and experimental results are given in section 4. Conclusion and future worksare given in section 5.

2 Literature Survey

The NMF based algorithm can beused inBlind Source Separation (BSS)methods [3]. BSS is theprocess of split a set of source signals from a set of mixed signals, without the help of information (or with very minute information) about thesource signals or the mixing process.

Like NMF the most normally used technique in BSS isIndependent Component Analysis (ICA). In ICA, the linear representation of non-Gaussian data is calculated so as to make the components statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation etc. But when both sources and mixing matrices are unknown; ICA cannot determine the variances (energies) of the independent components as well as the order of the independent sources because of the basic functions are ranked by non-gaussianities [4].

Lee and Seung [5] have recommended a solution for BSS problem with non negativity constraints. NMF does not require the independence assumption, and is not restricted to data lengths. It yields more important to the basis vectors for reconstructing the underlying signal than the activation vectors. In NMF the basis functions are not ranked.It is identified that NMF performs better thanICA in BSS environment. The spatial and temporal correlations between variables are more accurately taken into account by NMF which helps to makes NMF a useful tool for decomposition of multivariate data.

Various NMF algorithms are available such as Regularized Expectation Minimization MaximumLikelihood Algorithm (EMML), Regularized Image Space Reconstruction Algorithm (ISRA), Itakura Saito NMF algorithm (IS-NMF) and Sparse Non-Negative Matrix Factorization (SNMF) Algorithm.We have selected SNMF algorithm for our implementation. In order to optimize our system with respect to time we need to reduce the computation time. Based on this criteria we have selected SNMF algorithm, its computation timeis comparatively lesser [6] than other NMF algorithms.

To identify a musical instrument in a complex domain we need to detect the required signal either as a result of energy change in the magnitude spectrum or modification in phase values in the phase spectrum. To achieve this CSD method is appropriate for calculating the onset value [7]. This method is discussed in section 3.4.

3 Proposed system

The proposed system helps the music interested users to locate the video file they are interested in.Our system is able to identify (Guitar, Flute and Drums) three instruments in a video file. It performs the search based on video content. The entire details of the proposed system are explained here.

3.1 Proposed algorithm

To perform the contentbased searching in a video store a novel algorithm is proposed. This algorithm will extract the musical signal from the video file and split the signal into various sources using SNMF algorithm and identify it by onset value. The onset value varies for each musical instrument based on their frequencies.

Algorithm:

1. Get the musical instrument name to be searched in a video store.
2. Repeat the following steps (i) through (vi) for each video file until it reaches the end of archive:

(i) Without playing the video file extract the audio track.
(ii) Preprocess the audio signal.
(iii) Using SNMF algorithm the signal is separated into various sources.
(iv) Based on the onset values the instruments are classified.
(v) Using threshold value the presence or absences of the instruments are identified.
(vi) If the given instrument is present then Store the result in a vector.

3. Finally display the content of the vector.

The overall flow of the proposed system is depicted in Fig.1.

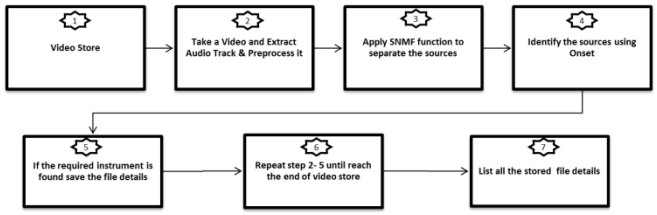

Fig.1. Block diagram of proposed system

3.2Preprocessing

The retrieved audio signals are pre-processed for getting a uniformed signal. After applying normalization the signal has been uniformly scaled. To obtain time localization we performed Fourier analyses. Most commonly used windowing functions are Hamming windows and Hanningwindows[8].In our system we used Hanning windowing method and it can be computed as:

$$hann(l) = 0.5\ (1 - \cos(\tfrac{2\pi l}{N-1}))\ 0 < l < N \qquad (1)$$

Where N is the length of the Hanning window. FFT is then applied on each pre-emphasized, Hanning windowed frame to obtain the corresponding spectrum. The audio samples are sampled at 44.1 KHz. Music samples are segmented into 70 ms frameswith 50 % overlapping of windows to get FFT [9] and the spectral features are extracted.

3.3NMF Algorithm

Instead of using predefined features, our system is using NMF algorithm to identify the musical instruments in the signal. NMF is a popular dimension reduction technique, active for non-subtractive, parts-based exemplification of nonnegative data. Given a data matrix M of dimensions I × J with nonnegative entries, NMF is the problem of finding a factorization

$$M \approx UV \qquad (2)$$

where U and V are non-negative matrices of dimensions I ×K and K × J, respectively. K is usually chosen such that IK + KJ≪I J, hence reducing the data dimension. The factorization (2) is generally obtained by minimizing a cost function defined by

$$D(M|UV) = \sum_{f=1}^{I} \sum_{n=1}^{J} d(M_{fn} \,|\, [UV]_{fn}) \qquad (3)$$

where d(m|n) is a function of two scalar variables. d is typically nonnegative and takes value zero if and only if (iff) m = n. d is typically popular cost functions for NMF are the Euclidean (EUC) distance, defined as :

$$d_{EUC}(m|n) = \tfrac{1}{2}(m-n)^2 \qquad (4)$$

And the generalized KullbackLeibler (KL) divergence, defined as

$$d_{KL}(m|n) = m\ log\ \left(\tfrac{1}{2}\right) - m + n \qquad (5)$$

One of the most useful properties of NMF is that it usually produces a sparse representation of the data. Sparse NMF is an extension of NMF, in which an additional sparsity constraint is enforced on either the matrix U or V.
The SNMF is calculated as:

$$\mathrm{Min}_{U,V \geq 0}\ D(M, U\ V) + \beta(V) \qquad (6)$$

where β is a penalty term that enforces the sparsity. In practice β (V = $\lambda \sum_{i,j} V_{i,j}$), where λis a parameter which controls the trade-off between sparsity and accuracy of approximation. To use this penalty function a normalization constraint on either U or V is introduced, since trivial solutions minimizing β can be found by letting V decrease and Uincrease accordingly.
For an audio application Let us take M as a time-frequency depiction of a polyphonic music signal, I being the number of frequency bins and J the number of time frames. Under an adequate additivityhypothesis [10], the NMF M ≈ UV may give a separation of the K notes appearing in the input signal, by interpreting U as a basis of note pseudo-spectra and V as the corresponding time envelopes. Following this idea, we can use NMF algorithm for musical signal separation.

3.4 Onset calculation

In order to identify the instrument, we need to calculate the onset of different (Guitar, Flute, and Drum) musical instruments.Onset denotes the beginning of a musical note or other sound, in which the amplitude rises from zero to an initial peak [11]. It is connected to the concept of a transient: all musical notes have an onset, but do not essentially include an initial transient. Various methods are available for detecting the onset of musical notes in audio signals. In this paper we used CSD method.

Onset can be calculated [10] as follows:

$$CSD(n) = \sum_{k=\frac{N}{2}}^{\frac{N}{2}-1} CSD(n,k) \quad (7)$$

$$Where\ CSD(n,k) = \begin{cases} |X(n,k) - X_T(n,k)|, if\ |X(n,k)| \geq |X(n-1,k)| \\ 0, Otherwise \end{cases}$$

X (n, k) represents the *k*th frequency bin of the *n*th frame, N is length of the hanning window.

3.5Wiener Filter

In our systemWiener filter algorithm is used for filtering noise and unwanted signals from the extracted source signals. It separate signals based on their frequency spectra [12]. The concept behind this filter is it will allow, the typically signal frequencies and block typically noise frequencies.

$$W[frq] = \frac{Sig[frq]^2}{Sig[frq]^2 + N[frq]^2} \quad (8)$$

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

3.6 Threshold

Threshold is a value. As soon as data is collected, it is compared with the related Threshold value [13]. If the collected data value does not suit the Threshold value then it indicates that this kind of data might lead to poor performance. The threshold value can be checked as maximum, minimum or equal with the collected data.

In our system after the statistical study the threshold value has been fixed to 0.010.This cut off value tells the presence or absence of a musical instrument.After executing the SNMF function, the input signal will be split into different sources of output signals. Based on the onset value these output signals are categorized into three instruments namely Guitar, Flute and Drums. Since we are concentrating only on these three instruments the remaining output signal sources are not processed.

If the *mean of standard deviation*of each identified signal is greater than 0.010 then it shows the presence of the instrument otherwise the absence of the instrument. For example, from the output signal, take the guitar line, and find the mean of standard deviation of this signal, if the calculated value is greater than the threshold value, then it indicates the presence of guitar in the audio signal

4 Experimental result

4.1 Dataset

From the internet and Video CD we have collected 300 video songs.100 songs for each instrument.Duration of all 300 video files is trimmed to 10 seconds.

4.2Implementation in Mat lab

We have implemented a novel simulator for search engine in Mat lab version 7.11.0 (2010b). Using Graphical User Interface Editor (GUIDE) in mat lab the User Interface (UI) was developed.The collected video songs are placed in a folder.Thesearch process will be performed inside this folder. It will take each video file and extract the audio. After preprocessing the audio signal, SNMF function is called to separate the available sources in the audio track. Initially Onset values for Guitar, Flute, and Drums are calculated.Using these onset values the separated signals are identified.

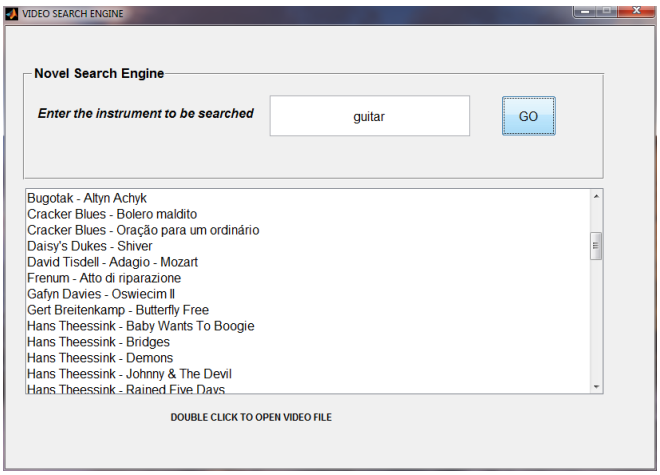

Fig.2. Search screen

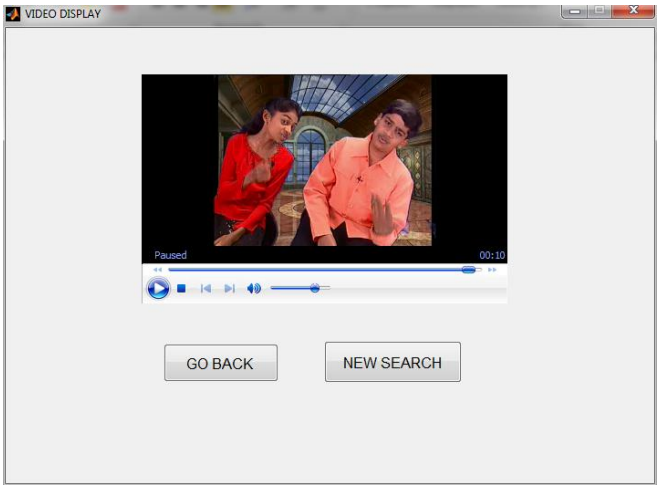

Fig.3. Video display screen

In fig. 2 the screen shot of the implemented system is depicted. In the textbox user can type the musical instrument they are interested in and press the *GO* button. This will internally call the *Call Back* function to handle the event. After

executing SNMF, Onset functions, the result will be displayed on the user interactive list box. For example, if we type guitar in the text box, it'll search the content of the video files in the video store, and list the entire video files list which contains guitar music in it. By double clicking the filename we can view the video song as shown in Fig 3. From that screen we can go back to previous screen or perform a new search.
The system gives upto 92% of accuracy for guitar, 90 % for drums and 89% for flute. For guitar it gives highest accuracy. The accuracies of three instruments are given in Fig.4.

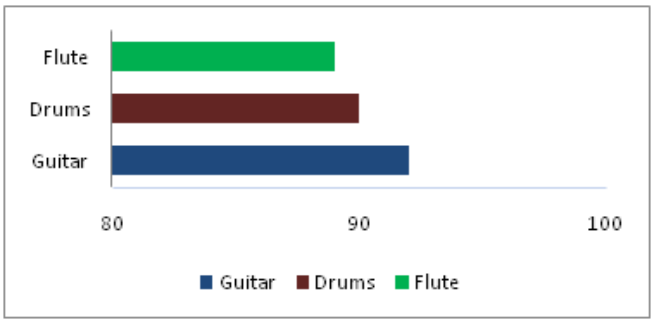

Fig.4. Result Accuracy of three Instruments

The original audio signal from one sample video file is given in Fig.5. In Fig.6 extracted guitar signal is depicted. Similarly we can extract Drums and Flute signals also.

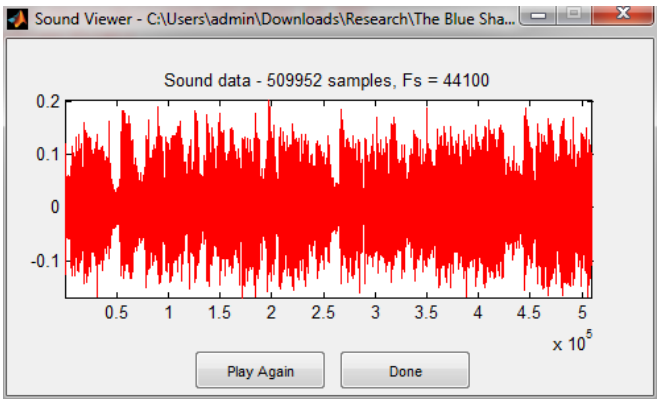

Fig.5. Original audio signal of a sample video file

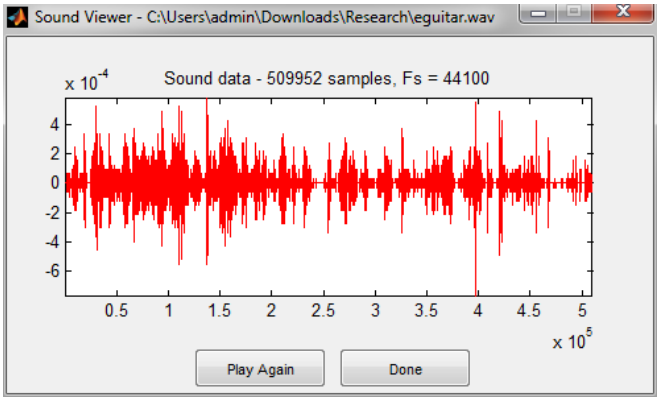

Fig.6. Seperated guitar signal from original signal

5 Conclusion

We have presented a novel and efficient approach for content based searching in a video file. The current system is capable of processing the content of video songs for 10 seconds duration and gives maximum accuracy of 92% for Guitar, 90% and 89% for Drums and Flute respectively. Duration of the song can be extended. In this paper we have identified Guitar, Flute and Drums. In future work we are planning to cover more musical instruments.This system is implemented as a window based application and also the dataset size is limited. It can be converted to web based and the dataset size also can be increased. We have used SNMF algorithm for source separation, and CSD method for calculating onset values. The same application can be performed with any other source separation algorithms like Independent Component Analysis (ICA) andOnset can also be calculated using other methods like Phase Deviation, Rectified Complex Domain etc.Extending these techniques to videos seems to be a promising and interesting direction for further research.

References

[1] Bertin, N. ; CNRS LTCI, TELECOM ParisTech, Paris ; Fevotte, C. ; Badeau, R.,"A tempering approach for Itakura-Saito non-negative matrix factorization. With application to music transcription" in Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference

[2] Barabasa, C. ; Jafari, M. ; Plumbley, M.D,"A robust method for S1/S2 heart sounds detection without ecg reference based on music beat tracking" in Electronics and Telecommunications (ISETC), 2012 IEEE Conference Publications

[3] Menaka Rajapakse and Lnnce Wyse, "NMF vs ICA for Face Recognition", Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis,2003

[4] F. Cong, Z. Zhang, I. Kalyakin, T. Huttunen-Scott, H. Lyytinen, and T. Ristaniemi, "Non-negative Matrix Factorization Vs. FastICA on Mismatch Negativity of Children", Proceedings of International Joint Conference on Neural Networks, June 2009.

[5] Daniel D. Lee and H. Sebastian Seung, "Algorithms for Nonnegative Matrix Factorization", Neural Inf. Process. Syst, 2001.

[6] Mona Nandakumar M,"An Experimental Survey on Non-Negative Matrix Factorization for Single Channel Blind Source Separation", International Journal of Computer Applications, August 2014

[7] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B.Sandler, "A Tutorial on Onset Detection in Music Signals" *IEEE Trans.Speech Audio Process*, vol. 13, no. 5, pt. 2, pp. 1035–1047, Sep. 2005.

[8] May Thu Myint,"Audio Indexing System based Myanmar Ethnic Music Signals using Simple Search Algorithm" in International Conference on Advances

in Engineering and Technology (ICAET'2014) March 29-30, 2014 Singapore.

[9] D.M.Chandwadkar, Dr. M.S.Sutaone "Role of Features and Classifiers on Accuracy of Identification of Musical Instruments" in Proceedings of CISP 2012.

[10] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), Oct. 2003, pp. 177–180.

[11] Lance Alcabasa, Nelson Marcos,"Automatic Guitar Music Transcription",International Conference on Advanced Computer Science Applications and Technologies, 2012

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

[13] Defining Thresholds:https://www.webnms.com/webnms/help/administrator_guide/performance/thresholds/perf_thresholdsintro.html