Approach for Retrieving and Mining Video Clips

Dr. Ala'a H. AL-Hamami
Amman Arab University/ Jordan
Email: alaa_hamami@yahoo.com
Dr. Soukaena Hassan
Computer Science Department, University of Technology/Baghdad
Email: soukaena_hassan@yahoo.com
Dr. Mazin Samer AL-Hakeem
Computer Science Department, University of Technology/Baghdad
Email: mazin_ictc@yahoo.com

Received on: 18/10/ 2011&Accepted on: 5/4/ 2012

ABSTRACT

Multimedia (include video, images, audio and text media) is characterized by its high dimensionality, which makes information retrieval and data mining even more challenging. This research proposesa method to build an indexes database for huge collection of video clips, to make the video retrieval and mining much more efficient and perfectthat by considering similarity in both text of sound and features of frames. The proposed method has the following steps: **First,** isolates video motion from sound in the video clips. **Second**, converts the sound to text and index the result with database. **Third** converts video motion to shots, then select the master frame for each one and extracts the feature vector for them such as color, texture, shape and others and finally index the result with database. **Fourth,**combines the two resulted indexed database (Second and Third steps)into one database and make it the final and standard for both retrieval and mining.

Keywords: Video Clips, Mining, Indexed Database.

طريقة لاسترجاع وتنقيب مقاطع الفيديو

الخلاصة

تتميزالوسائط المتعددة (تشمل الفيديو والصور والصوت والنص) بأنها ذات أبعاد عالية، الأمر الذي يجعل من استرجاع المعلومات والتنقيب عن البيانات أكثر صعوبة. في هذا البحث تم اقتراح طريقة لبناء قاعدة بيانات لمجموعة ضخمة من مقاطع الفيديو، بهدف جعل استرجاع مقاطع الفيديو والتنقيبعنها أكثر كفاءة من خلال النظر في التشابه في كل من نص الصوت وخصائص الإطارات. الطريقة المقترحة تتضمن الخطوات التالية: أولا،تعزل حركة الفيديو عن الصوت في مقطع الفيديو.الخطوة الثانية، تحويل الصوت الى نص وتأشير النتيجة مع قاعدة البيانات. الخطوة ثالث،تحويل الحركة إلى لقطات فيديو، ثم تحديد الإطار الرئيسي لكل واحده واستخراج المميزات والخصائص بالنسبة لهم، كاللون والشكل والملمس وغيرها من الخصائص، وأخيراًتأشير النتيجة مع قاعدة البيانات. الخطوة الرابعة، دمج قاعدة

البيانات المفهرسة (في الخطوات الثانية والثالثة) في قاعدة بيانات واحدة وجعلها المعيار النهائي لكل من الاسترجاع والتنقيب.

INTRODUCTION

Currently text-based search engines are commercially available, and they are predominant in the *World Wide Web* for search and retrieval of information. However, demand for search and mining multimedia data based on its content description is growing. Search and retrieval of contents is no longer restricted to traditional database retrieval applications. As an example, it is often required to find a video clip of a certain event in a television studio. In the future the content customers will demand to search and retrieve video clips based on content description in different forms. It is not difficult to imagine that one may want to mine and download the images or video clips containing the presence of Mother Teresa from the Internet or search and retrieve them from a video archival system. It is even possible to demand for retrieval of a video which contains a tune of a particular song. In order to meet the demands for retrieval of audio-visual contents, there is a need for efficient solution to search, identify and filter various types of audio-visual content of interest to the user using non-text based technologies [8].

RELATED WORKS

The closely related works to our work are addressing as the following:

- Oh J. and Bandi B., [8],retrieve and mine the video clips depending on MPEG, concentrate with the metadata only which has at most the general information of thefeature and keywords of sound in the related video clip.
  JPEG2000 is the new standard for still picture compression and has been developed in such a way that metadata information can be also stored in the file header for access and retrieval by users as well.
  All these developments still influence effective mining of video data.
- Minami K., [3], Pfeiffer S., [4], Tonomura Y [5], Saraceno C.[6],all of them retrieve and mine the video clips depending on the words contained in the sound of the clips only without any care with the content of pictures of the video clips.

However, our work is addressingto make the video retrieval and mining much more efficient and perfect that by considering similarity in both text of sound and features of frames together.

THE PROPOSED SYSTEM

Content-based image retrieval techniques can be extended, in principle, to video retrieval systems. However, this is not verystraightforward because of the temporal relationship of video frames and their inherent structure. A video isnot only asequence of pictures, it represents the actions and events in a chronological order to convey a story and represent moving visual information [1, 2]. For that the following algorithm has been proposed for indexing video clips in a database to be suitable and efficient for retrieval and mining.

Input:Huge no. of video clips.
Output:Indexes database forall the video clips which provide efficient retrieval and mining.
Step one:Take the next video clip.
Step two:Separate the video motion from the sound along of the clip: this made easy by recording the video motion when muting the sound in the video clip and then record the sound only without the video motion, so we will have the video motion separated from video sound.
Step three:Take the sound of the clip then do the separation of music, voice, silence and noise: by using frequency spectrum, relative loudness and other parameters of traditional sound analysis, it is possible to filter the sound track into music, voice. By using similar techniques, it is even possible to segment the music stream into different beats. These are currently available as either commercial products or research prototypes [3, 4, 5,6].
Step four:Take the separated voice in the previous step, thentranslates the voice to text, by applying speech recognition techniques on the sound, these are currently available as either commercial products or research prototypes [6, 7] (in this paper we use Audioomine System [7] as speech recognition technique). After extracting the words keep them to index the video with its corresponding words which composed the sound see figure (1) and figure (2).

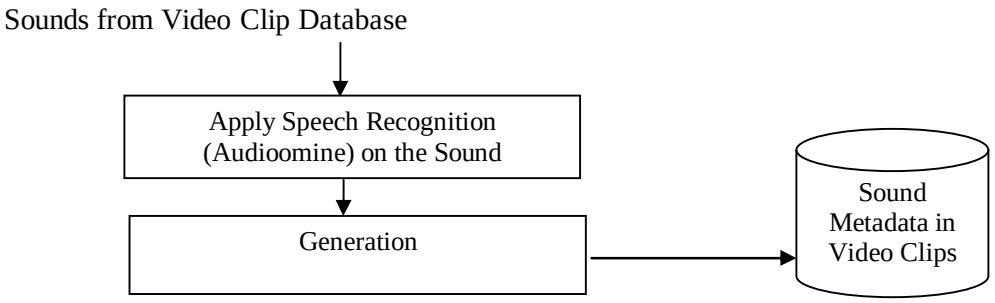

Figure (1) Build Indexes Database for the Sounds.

Video Clip ID	Words Corresponding its Sound
V1	Every night in my dream I see
V2	……
V3	……
V4	……
.	……
.	

Figure (2) The Indexes Database which Contains VideoClip and its Corresponding Words.

Step five:After separating the video motion from sound, then take videomotion and first temporarily segmented into video shots. A shot is a piece of a video motion (a group of frames or pictures) where the video content from one frame to the adjacent frames does not change abruptly. One of these frames in a shot is considered to be a master frame***(the selection of master frame will be depend on no. of objectsso the master frame of the shot will has the maximum no. of objects)***. This master frame is considered to be a representative for the picture content in that shot. Sequence of master frames can define the sequence of events happening in the video clip. This is very useful to identify the type and content of the video, see figure (3).

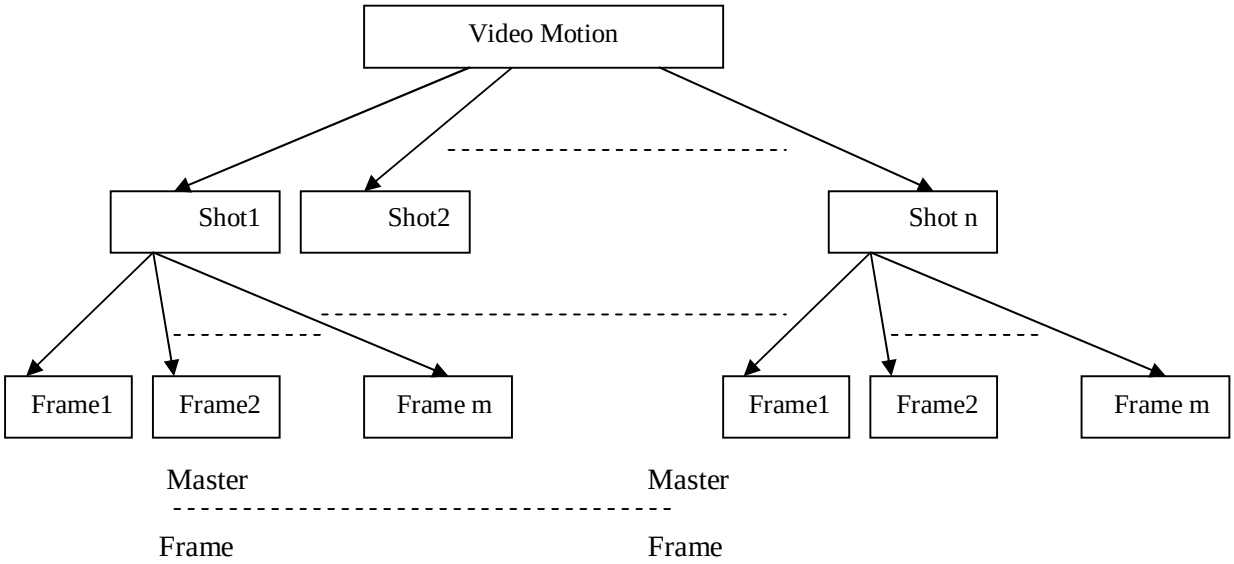

Figure (3) Video Segmentation to Shot, then Frames and Detecting the Master Frame.

Step six:Now collect the master frames of each video clip, since master frames are images, then we will propose that: the images in an image database are indexed-based on extracted inherent visual contents (or features) such as color features (color histogram, color coherence vector, color moment, linguistic color tag), texture, image shape and topology. The feature vector actually acts as the *signature* of the image, and extraction of these features depend on content based method, [9], figure (4) representsthe main architecture to build indexes database for the master frames of the video clips in each shot. Figure (5) represent the indexes database for each video clip and the corresponding master frames of that clip.

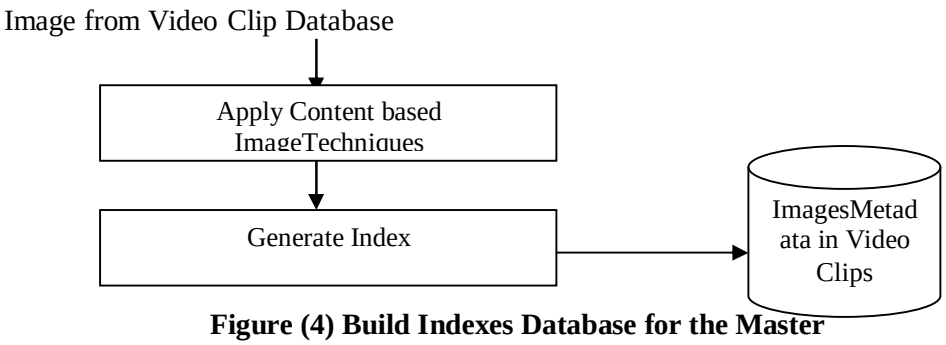

Figure (4) Build Indexes Database for the Master Frames of the Video Clips

Video ID	Master frame ID	Color texture shape topologyhisto mom coher tag
V1	I1	116.84 122.84 117.84 116.86 0.87 0.58 0.61
	I2	……
	I3	……
V2	I1	……
	I2	……
V3	.	
V4	.	

Figure (5) The Indexes Database for each Video Clip and the Corresponding Master Frames.

Step seven:Now combine the two indexes databases (the sound and image) into one comprehensive database, see figure (6).

Video ID	Words corresponding it is sound	master frame ID	color text. shape topologyhisto mom coher tag
V1	Every night in my dream I see	I1	116.84 122.8 117.84116.86 0.87
	……	I2	0.580.61
	……	I3	……
V2	……	I1	……
	……	I2	……
etc	……	.	……
	……		……
	…		……
			……
			……

Figure (6) The Final Video Clip Indexes Database.

Step eight: After completing the indexes database for video clips, now the retrieval and mining process is depicted. The query video clip is analyzed to extract the visual features (according to the features depended in the proposed indexes database) and analyze the sound to text, then these features and text are used to retrieve and mine the similar video clips from the comprehensive video clips database. Rather than directly comparing two video clips, similarity of the visual features of the query video clip is measured with the features of each video clip stored in the comprehensive database as their signatures. Often the similarity of two video clips is measured by computing the distance between the feature vectors of the two video clips. The retrieval and mining system returns the first *k* video clips, whose distance from the query video clips is below pre-defined threshold, see figure (7).

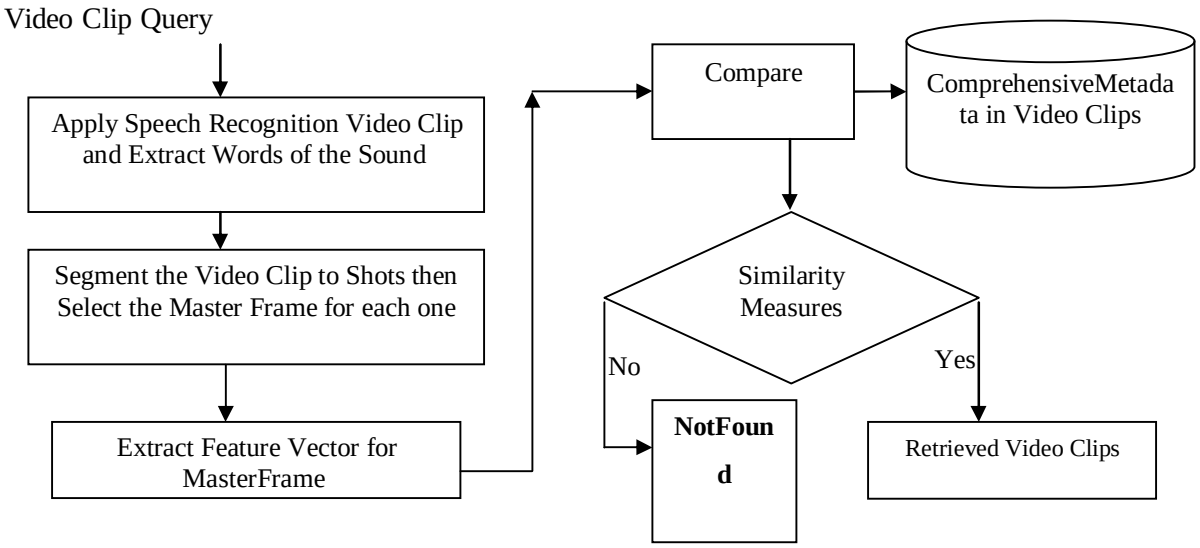

Figure (7) Retrieve or Mine the Comprehensive Database According to theQuery.

The Implementation of theProposed System

The implementation of the proposed system has two sub programs the first one will be used if the user has the clip of retrieval and the second will be used if the user has only information about the clip he wants to retrieve. ***The first sub program*** will be as the following: the initial main window of the implementation will be shown in figure (8).

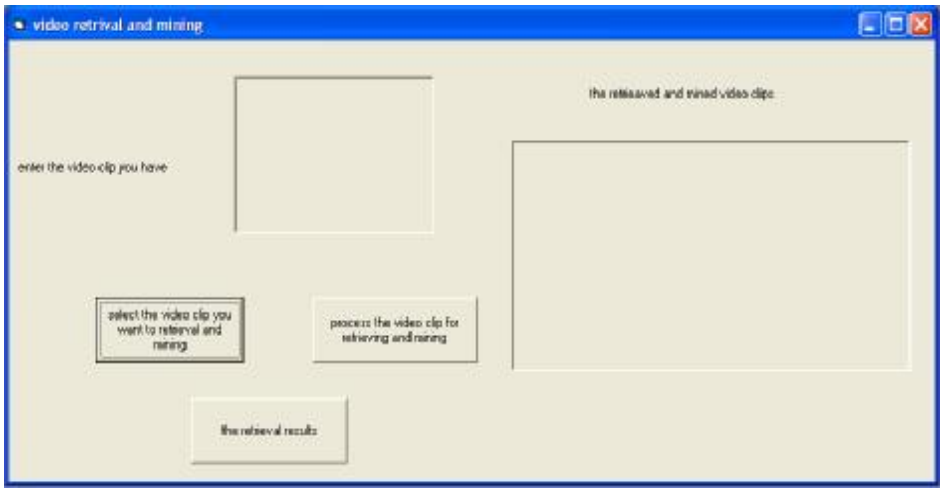

Figure (8) The Initial Main Window.

Figure (8) has three commands, when the first one is clicked then figure (9) will appear to enable the user for loading the desired video clip, see figure (9).

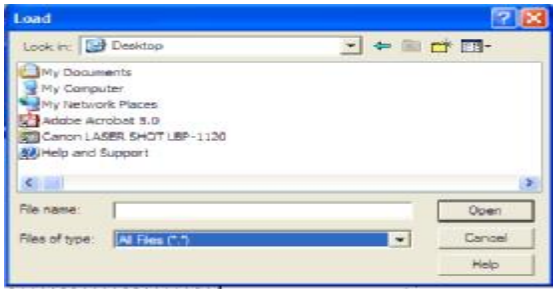

Figure (9) the Window Responsible to Loading the Query Represented by Video Clip

After loading the desired video clip by the user, the user must click the second command to analyze the desired video clip and present the results of that clip, as shown in figure (10).

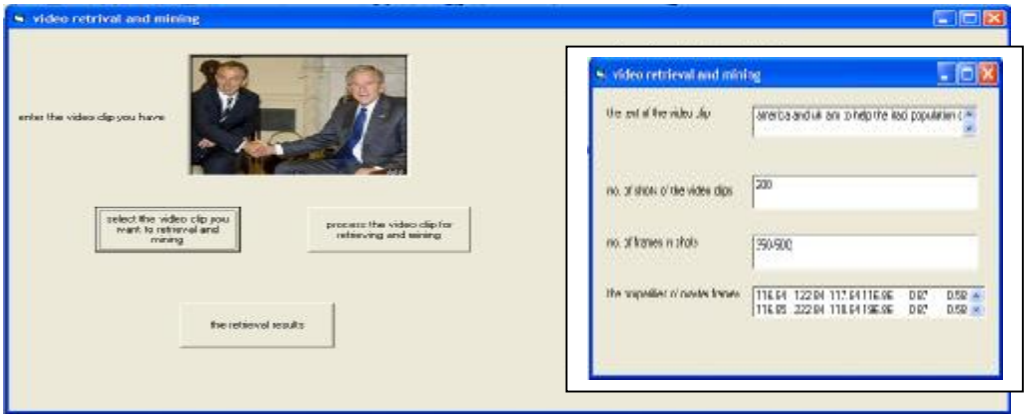

Figure (10) The*Main Window* withWindowwhich Display the Results of the Analyzed Clip

The last step is to click the third command to take the results of the analyzed clip and compare them to all analyzed clips in the comprehensive databases to display the similar clips, see figure (11).

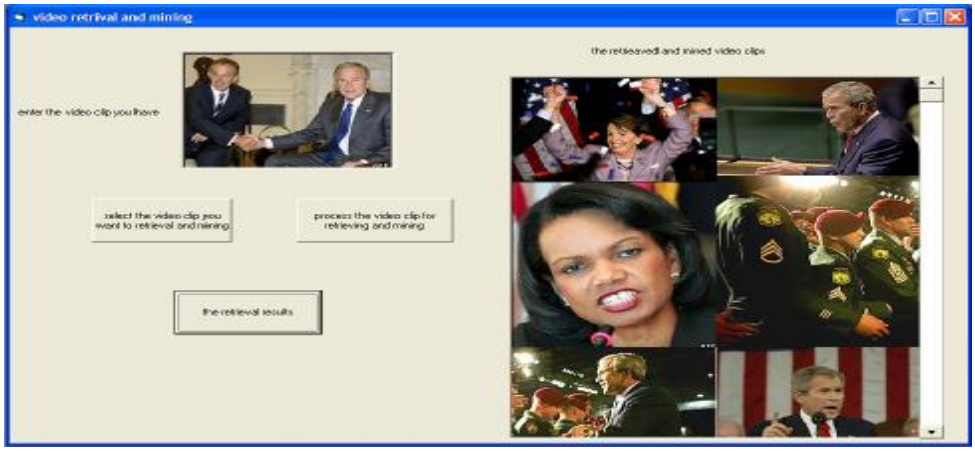

Figure (11):The Final Appearance of the Main Window which have the Desired Clip and all the Similar Clips

The second sub program will be as the following: the main window of the implementation will be shown in figure (12). In this figure the user will enter the keyword of the sound of the video clips he wantsin the first text box and in the second text box will enter all the Meta data he knows about the desired clips. Finally, by pressing the command below both text boxes then the mined and the retrieved clips will be displayed.

Figure (12) The MainWindow which has the Entered Sound Keywords and Metadata for the Wanted Clip and all the Results of Similar Clips

CONCLUSIONS

The following items represent the important conclusion which is drawing through the development of our proposed system.

1. Consider both sound and video motion of video clips in building the indexed database of video clips make the retrieval and mining of the video clips much more efficientthan using metadata of video only or depending on the words contained in the sound of the video clips only as shown in the work [3, 4, 5, 6] which care with the words of the sound only and cancel the features of pictures which represent crucial point in retrieve and mine the video clips; and the work [8] which depend in it is retrieve only on the general features and keywords mentioned in metadata and that surely un sufficient to optimize the mining.
2. Segmenting the video motion in to shots then each shot contains no. of frame and finally select a master frame for each shot by taking the one has maximum no. of objects.This will give optimization for the selection since that the selected frame will be ensured that has all the objects in the shot.
3. Taking the color features (color histogram, color coherence vector, color moment, linguistic color tag), texture, image shape and topology will give the optimized signature of the frame, that will make the retrieve and mine process perfect.
4. Build a comprehensive database will unify the search process with one database for both sound and video motion.

REFERENCES

[1].Kantardzic M.; ***"DM Concepts, Models, Methods and Algorithms",*** jhon wiley &Sons, 2003.

[2].Sakurai S., Ichimura Y., Suyama A., and Orihara R.; ***"Inductive Learning of a Knowledge Dictionary for a Text Mining System,"*** in Proceedings of 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, 2003.

[3].Minami K., Akutsu A., and H. Hamada, "***Video Handling with Music and Speech Detection"***, IEEE multimedia, page 17-25, July-September 1998.

[4].Pfeiffer S., Fischer S., and Effelsberg W., ***"Automatic Audio Content Analysis"***, In proc of ACM multimedia, pages 21-30, ACM press, New Yourk, 1996.

[5].Tonomura Y., Akutsu A., Taniguchi Y., and Suzuki G., ***"Structured Video Computing",*** IEEE multimedia, pages 34-43, fall 1994.

[6].SaracenoC. and Leonardi R., ***"Audio as a Support to Scene Change Detection and Characterization of Video Sequence"***, in proceeding of the IGASSP, IEEE computer society press, 1997.

[7].Audioomine by dragon system inc. Available at http://dragonsys.com.

[8]. Bandi B., Oh J., ***"Multimedia Data Mining Framework for Raw Video Sequences"***, Proceedings Third International Workshop on Multimedia Data Mining MDM/KDD'2002, July 23rd 2002, Edmonton, Alberta, Canada.

[9]. Vailaya A., Figueiredo A. T., Jain A. K., and Zhang H. J., ***"Image Classification for Content-Based Indexing"***, *IEEE Transactions on Image Processing*, Volume: 10 Issue: 1, pp 117 –130, Jan. 2001.