Music Videos Miner

Lalitha Agnihotri[1], Nevenka Dimitrova[2], John Kender[1], John Zimmerman[3]

[1]Dept. of Computer Science
Columbia University
New York, NY 10027

[2]Philips Research
Briarcliff Manor, NY
+1 914 945 6059

[3]Human-Computer Interaction Inst.
Carnegie Mellon University
Pittsburgh, PA 15213

lalitha@cs.columbia.edu , nevenka.dimitrova@philips.com , jrk@cs.columbia.edu, johnz@cs.cmu.edu

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]. Image/video retrieval

General Terms: Algorithms, Human Factors, Video Analysis.

Keywords: Music video summarization, multimedia content analysis, user needs analysis, chorus detection, music databases.

1. INTRODUCTION

In this demonstration we present a music video summarization system. Music videos are becoming available via video on demand services in addition to the traditional broadcast channels such as VH1, MTV and a variety of local channels. The system performs video segmentation and feature extraction to create an abstract representation of video. The abstracted video can be stored and used in a variety of applications, such as a personalized music channel, party list creator, etc. Content analysis methods exist that aim at providing high level access to specific parts of the program (e.g. highlights). Video summarization methods have been developed for news, sports, movies, etc. While music content analysis is an active area of research [2][4], music videos analysis and summarization has been neglected amongst the existing work. We are presenting a system for browsing and searching music videos based on song summaries. Access to the individual song summaries is provided via a Web-based interface.

2. USER NEEDS ANALYSIS

In order to ascertain the situational utility of music videos summaries, we performed a user-needs analysis. Our test group consisted of 20 people. For the question on whether viewers need music summarization, people liked the idea of watching what they want, when the want, where they want. Viewer-ship of music videos moderated by excessive talk, non-music related shows, or ill-matched choices. Access to a music videos library instead of watching music video channels was welcomed. The summary should include i) title, artist, album, year of the song; ii) chorus of the song for the audio summary; iii) a shot of the performing artist or the band.

3. MUSIC VIDEOS SUMMARIZATION

We used the output from the user needs analysis study to design and implement a music summarization and browsing system.

MM'03, November 2-8, 2003, Berkeley, California, USA.
ACM 1-58113-722-2/03/0011..

3.1 System overview

We present in Figure 1 the overall system architecture of the video music summarization. We assume that the system is receiving a video feed either from a broadcast/cable/satellite source, Internet streaming, or from a file stored in a video library. Also, we assume that connection to the Web is available in order to access song information such as title, artist, genre, and lyrics.

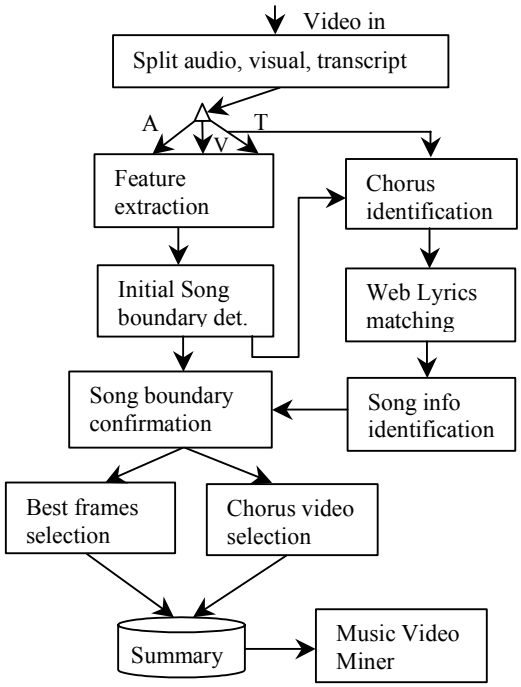

Figure 1. Overview of the music summarization system.

Music video summarization is based on identification and summarization of individual songs. Multimedia elements that are included in the summary the best representative frames include close-ups from the artist or the title image with the song information. The most representative video, audio, and text segment is the chorus of the song.

3.2 Boundary Detection

There are two types of boundary problems present in music video summarization. The first problem is to detect the song boundary. The second one is to detect the boundary of the chorus.

We use audio, visual and transcript features. Visual features include: presence of videotext [3], face presence, abrupt cuts,

color histograms [1]. In order to determine the breaks we use the approximate boundaries from all the different features: videotext, superhistograms, audio, and transcript. Then we use the fact that the transcript starts later than the visual and audio. From visual point of view we also get the videotext title page which normally appears after the start of the song and before the end of the song. The boundary is aligned with the visual color boundaries and the start (or end) of music classification in the audio domain.

3.3 Chorus Detection

Different sites (e.g. amazon.com) on the Web that sell CDs, offer samples for viewers to hear before deciding to buy music. Almost all of them include the chorus of the song. The chorus of a song is ideal to be presented in a summary. We use transcripts generated using closed captions in order to find the chorus of the song. Chorus contains the lyrics in a song that are repeated most often. We can identify the chorus segments by detecting and clustering repetitive phrases We generate time stamped closed captions in order to link back the identified chorus to the video. For the summarization, we choose the chorus that is of medium length.

4. USER INTEFACE & APPLICATIONS

The summaries can be used either in a browsing mode or in a searching mode. In the browsing mode a sample of a program summary is shown in Figure 2. Program level summary consists of the list of songs that were shown. At the next level, title, artist information, transcript from the chorus and selected video frame summarizes each song. The frames are hyper linked to audio and video of the chorus. The users can see if they are interested in certain songs and listen to a short clip before deciding to view it. This kind of interface gives a quick overview of a music videos program and makes the video transparent to the users.

Figure 2 Interactive Music Video Summaries Interface

4.1 Music Video Miner

We have developed Music Video Miner Web application for searching music videos. Using this interface, people can interactively search for music videos in the database. All the information generated for the summary is stored in an SQL database. Users can search based on a number of attributes of the summary: name of an artist or group, the title of a song, or genre. Based on the search, the system brings back the summary of songs that meet the criteria. The users have an option to play the summary and, if they like it, they have the ability to download the entire of the song. The result of a search on Shania Twain is shown in Figure 3. The summary includes the image from video that contains the title text; name of the song along with the duration; name of the artist; track and CD information; text from the chorus of the song.

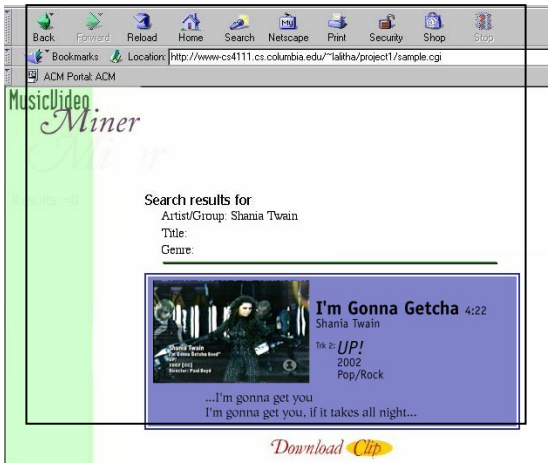

Figure 3 Music Video Miner Result Screen.

4.2 Scenarios

There are different usage scenarios for this application. The most compelling application that came up frequently in the user needs analysis was for preparing a play list for a party. Also, Music Video Miner can help in creating services for music videos on demand as well as making music purchases. Another scenario is to use the Music Video Miner coupled with automatic audio/video recommenders. Automatic recommender systems can use the information in the summary for clustering the music videos and selecting songs to compile a playlist and recommending new music to the user. We envision many other applications such as music visualization, copyright infringement detection, tracking of content distribution recording user behavior and others.

5. REFERENCES

[1] L. Agnihotri and N. Dimitrova, "Video Clustering using superhistograms in large video archives," Visual 2000.

[2] M. A. Bartsch, Gregory H. Wakefield, "To Catch a Chorus: Using Chroma-based Representations for Audio Thumbnailing," IEEE WASPAA, Oct. 2001.

[3] N. Dimitrova, L. Agnihotri, C. Dorai, R Bolle, "MPEG-7 VideoText Description Scheme for Superimposed Text," International Signal Processing and Image Communications Journal, Sept. 2000.

[4] D. Li, I. K. Sethi, N. Dimitrova, McGee, "Classification of general audio data for content-based retrieval," Pattern Recognition Letters 2000.

[5] B. Logan and S. Chu, "Music summarization using keyphrases," in International Conference on Acoustics, Speech and Signal Processing 2000.