

Novel DSP Tools in CBIR MUVIS Framework

Moncef Gabbouj

Institute of Signal Processing, Tampere University of Technology

FIN-33101 Tampere, Finland

Email : moncef.gabbouj@tut.fi

Summary

MUVIS [1] is a content-based indexing and retrieval system for large multimedia databases. MUVIS is developed and maintained at Tampere University of Technology and has recently been adopted as the reference framework by EU COST 292 Action WG2. MUVIS brings a unified and generic solution to content-based multimedia indexing and retrieval since it is designed for various formats, representations and different parameters in today's digital multimedia world, such as codec types, file formats, capture and encoding parameters. By covering such a wide-range of multimedia family and especially the last generation multimedia codecs, MUVIS is developed to provide an efficient framework structure upon which robust algorithms can be implemented, tested, configured and compared against each other. Furthermore, it supports three types of browsing, five levels of hierarchic video representation and summarization and most important of all, MUVIS framework supports integration of audio and visual feature extraction algorithms explicitly. Several novel DSP algorithms and tools have been developed and used within the context of the following processing capabilities and properties of MUVIS [2-7]:

- An effective framework structure, which provides an application independent basis in order to develop audio and visual feature extraction (AFeX & FeX) techniques that are dynamically integrated to and used by the MUVIS applications for indexing and retrieval. Some of the FeX modules developed include:
 - Shape features from multi-scale edge detection and object extraction for image retrieval,
 - Histogram-based color features,
 - Texture features from gray-level co-occurrence matrix, ordinal correlation techniques and Gabor wavelet transform,
 - Mel-frequency cepstrum coefficients (MFCC) based audio features.
- Real-time audio and video capturing, encoding and recording. Particularly the encoding is performed with last generation audio & video codecs, such as:
 - H.263+ video encoder (H.264 will shortly be available)
 - MPEG-4 SP video encoder
 - MPG-2,4 Advanced Audio Codec (AAC)
 - MPEG-1,2 Layer-3 (MP3) audio encoder
 - AMR audio encoder.
- Multimedia conversions into one of the convertible formats that MUVIS supports. The following signal processing methods are applied:
 - Sampling frequency conversion for audio (Re-sampling) using a polyphase filterbank.
 - Video frame size, rate and bit-rate conversions.
- Scalable video management using:
 - Video summarization via shot-boundary detection via histogram difference and key-frame extraction.

- Scene analysis and scene frame extraction via MST clustering and NN elimination.
- A novel *Progressive Query* mechanism [2], which provides faster query results along with the query process and lets the user browse among the queries obtained and stops an ongoing query in case the results obtained so far are satisfactory. It uses database fractioning, sub-set fusing, linear time estimation and error correction.
- A novel indexing and browsing method: *Hierarchical Cellular Tree (HCT)* [6] and the implementation of *PQ* over *HCT*. *HCT* is based on the following techniques:
 - Unsupervised machine learning,
 - Dynamic (incremental) MST construction based on leaf node (vertex) pruning and branch (edge) construction.
 - Regularization of an energy function to maximize cell compactness.
- Audio content analysis and audio-based multimedia indexing and retrieval [3]. This unsupervised audio classification and segmentation technique uses several signal processing tools over raw and encoded (by MP3 and AAC) audio stream, such as:
 - Formation of a spectral template over modified DCT (MDCT) coefficients obtained only by Huffman Decoding and De-quantization.
 - Feature extraction based on:
 - Total frame energy
 - Band energy ratio
 - Fundamental frequency estimation: using modified HPS (Harmonic Product Spectrum) and adaptive peak extraction
 - Sub-band centroid
 - A novel segment feature extraction method: transition rate
 - Fuzzy perceptual modelling and generic audio classification and segmentation.

The presentation shall focus on the novel DSP algorithms and tools used in MUVIS system and show some experimental results and demos.

References

- [1] <http://muvis.cs.tut.fi/>
- [2] S. Kiranyaz and M. Gabbouj, "A Novel Multimedia Retrieval Technique: Progressive Query", in *IEE Proc. Vision, Image & Signal Processing*, to appear. (a copy can be downloaded from <http://sp.cs.tut.fi/publications/inpress.en.shtml>)
- [3] S. Kiranyaz, A.F. Qureshi, and M. Gabbouj, "A Generic Audio Classification and Segmentation Approach for Multimedia Indexing and Retrieval," *IEEE Transactions on Speech and Audio Processing*, in press. (a copy can be downloaded from <http://sp.cs.tut.fi/publications/inpress.en.shtml>)
- [4] Faouzi Alaya-Cheikh, Bogdan Cramariuc, Mari Partio, Pasi Reijonen and Moncef Gabbouj, "Ordinal Measure Based Shape Correspondence," *Journal of Applied Signal Processing*, Special issue on *Image Analysis for Multimedia Interactive Services – Part I*, Vol. 2002, No. 4, pp. 362-371, April 2002
- [5] Bogdan Cramariuc, "Image Analysis and Understanding Techniques with Application to Content Based Image Manipulation," Ph.D. Thesis, Tampere University of Technology, Tampere, December 2005.
- [6] Serkan Kiranyaz, "Advanced Techniques for Content-Based Management of Multimedia Databases," Ph.D. Thesis, Tampere University of Technology, Tampere, June 2005.
- [7] Faouzi Alaya Cheikh, "MUVIS: A System for Content-Based Image Retrieval", F. Alaya Cheikh, Ph.D. Thesis at Tampere University of Technology, Tampere, Finland, April 2004.