Redundancy Reduction Techniques and Content Analysis for Multimedia Services – the European COST 211quat Action

Moncef Gabbouj[1], Geoff Morrison[2], Faouzi Alaya-Cheikh[1], Roland Mech[3]
[1] Signal Processing Laboratory, P.O. Box 553, FIN-33101 Tampere, Finland
[2] BT Labs, Martlesham Heath, Ipswich IP5 3RE, United Kingdom
[3] University of Hannover, Appelstraße 9A, 30167 Hannover, Germany

Abstract. The general philosophy of COST projects is introduced before narrowing the focus to the COST 211 series. For more than 20 years COST 211 has been concerned with image coding and compression using redundancy removal techniques. Within COST 211ter an Analysis Model was developed for the automatic and semi-automatic segmentation of motion video and significant parts of this work have been endorsed by ISO MPEG-4. The current phase, COST 211quat, continues to improve the Analysis Model with the additional aim of advancing the state of the art in object segmentation and recognition technologies. Such content analysis technologies will be essential to make MPEG-7 and related multimedia content description standards viable for easy and widespread use.

1. COST and the COST 211quat Action

1.1 COST

Coopération Européenne dans la recherche scientifique et technique, commonly known as COST, is an open and flexible framework for cooperation in pre-competitive and basic research and development. R&D fields, called "Actions" in COST, can be proposed by researchers and the European Communities. There are currently some 130 such Actions. The EC finances management committee meetings and workshops organised by the Action, e.g. WIAMIS'97 and WIAMIS'99 organised by COST 211ter and COST 211quat, respectively. The participants, on the other hand, are responsible for funding their own research, including technical meetings.

A basic principle in COST is that participation of COST countries is voluntary and is accomplished by the COST country signing the "Memorandum of Understanding". The MoU is the legal basis of the Action. Once a country has signed it, any organisation from that country becomes eligible to participate in the technical activities of the Action. The MoU defines the scope and objectives of the Action, as well as the terms of participation including compliance with intellectual property rights.

1.2 COST 211quat

The present COST 211quat Action started in 1998 and will end in 2003. It is a follow-up on Actions COST 211, 211bis and 211ter. The main objective of the Action is to improve the efficiency of redundancy reduction techniques and to develop content analysis for video signals to assist future multimedia applications. Furthermore, it is the objective to strongly influence the standardisation activity in this field. It will, in particular, focus on content-oriented processing for emerging interactive multimedia services, such as the ongoing ISO MPEG-4 standardisation phase as well as the new ISO MPEG-7 initiative. The aim is to define and develop a set of tools assisting these new services in the analysis, characterisation and processing of their video and audio signals.

Interactive multimedia services will strongly influence and even dominate the future of communications and telecommunications. Both the flexibility and efficiency of the coding systems used, as well as the ability to efficiently search for particular content of interest on distributed data-bases are essential for the success of these emerging services. A major outcome of this Action will be a valuable contribution towards new and economic solutions for interactive multimedia services.

Redundancy reduction has been an ongoing topic from the beginning of the original Action 211 [Sikora 97]. At that time compression schemes were blind to the signals being coded, the pictures being regarded merely as rectangular arrays of numbers, albeit with some inherent properties which could be exploited. More recently, to enable higher reduction rates (for an equivalent visual quality) to be obtained and to introduce possibilities of content interactivity, attention has turned to object coding methods, in which scenes are first decomposed into appropriate arbitrarily shaped regions that are then separately coded. Each region may be compressed with an optimum method. Such a concept is at the heart of MPEG-4 to which 211ter contributed. However, the MPEG-4 standard does not cover the decomposition itself and the latter stages of 211ter focussed on this segmentation problem which is still not completely solved in 211 or elsewhere.

Allied to the above is the great interest in multimedia content description. Notably the intention of MPEG-7 to standardise a multimedia content description interface. MPEG will not standardise the method by which a description is obtained. Thus, unless purely manual methods are to be used, there is a need for algorithms which can "understand" the content of images. Hence the core issue for 211quat is content analysis and the following lists the main items identified for study during the Action:

- Audio-visual Content Identification - e.g. the automatic detection of content of interest in video scenes, such as a person, car, etc.
- Content and/or Feature Extraction - e.g. the automatic extraction of content of interest in complex video or audio scenes.
- Content-Based Visual Database Query and Indexing - e.g. visual query for content of interest.
- Tracking of Content and/or Features - e.g. the tracking of content or features of interest over time in an audio or video scene allowing user interaction with the tracking process.
- Selective Coding of Objects - for the separate compression of content data to allow separate access to, and query of, the content in a data-base.
- Improved Coding Efficiency - for the reconstruction of content with sufficient quality by means of sophisticated compression technology.
- Selective Error Protection - for robust storage or transmission of important content in error prone environments.

Collaboration within COST 211quat is achieved by technical contributions and presentation of results at meetings, by remote working via email, and by simulation software exchange. Focusing on important areas in the multimedia domain and stimulating collaborative research has allowed the group to play a major role in a number of standards, (e.g. H.263, MPEG-4 and MPEG-7). However COST 211 does not limit itself to contributions specifically related to its focus at a particular time. Rather, COST 211 is an open and flexible collaborative framework which welcomes contributions from members on any topic of interest within the field of image/video processing. This includes, but is not limited to, compression (block-based or otherwise), image/video segmentation, feature extraction, content-based retrieval, evaluation methodologies and hardware considerations.

1.3 Collaboration with other EU Programmes and Projects

COST 211 collaborates with the ACTS projects MoMuSys, MODEST, and DICEMAN. This collaboration includes the exchange of meeting reports, software executables, and test sequences. Furthermore, parts of the software tools developed in the MoMuSys Work Package 5.2 [Marcotegui 99] and the MODEST project are based on algorithms from the COST 211 Analysis Model. Therefore, additionally a large amount of knowledge and experience is exchanged between the different projects.

2. Recent Developments within COST 211quat

2.1 The COST Analysis Model [Alatan 98]

Recently, the focus has been on improving the COST 211 Analysis Model, called the COST AM, and equipping it with objective evaluation measures to assess the segmentation quality [Wollborn 97]. Another major effort has been on content-based indexing and retrieval over image, video and music databases.

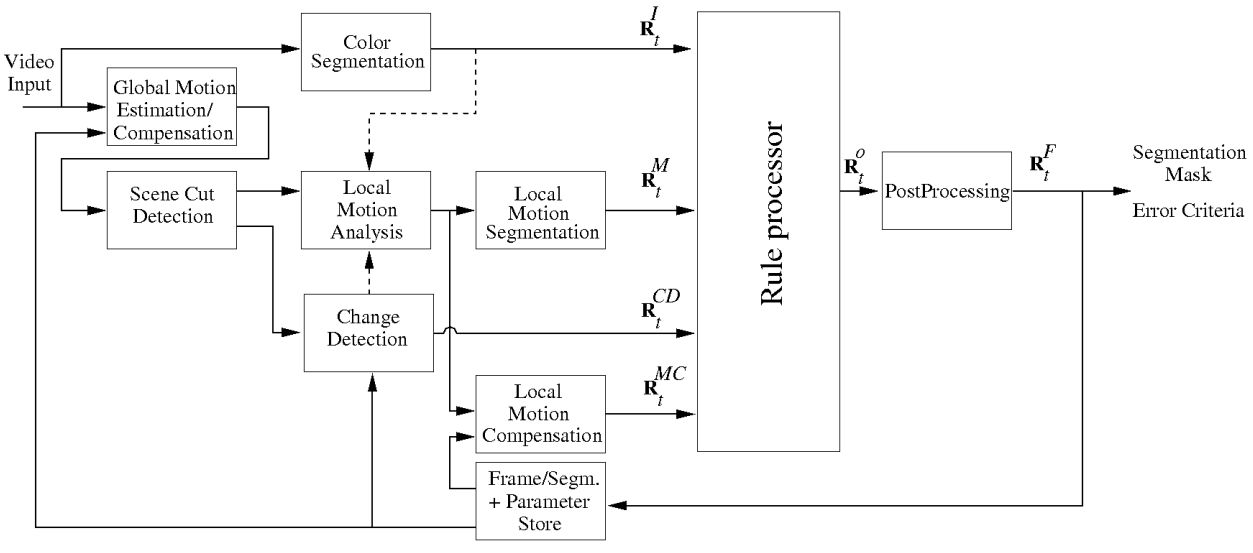

Fig. 1: The COST Analysis Model

The COST AM is a realisation of the COST 211 kANT (kernel of Analysis for New multimedia Technologies). The COST AM consists of a full description of tools and algorithms for automatic and semiautomatic segmentation of video signals, including object detection, extraction and tracking as shown in Fig. 1. The COST AM has been implemented in software and shared by the COST 211 Group. It has now reached version 4 and will continue to be improved. Basic modules [Mech 98] of the VOP generation tool described in an informative annex of the ISO/MPEG-4 standard [MPEG 98] originated in the COST AM.

2.2 Content-Based Retrieval

Since the early 1990s, content-based retrieval of digital imagery has become a very active area of research. Traditional methods for image indexing by keywords associated with each image are inadequate in terms of effective description of image contents, since they are subjective (relative to the person indexing the images) and are not even possible in some applications due to the fast growth of the digital material collections.

Both industrial and academic systems for image retrieval have been built. Most of these systems (e.g. QBIC from IBM, *Netra* from UCSB, *Virage* from Virage Inc.) support one or more of the following options: browsing, searching by example, search based on a single or a combination of low level features such as colour, shape, texture, spatial layout of objects in the scene and keywords. Some systems may allow other options such as facial feature extraction and semi-automatic annotation, e.g. *Photobook* (MIT Media Lab.), HHI-search engine; or compressed domain visual feature extraction, e.g. *VisualSEEk* and *WebSEEk* (Columbia University). Other systems follow different approaches, e.g. *MARS* (UI at Urbana-Champaign) focuses on how to organise various visual features into a meaningful retrieval architecture which can dynamically adapt to different applications and different users rather than finding a single best feature [Rui 99].

The contribution of the COST 211quat partners to this active research area consists of three systems being developed independently by HHI, Telefónica I+D and TUT (*MUVIS*). The first, from HHI is a system for automatic detection and facial tracking. The technique used may be applied in 3-D multimedia displays; object oriented image coding, security control, expression recognition and other interactive user interfaces. The second is a video indexing system being developed at Telefónica I+D which represents the audio-visual data as a hierarchy of data units (sequence, topic, scene, shot, image, and object). To each unit, descriptors selected from a fixed set are assigned. These descriptors can be high level attributes (title, author, subject...), low level attributes (texture, colour...) or context descriptors (temporal position, duration, spatial position). The values of these descriptors may propagate up or down the hierarchy in various ways. The third, called *MUVIS* developed at TUT [MUVIS 99], is a platform independent system composed of several modules, see Fig.2. Each of these modules performs one or a set of specific tasks, such as image processing, feature extraction, similarity estimation, index creation, database management and visualisation of query results. *MUVIS* allows querying by example images, sketches (currently only one object contour per query is supported), drawings, keywords, colour features (dominant colours and a *k*-element colour histogram in the HSV colour space), texture features (Zernike moments and Gabor filters), and shape features based on contour scale-space images.

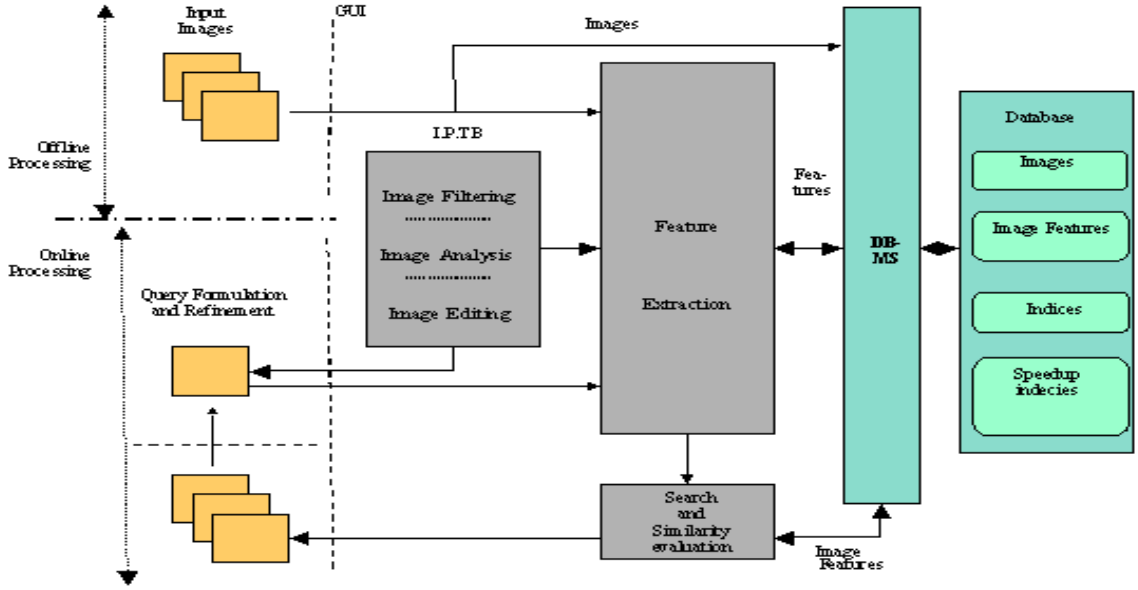

Fig. 2: MUVIS Block Diagram

Indexing techniques for searching in high-dimensional feature spaces, such as the pyramid and the *R**-tree technique, are being investigated. MUVIS can be extended to video indexing and retrieval by adding new modules to extract additional features, such as scene and shot boundaries, events inside a shot, key frames, etc.

2.3 Other activities

A major problem in segmenting a video sequence into moving objects and static background is that the estimation is disturbed by moving cast shadows. Therefore, a technique for the detection of image regions changed by a moving cast shadow is currently included in the COST 211 Analysis Model. Furthermore, these cast shadows are tracked as special objects [Stauder 99].

Two research projects at TUT are dealing with content-based analysis and retrieval of music. The first is concerned with content analysis of musical signals by automatic transcription of music. In this project, methods of observing sounds in polyphonic signals are investigated. Specifically, algorithms for multi-pitch estimation are studied. The second project is focusing on content-based indexing of music using metadata, e.g. MIDI files.

3 Call for Comparison

COST 211quat is planning to issue an Open Call for Comparison with the COST AM. A set of image sequences will be selected along with their reference segmentation masks for the comparison. Both visual and objective evaluation of the results will be conducted and the results will be published in a special journal issue. The details are currently being worked out. Look at the COST website [http://www.teltec.dcu.ie/cost211/] for details.

4 Conclusions

The COST 211quat Action has been briefly reviewed in this paper. The COST AM, which is the testbed for image sequence analysis and segmentation has been summarised and the new activities concerning content-based analysis have been briefly described. An Open Call for Comparison with the COST AM is expected to be published shortly.

5 References

[Alatan 98] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel and T. Sikora, “Image sequence analysis for emerging interactive multimedia services- The European COST 211 framework”, IEEE Trans. on Circuits and Systems for Video Technology, 1998.

[Sikora 97] T. Sikora and H. Seguin, “The European COST 211ter Group – Research on redundancy reduction techniques and content analysis for multimedia services”, Proc. of Workshop on Image Analysis for Multimedia Services, WIAMIS’97, 24-25 June 1997, Louvain-la-Neuve, Belgium, pp. 127-134.

[Mech 98] R. Mech, M. Wollborn, “A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera”, Signal Processing, Vol. 66, No. 2, April 1998, pp. 203-217.

[Wollborn 97] M. Wollborn, R. Mech, “Procedure for Objective Evaluation of VOP Generation Algorithms”, Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/2704, Fribourg, Switzerland, October 1997.

[Stauder 99] J. Stauder, R. Mech, J. Ostermann, “Detection of Moving Cast Shadows for Object Segmentation”, IEEE Transactions on Multimedia, Vol. 1, No. 1, March 1999, pp. 65-76.

[Marcotegui 99] B. Marcotegui, P. Correia, F. Marques, R. Mech, R. Rosa, M. Wollborn, and F. Zanoguera, “VOGUE: The MoMuSys Video Object Generator with User Environment”, Proc. of Workshop on Image Analysis for Multimedia Interactive Services 1999 (WIAMIS’99), Berlin, Germany, May/June 1999.

[MPEG 98] ISO/IEC JTC1/SC29/WG11 Doc. N2502, “Information Technology – Generic Coding of Audio-visual Objects, Part 2: Visual, ISO/IEC 14496-2”, Final Draft of International Standard, October 1998.

HYPERLINK[MUVIS 99] F. Alaya Cheikh, B. Cramariuc, C. Reynaud, M. Quinghong, B. Dragos-Adrian, B. Hnich, M. Gabbouj, P. Kerminen, T. Mäkinen and H. Jaakkola, "MUVIS: A System for Content-Based Indexing and Retrieval in Large Image Databases," Proceedings of the SPIE/EI'99 Conference on Storage and Retrieval for Image and Video Databases VII, Vol. 3656, San Jose, California, 26-29 January 1999.

[Rui 99] Y. Rui, T.S. Huang and S-F. Chang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues”, Journal of Visual Communication and Image Representation, v 10, n 1, p 39-62, March 1999.