

AUDIO-VISUAL CONTENT-BASED MULTIMEDIA INDEXING AND RETRIEVAL – THE MUVIS FRAMEWORK

Moncef Gabbouj and Serkan Kiranyaz

Institute of Signal Processing
Tampere University of Technology
Tampere, Finland
moncef.gabbouj@tut.fi

ABSTRACT

MUVIS is a collaborative framework that supports indexing, browsing and querying of various multimedia types such as audio, video, audio/video interlaced in several formats. It allows real-time audio and video capturing, encoding by last generation codecs such as MPEG-4, H.263+, MP3 and AAC. MUVIS also supports several audio/video file format such as AVI, MP4, MP3 and AAC. MUVIS achieves a global and unified solution for content-based indexing and retrieval problem and provides user-friendly applications and a generic framework especially for third parties to develop their feature extraction modules. In this paper, we present an overview of the MUVIS system and we shall especially focus on the overall audio-based multimedia indexing and retrieval scheme within MUVIS framework.

1. INTRODUCTION

The growth in the size of available multimedia both audio and visual requires proper management, indexing and retrieval schemes. In order to overcome such problems several content-based indexing and retrieval techniques and applications have been developed such as MUVIS system [1], [2], [13], Photobook [3], VisualSeek [4], Virage [5], VideoQ [6] and VideoAL [15]. The common feature of all such systems is that they all provide some kind of framework and several techniques for indexing and retrieving either still images or audio-video files. MPEG-7 [7] is a recent standard for multimedia content description.

We have recently developed a PC-based MUVIS system, which is further capable of content-based indexing and retrieval of video and audio information in addition to several image types. Table 1 shows the types of multimedia that the MUVIS system so far supports.

MUVIS Audio			
Codecs	Sampling Freq.	Channel No	File Formats
MP3 [11]	16, 22.050, 24 KHz	Mono	MP3
AAC [12]	32, 44.1 KHz	Stereo	AAC
G721	Any for G721,		AVI
G723	G723 & PCM		MP4
PCM			

MUVIS Video			
Codecs	Frame Rate	Frame Size	File Formats
H263+ [10]	1 - 25 fps	Any	AVI
MPEG-4 [9]			MP4
YUV 4:2:0 RGB 24			

MUVIS Image Types							
Convertible Formats							
	JPEG	JPEG 2K	BMP	TIFF	PNG		
Inconvertible Formats							
PCX	GIF	PCT	TGA	PCX	EPS	WMF	PGM

Table 1: MUVIS Multimedia Family

The current version of the MUVIS framework supports the following multimedia processing capabilities and features:

- Real-time audio and video capturing, encoding and recording,
- Hierarchic video handling and representation,
- Video summarization via scene detection [8],
- An effective framework structure, which provides an application independent basis in order to develop audio and visual feature extraction techniques that are used dynamically by MUVIS applications for indexing and retrieval.
- The retrieval based on distinct visual and aural queries initiated from any MUVIS database that includes audio/video clips and still images.
- Conversion of alien formats to those supported in MUVIS,

The rest of this paper is organized as follows: in section 2, we outline the system philosophy of MUVIS and the general MUVIS framework with underlying applications. Section 3 presents the overall audio-based indexing and retrieval scheme in the MUVIS framework. In section 4, we demonstrate some experimental results on audio-based multimedia retrieval via query and conclude the paper.

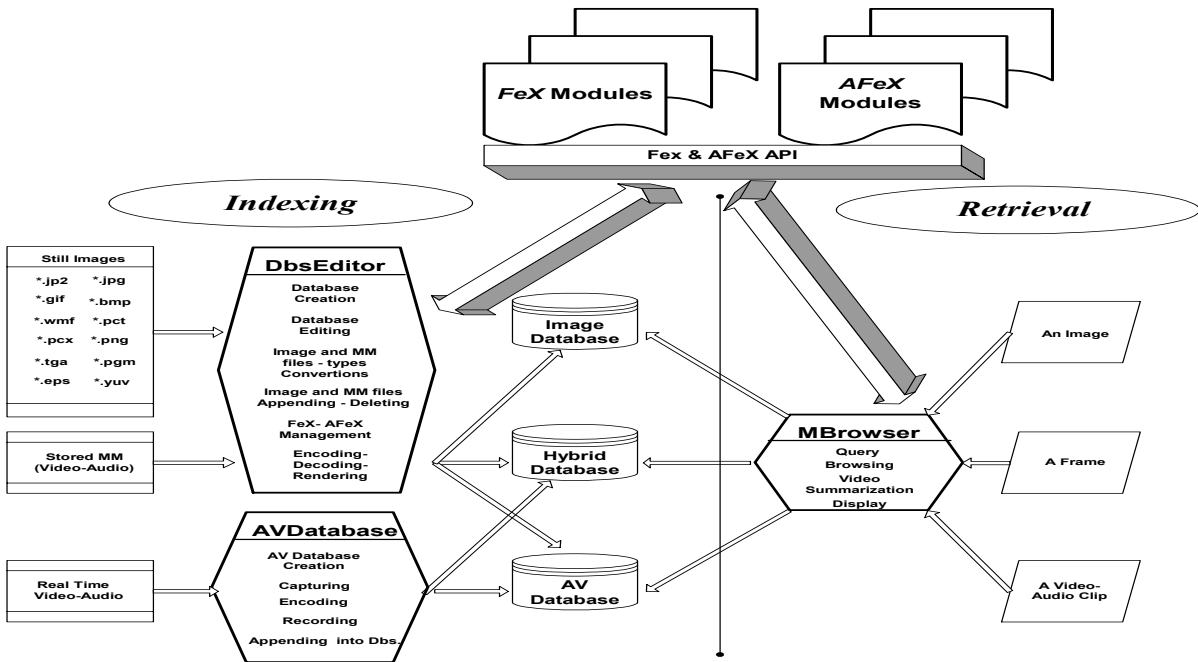


Figure 1: General structure of MUVIS framework

2. APPLICATIONS OF MUVIS FRAMEWORK

As shown in Figure 1, MUVIS framework is based upon three main applications, each of which has different responsibilities and functionalities. *AVDatabase* is mainly responsible for video and/or audio database creation. Several video and audio encoding techniques can be used with any encoding parameters specified in Table 1. More detailed information about *AVDatabase* application can be found in [2].

DbsEditor performs the general tasks of indexing of multimedia databases, and therefore, offline feature extraction processing is its main task. The main functionalities of *DbsEditor* can be listed as follows:

- Dynamic integration and management of feature extraction modules.
- Extracting new features or removing existing features of a database using available *FeX* and *AFeX* modules,
- Appending/Removing any audio/video clips and still images to/from any existing database,
- Converting one image format to any convertible image format given in Table 1.
- Appending alien audio/video files into any MUVIS database.
- Preview of any audio/video clip or image in a database.

MBrowser is the main media browser and retrieval terminal, which works with any kind of MUVIS database. In the most basic form, it has the capabilities of an advanced multimedia player and a simple database browser. Furthermore, it allows

users to reach any kind of multimedia easily, efficiently and, for the case of video clips, in any of the available hierarchic levels. *MBrowser* supports a 4-level video display hierarchy: single frame, shot frames (key-frames), scene frames and the entire video clip.

MBrowser has a built-in search and query engine, which is capable of finding multimedia primitives in a database and for any multimedia type that is similar to a queried media item (a video clip, a frame or an image). Retrieval is based on comparing the similarity distances between the queried media item's feature vector(s) with the feature vectors of multimedia primitives available in the database. The similarity distances are calculated by the corresponding functions implemented in the corresponding feature extraction module.

Furthermore, *MBrowser* provides the following additional functionalities:

- Video summarization via scene detection,
- Key-frame browsing during video playback,
- Random access support for video playback,
- Displaying information related to the extracted features of the active database,
- Visualizations of feature vectors of the multimedia items.

3. AUDIO INDEXING AND RETRIEVAL FRAMEWORK IN MUVIS

Audio is an important source of information for content-based multimedia indexing and retrieval. Furthermore, it can sometimes be even more important than the visual part since it is mostly unique and significantly stable within the content. However, when dealing with digital audio there are several requirements to be fulfilled and the most important of them is

the fact that the content is totally independent from the digital audio capture parameters (i.e. sound volume, sampling frequency, etc.), audio file type (i.e. AVI, MP3, etc.), encoder type (MP3, AAC, etc.) or encoding parameters (i.e. bit-rate, etc.). So the overall structure of the audio-based indexing and retrieval framework has been developed to provide a pre-emptive robustness (independency) to such variations.

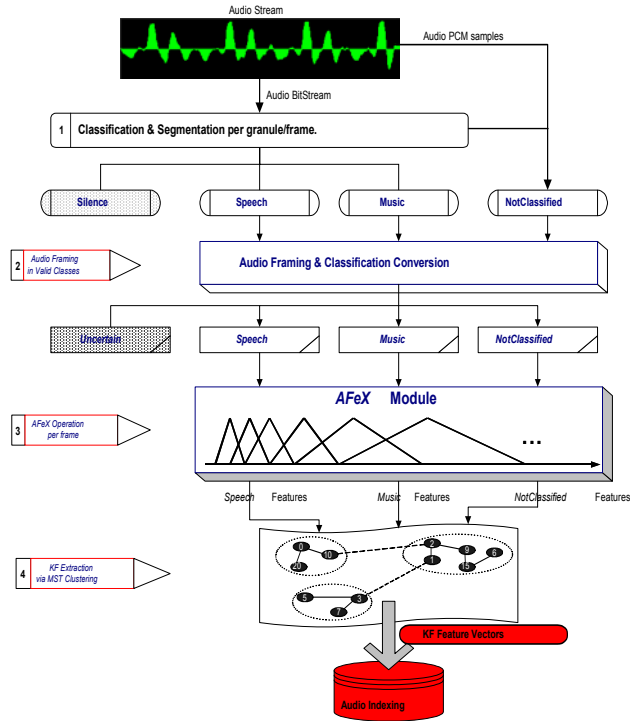


Figure 2: MUVIS Audio Indexing Operation Flowchart

As shown in Figure 2, audio indexing is accomplished in several stages: classification and segmentation using band energy ratio, pause rate, subband centroid and fundamental frequency, see Figs. 3-5; audio framing within segments with certain types, see Fig. 6; audio feature extraction via available *AFex* modules (Fig. 7 shows how this module interacts with the MUVIS applications); key-framing via minimum spanning tree (MST) clustering [14], see Fig. 8; and finally the indexing over the extracted key-frames (KFs).

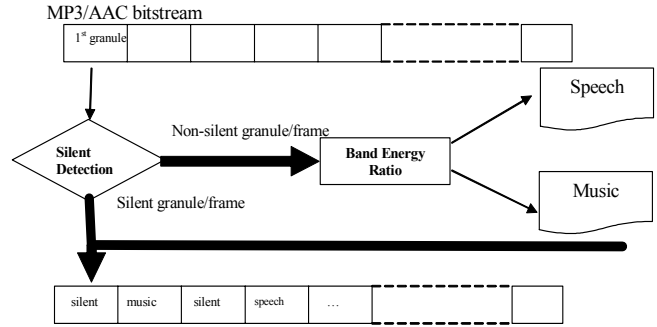


Figure 3: Feature Extraction per granule

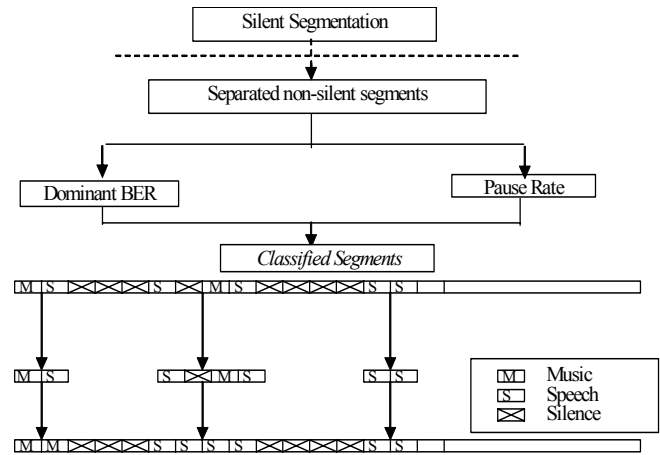


Figure 4: Segmentation and Classification per Segment

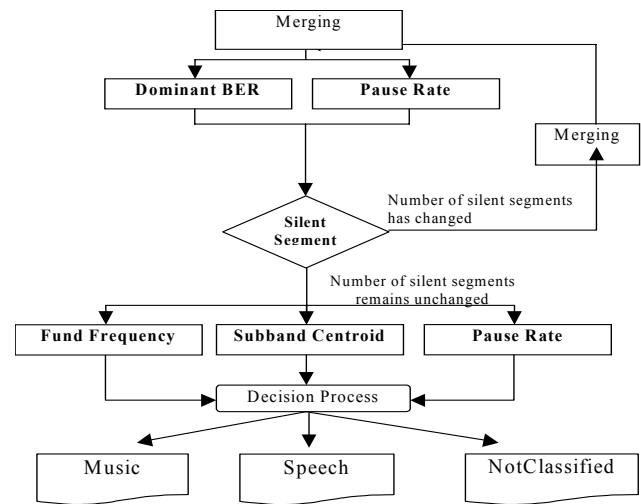


Figure 5: Merging and final classification

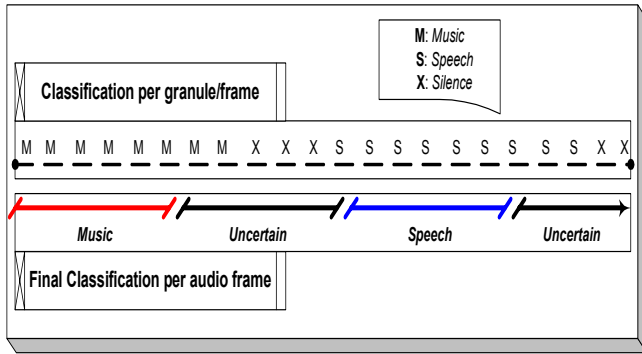


Figure 6: A sample audio classification conversion

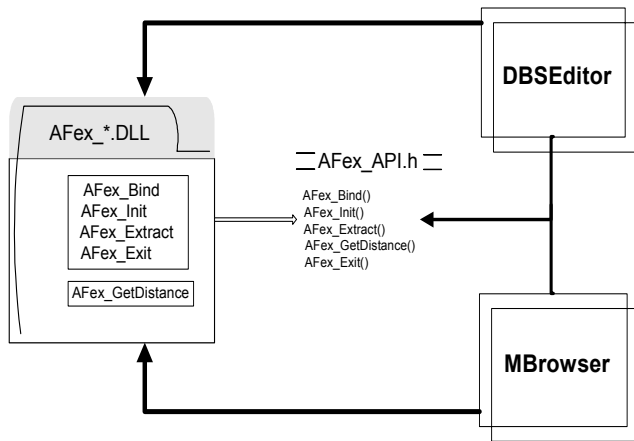


Figure 7: Basic AFeX Module interface with MUVIS applications

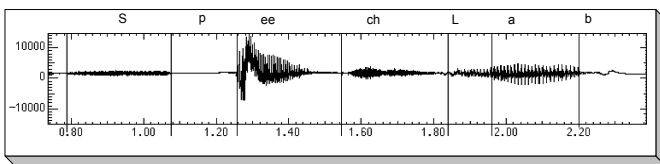


Figure 8: An illustrative MST-based clustering scheme

4. AURAL RETRIEVAL-BY-QUERY SCHEME

As explained in detail in the previous sections, the audio part of any multimedia item within a MUVIS database is indexed using one or more AFeX modules that are dynamically linked to the MUVIS application. The indexing scheme uses the audio classification per segment information to improve the effectiveness in such a way that during an audio-based query scheme, the matching (same audio class types) audio frames will be compared against each other via the similarity distance. In order to accomplish an audio based query within MUVIS, an audio clip is chosen from a multimedia database and queried through the database if the database includes at least one audio feature. Figure 9 illustrates a class based audio query. Details of the specific distance measures used can be found in [16].

5. CONCLUSIONS

Audio-based retrieval will be demonstrated in the presentation. Two particular experiments can be mentioned. In the first one, a video clip (with MP3 audio) is queried over a database containing over 800 audio and video (with audio) clips with a total duration around 28 hours. The second experiment involves a database of 181 video (with audio) clips with a total duration of around 12 hours. The capturing/encoding parameters (i.e. sampling frequency, sound volume level, bit-rate, etc.), encoder types (MP3, AAC, G721, etc.), clip duration and file formats of the clips are varying among the values given in Table 1 in both databases. The best matches in both experiments are video (with audio) and audio clips from the database with a significant relevance to the query clip.

Several retrieval experimental results show the effectiveness of the overall FeX - AFeX indexing and retrieval framework and the available FeX - AFeX modules. The audio indexing framework achieves a significant query performance and it provides a robust and generic solution for several multimedia types, capture parameters, coding methods, file formats and several other factors that MUVIS system so far supports.

MUVIS framework is designed to bring a unified and global solution to content-based multimedia indexing and retrieval problem. The unified solution is basically achieved by designing the whole system of applications, which are handling the media during its life-time starting from capturing till indexing in such a way that the media can be indexed as efficiently as possible. The framework supports browsing, hierarchic video representation and summarization. Most important of all, MUVIS framework supports integration of the aural and visual feature extraction algorithms explicitly. This brings a significant advantage for third parties to develop and test several feature extraction modules that are independent from the MUVIS applications.

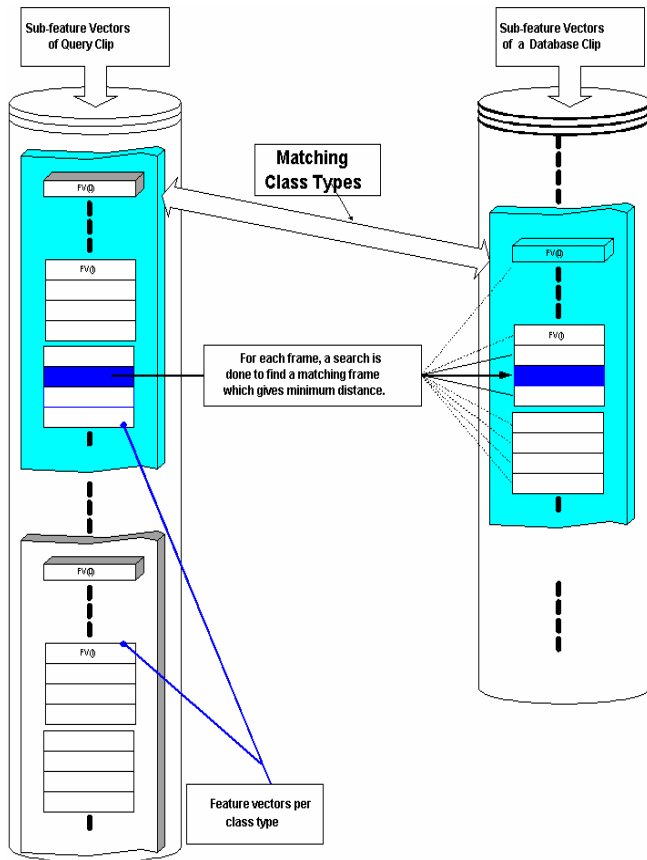


Figure 9: A class-based audio query

6. REFERENCES

- [1] M. Gabbouj, S. Kiranyaz, K. Caglar, B. Cramariuc, F. Alaya Cheikh, O. Guldogan, and E. Karaoglu, "MUVIS: A Multimedia Browsing, Indexing and Retrieval System", *Proc. of the IWDC 2002 Conference on Advanced Methods for Multimedia Signal Processing*, Capri, Italy, September 2002.
- [2] S. Kiranyaz, K. Caglar, O. Guldogan, and E. Karaoglu, "MUVIS: A Multimedia Browsing, Indexing and Retrieval Framework", *Proc. Third International Workshop on Content Based Multimedia Indexing, CBMI 2003*, Rennes, France, 22-24 September 2003.
- [3] A. Pentland, R.W. Picard, S. Sclaroff, "Photobook: tools for content based manipulation of image databases", *Proc SPIE (Storage and Retrieval for Image and Video Databases II)* 2185:34-37, 1994.
- [4] J.R. Smith and Chang, "VisualSEEk: a Fully Automated Content-Based Image Query System", *ACM Multimedia*, Boston, Nov. 1996.
- [5] Virage. [URL:www.virage.com](http://www.virage.com)
- [6] S.F. Chang, W. Chen, J. Meng, H. Sundaram and D. Zhong, "VideoQ: An Automated Content Based Video Search

System Using Visual Cues", *Proc. ACM Multimedia*, Seattle, 1997.

[7] ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-7 Standard Version 5.0", March 2001.

[8] S. Kiranyaz, K. Caglar, B. Cramariuc, and M. Gabbouj, "Unsupervised Scene Change Detection Techniques In Feature Domain Via Clustering and Elimination", *Proc. IWDC 2002 Conference on Advanced Methods for Multimedia Signal Processing*, Capri, Italy, September 2002.

[9] ISO/IEC JTC1/SC29/WG11, "Coding of Moving Pictures and Audio: Overview of MPEG-4 Standard", V. 21, March 2002.

[10] ITU-T Recommendation H.263, "Video Coding For Low Bit Rate Communication", February 1998.

[11] ISO/IEC 13818-3:1997, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information – Part3: Audio, 1997.

[12] ISO/IEC CD 14496-3 Subpart 4: 1998, Coding of Audiovisual Object Part3: Audio, 1998.

[13] F. Alaya Cheikh et. al. "MUVIS: a System for Content-Based Indexing and Retrieval in Large Image Databases", *Proc. SPIE/EI'99 Conference on Storage and Retrieval for Image and Video Databases VII, Vol.3656*, San Jose, California, 26-29 January 1999.

[14] Graham, R.L., and O. Hell, "On the History of the Minimum Spanning Tree Problem," *Ann. Hist. Comput.* 7, pp. 43-57. 1985.

[15] Ching-Yung Lin, Belle L. Tseng, Milind Naphade, Apostol Ntsev and John R. Smith, "VideoAL: A Novel End-to-End MPEG-7 Video Automatic Labeling System," *Proc. IEEE ICIP 2003*, Barcelona, Spain, 14-17 September 2003.

[16] M. Gabbouj, S. Kiranyaz, K. Caglar, E. Guldogan, O. Guldogan and F.A. Qureshi, "Audio-Based Multimedia Indexing and Retrieval Scheme in MUVIS Framework," *Proceedings of ISPACS 2004*, 8-10 December 2003, Awaji Island, Japan.