

**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11
MPEG 98/N2462
October 1998/Atlantic City**

Title: MPEG-7 Applications Document v.7
Source: MPEG Requirements
Status: Approved

MPEG-7 Applications

1. INTRODUCTION.....	2
2. MPEG-7 FRAMEWORK.....	2
3. MPEG-7 APPLICATION DOMAINS.....	3
4. “PULL” APPLICATIONS.....	3
4.1 STORAGE AND RETRIEVAL OF VIDEO DATABASES.....	3
4.2 DELIVERY OF PICTURES AND VIDEO FOR PROFESSIONAL MEDIA PRODUCTION.....	5
4.3 COMMERCIAL MUSICAL APPLICATIONS (KARAOKE AND MUSIC SALES).....	7
4.4 SOUND EFFECTS LIBRARIES.....	8
4.5 HISTORICAL SPEECH DATABASE.....	9
4.6 MOVIE SCENE RETRIEVAL BY MEMORABLE AUDITORY EVENTS.....	9
4.7 REGISTRATION AND RETRIEVAL OF MARK DATABASES.....	10
5. “PUSH” APPLICATIONS.....	12
5.1 USER AGENT DRIVEN MEDIA SELECTION AND FILTERING.....	12
5.2 PERSONALISED TELEVISION SERVICES.....	13
5.3 INTELLIGENT MULTIMEDIA PRESENTATION.....	14
5.4 INFORMATION ACCESS FACILITIES FOR PEOPLE WITH SPECIAL NEEDS.....	15
6. SPECIALISED PROFESSIONAL AND CONTROL APPLICATIONS.....	16
6.1 TEleshopping.....	16
6.2 BIO-MEDICAL APPLICATIONS.....	17
6.3 REMOTE SENSING APPLICATIONS.....	18
6.4 SEMI-AUTOMATED MULTIMEDIA EDITING.....	19
6.5 EDUCATIONAL APPLICATIONS.....	20
6.6 SURVEILLANCE APPLICATIONS.....	21
6.7 VISUALLY-BASED CONTROL.....	22
7. REFERENCES.....	23

ANNEX A: SUPPLEMENTARY REFERENCES, BY APPLICATION	24
4.1 & 4.2 STORAGE AND RETRIEVAL OF VIDEO DATABASES & DELIVERY OF PICTURES AND VIDEO FOR PROFESSIONAL MEDIA PRODUCTION	24
5.1 USER AGENT DRIVEN MEDIA SELECTION AND FILTERING.....	27
5.2 INTELLIGENT MULTIMEDIA PRESENTATION.....	27
6.5 EDUCATIONAL APPLICATIONS.....	27
ANNEX B: AN EXAMPLE ARCHITECTURE FOR MPEG-7 PULL APPLICATIONS	28

1. Introduction

This ‘MPEG-7 Applications Document’ lists a number of applications that should be enabled by MPEG-7 tools. It does certainly not list *all* the applications enabled by MPEG-7, but rather gives an idea of what should be possible using MPEG-7 technology, including improving existing applications as well as presenting completely new ones.

The purpose of the document is:

- to provide a better understanding of what MPEG-7 should be, and what functionality it should deliver,
- to be a ‘Public Relations’ instrument that can help explain what MPEG-7 is, and
- to be of use when writing the concrete requirements for MPEG-7.

For each of the applications, four sections are given:

1. The description of the application,
2. The application-specific requirements
3. The requirements that the application places on MPEG-7, and
4. Relevant work and references for the application.

2. MPEG-7 Framework

Nowadays, more and more audio-visual information is available from many sources around the world. Also, there are people who want to use this audio-visual information for various purposes. However, before the information can be used, it must be located. At the same time, the increasing availability of potentially interesting material makes this search more difficult. This challenging situation led to the need of a solution to the problem of quickly and efficiently searching for various types of multimedia material interesting to the user. Moreover, MPEG-7 is not only enables this type of search, but also enables filtering. Thus, MPEG-7 will support both push and pull applications. MPEG-7 wants to answer to this need, providing this solution.

MPEG-7, formally called ‘Multimedia Content Description Interface’, will standardise:

- A set of description schemes and descriptors
- A language to specify description schemes, i.e. a Description Definition Language (DDL).
- A scheme for coding the description

For more details regarding the MPEG-7 background, goals, areas of interest, and work plan please refer to document N2460, “MPEG-7: Context and Objectives” [1]. MPEG-

7's initial requirements are indicated in document N2461, "MPEG-7 Requirements" [2].

3. MPEG-7 Application Domains

The increased volume of audio-visual data available in our everyday lives requires effective multimedia systems that make it possible to access, interact and display complex and inhomogeneous information. Such needs are related to important social and economic issues, and are imperative in various cases of professional and consumers applications such as:

- Education,
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face),
- Tourist information,
- Cultural services (history museums, art galleries, etc.),
- Entertainment (e.g. searching a game, karaoke),
- Investigation services (human characteristics recognition, forensics),
- Geographical information systems,
- Remote sensing (cartography, ecology, natural resources management, etc.),
- Surveillance (traffic control, surface transportation, non-destructive testing in hostile environments, etc.),
- Bio-medical applications,
- Shopping (e.g. searching for clothes that you like),
- Architecture, real estate, and interior design,
- Social (e.g. dating services), and
- Film, Video and Radio archives.

4. "Pull" Applications

A preliminary note on the division of this document:

There is a multitude of ways of dividing this group of applications into different categories. Originally, applications were divided by medium, but later were categorised by delivery paradigm. This is not to imply an ordering or priority of divisions, but is simply a reflection of what was convenient at the time. Other means of dividing the list of applications may be done by content type, user group, and position in the content.

MPEG-7 began its life as a scheme for making audio-visual material "as searchable as text is today." Although the proposed multimedia content descriptions are now acknowledged to serve much more than search applications, they remain for many the primary applications for MPEG-7. These retrieval, or "pull" applications, involve databases, audio-visual archives, and the web-based internet paradigm (a client requests material from a server.)

4.1 Storage and retrieval of video databases

Application Description

Television and film archives store a vast amount of multimedia material in several different formats (digital or analogue tapes, film, CD-ROM, etc.) along with precise

descriptive information (meta-data) which may or may not be precisely timecoded. This meta-data is stored in databases with proprietary formats. There is an enormous potential interest in an international standard format for the storage and exchange of descriptions that could ensure:

- interoperability between video archive operators,
- perennial relevance of the meta-data, and
- a wider diffusion of the data to the professional and the general public.

MPEG-7, in short, *must* accommodate visual and other search of such existing multimedia databases.

In addition, a vast amount of the older, analogue audio-visual material will be digitised in years to come, which creates a tremendous opportunity to include content-based indexing features (which can be extracted during the digitisation/compression process) into those existing data-bases.

In the case of new audio-visual material, the ability to associate descriptive information within video streams at various stages of video production can dramatically improve the quality and productivity of manual, controlled-vocabulary annotation of video data in a video archive. For example, pre-production and post-production scripts, information captured or annotated during shooting, and post-production edit lists would be very useful in the retrieval and re-use of archival material.

Essential associated activities to this one are cost-efficient video sequence indexing and shot-level indexing for stock footage libraries [4].

A sample architecture is outlined in Annex B.

Application requirements

Specific requirements for those applications are:

- Support of full-text descriptions as well as structured fields (database descriptions);
- Multi-language support;
- We desire the ability to interoperate between different content description semantics (e.g. different database schemas, different thesauri, etc.) or to translate from each content description semantic into MPEG-7 semantics;
- The ability to reference audio-visual objects or object instances and time references, even in analogue format;
- The ability to include descriptions with incomplete or missing time references (a shot description that has not been timecoded);
- The ability to handle multiple versions of the same document at several stages in the production process, and descriptions that apply to multiple copies of the same material.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- There must exist a class of descriptors that support unstructured (free-) text in multiple languages.

A note about text: Descriptors should depend as little as possible on a specific language. If text is needed as a descriptor, the language used must be specified in the text description and a text description may contain several translations. The character set chosen must enable the use of all languages (as appropriate to ISO).

- There must exist a class of descriptors that support structured text.
- There must exist a mechanism by which different MPEG-7 DS's can interoperate.
- There must be a robust linking mechanism that allows for temporal references to material with incomplete timecode.
- There must be a mechanism by which document versions may be identified.

Application relevant work and references

Bloch, G. R. (1988). From Concepts To Film Sequences. In *RIAO 88*, (pp. 760-767). MIT Cambridge MA.: March 21-24, 1988.

Davis, M. (1993). Media streams: An iconic visual language for video annotation. *Teletronikk*, 89(4), 59 - 71.

EBU/SMPTE Task Force, First Report : User Requirements, version 1.22, chapitre 1 : Compression, chapitre 2 : Metadata and File Wrappers, chapitre 3 : Transfer Protocols. *SMPTE Journal*, April 1997.

Parkes, A. P. (1989b). The Prototype CLORIS system: Describing, Retrieving and Discussing Videodisc Stills and Sequences. *Information Processing and Management*, 25(2), 171 - 186.

ISO/TC 46 /SC 9, Information and documentation - Presentation, identification and description of documents
<<http://www.nlc-bnc.ca/iso/tc46sc9/index.htm>>

ISO/TC 46 /SC 9, Working Group 1 (1997) Terms of reference and tasks for the development of an International Standard Audio-visual Number (ISAN). Document ISO/TC 46/SC 9 N 235, May 1997.

See also Annex A.

4.2 Delivery of pictures and video for professional media production

Application description

[note: this section is still to be re-evaluated]

Studios need to deliver appropriate videos to TV channels. The studio may have to deliver a whole video, based on some global meta-data, or video segments, for example to edit an archive-based video, or a documentary, or advertisement videos.

In this application, due to the users' expertise, one formulates relevant and possibly detailed "pull" queries, which specify the desired features of some video segments. With present video databases, these queries are mainly based on objective characteristics at segment level. However, they can also take advantage of subjective characteristics of these segments, as perceived by one or several users.

The ability to formulate a single query on the client side, and send it to many distributed databases is very important to many production studios. The returned items should include visual abstracts, copyright and pricing information, as well as a measure of the technical quality of the source video material.

In this application, one should separate news programs, which must be made widely and instantly available for a short period of time, from other production programs, which can be retrieved on a permanent basis, usually from secondary or tertiary

storage. On-line news services providing instant access to the day's news footage are being built by many broadcasters and archives (including INA, BBC, etc), using proprietary formats, and would benefit from standardisation if they were to be consolidated into common services such as the Eurovision News Exchange (which currently uses broadcast channels, not databases).

Still pictures have similar applications and requirements as pertaining to design. The web designer must not only make new designs but also collect the already available graphics on the net for use in the designed web sites. Other design fields have similar uses for visual search.

Application requirements

Requirements are similar to the previous application. They are mainly characterised by:

- Support of feature-based and concept-based queries at segment level, and
- Support for similarity queries.
- Support for different data formats.
- Support for art & design specific parameters.
- Media summaries for fast browsing.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- There must exist a mechanism that embodies conceptual knowledge.
- There must exist a mechanism that allows for links to media summaries.

Application relevant work and references

Aigrain, P., Joly, P., & Longueville, V. (1995). Medium Knowledge-Based Macro-Segmentation of Video into Sequences. In M. Maybury (Ed.) (pp. 5-16), *IJCAI 95 - Workshop on Intelligent Multimedia Information Retrieval*. Montréal: August 19, 1995

The Art Teacher Connection
<http://www.primenet.com/~arted>

Cohen, A., Levy, M., Roeh, Itzhack, Gurevitch, M. (1995) *Global Newsrooms, Local Audiences, A study of the Eurovision News Exchange*, Acamedia Research Monograph 12, John Libbey.

European League of Institutes of the Arts
<http://www.elia.ahk.nl>

Pentland, A. P., Picard, R., Davenport, G., & Haase, K. (1994). *Video and Image Semantics: Advanced Tools for Telecommunications* (Technical Report No. 283). MIT.

Sack, W. (1993). Coding News And Popular Culture. In *The International Joint Conference on Artificial Intelligence (IJCA93) Workshop on Models of Teaching and Models of Learning*. Chambery, Savoie, France.

Zhang, H., Gong, Y., & Smoliar, S. W. (1994). Automated parsing of news video. In *IEEE International Conference on Multimedia Computing and Systems*, (pp. 45 - 54). Boston: IEEE Computer Society Press.

see also 4.1 'Application relevant work and references', and Annex A.

4.3 Commercial musical applications (Karaoke and music sales)

Application Description

The Karaoke industry is extremely large and popular. One of the aims of the pastime is to make the activity of singing in public as effortless and unintimidating as possible. Requiring a participant to recall the name and artist of a popular tune is unnecessary when one considers that the amateur performer must know the song well enough to sing it. A much friendlier interface results if you allow someone to hum a few memorable bars of the requested tune, and to have the computer find it (or a short list of alternatives, if the brief segment under-specifies the intended selection).

A similar application dealing with Karaoke, but also relating to music sales, below, is enabling solo Karaoke-ists to expand their repertoire in the privacy of their own home. Much of the industry is currently driven by people wishing to practice in their own homes. One can easily imagine a complete on-line database, in which someone selects a song she knows from the radio, sings a few bars, and the entire arrangement is downloaded to their computer, with appropriate payment extracted.

The consumer music industry is currently struggling with how to reach consumers with increasingly fragmented tastes. Music, as with all broadcast media artefacts, is undergoing the same internet-flavoured transformation as cable television: tastes are changing to prefer narrowcast over broadcast. An ideal way of presenting consumers with available music is to allow them effortless search.

The mechanics are similar to the above Karaoke example. Querents may hum approximate renditions of the song they seek from a kiosk or from the comfort of their own home. Alternately, they may seek out music with similar features (musicians, style, tempo, or year of creation) to those that they already know. From there, they may listen to an appropriate sample (and perhaps view associated information such as lyrics or a video), and choose to buy the music on the spot.

Application requirements

- Robust representations of melody and other musical features which allow for reasonable errors on the part of the indexer in order to accommodate query-by-humming,
- Associated information, and
- Cross-modal search.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- There must exist mechanism that supports melody and other musical features.
- There must exist a mechanism that supports descriptors based on information associated with the data (e.g., textual data).
- Support description schemes that contain descriptors of visual, audio, and/or other features, and support links between the different media.

Application relevant work and references

The area of query-by-humming may be the most-researched field within auditory query by content. A few example papers include:

- Ghias, A., Logan, J., Chamberlain, D., Smith, B. C. (1995). "Query by humming-musical information retrieval in an audio database," ACM Multimedia '95 San Francisco.
<<http://www.cs.cornell.edu/Info/People/ghias/publications/query-by-humming.html>>
- Kageyama, T., Mochizuki, K., Takashima, Y. (1993). "Melody retrieval with humming," ICMC '93 Tokyo proceedings, 349-351.
- Lindsay, A. (1996). "Using Contour as a Mid-Level Representation of Melody," S.M. Thesis, MIT Media Laboratory, Cambridge, MA.
<<http://sound.media.mit.edu/~alindsay/thesis.html>>

In addition, an interesting example of search for musical (and other) products is Firefly's BigNote. (<<http://www.firefly.com/>>, <<http://www.firefly.net/>>, and <<http://www.bignote.com/>>) Rather than using extensive meta-information (although it does allow for some search on production and genre information), Firefly's engine derives its power from a shared user base, performing automatic collaborative filtering.

4.4 Sound effects libraries

Application Description

Foley artists, sound designers, and the like must deal with extremely large databases of sound effects to be used for a variety of applications daily. Existing database management and search solutions are typically proprietary and therefore closed, or open, and unsuitable for any serious, orderly work.

A sound designer may specify a sound effect type, for example, naming the source of the sound, and select from variations on that sound. A designer may provide a prototypical sound, and detail features such as, "bigger, more distant, but keeping the same brightness." One may even vocalise the type of abstract sound one seeks, in an onomatopoeic variation of query-by-humming. Essential to the application is the ability to navigate a space of similar sound effects.

Application requirements

- Compact representation of sound effects,
- Sound source name and characteristics, and
- Ability to specify classes of audio-visual objects, with features to accommodate selection.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for hierarchical descriptors and description schemes.
- Support for text based descriptors.
- Efficient coding of descriptors.

Application relevant work and references

Blum, T, et al, "Content-based classification, search, and retrieval of audio," in *Intelligent multimedia information retrieval*, Maybury, Mark T. (ed) (1997). Menlo Park, Calif.

Mott, R.L. (1990) "Sound Effects: Radio, TV, and Film," Focal Press, Boston, USA.

4.5 Historical speech database

Application Description

One may search for historical events through key words spoken ("We will bury you"), key events ('shoe banging'), the speaker ('Nikita Krushchev'), location and/or context ('address to the United Nations'), date (12 October 1960), or a combination of any or all of the above in order to call up an audio recording, an audio-visual presentation, or any other associated facts. This application can aid in education (See also 6.4-Film music education) or journalistic research.

Application requirements

- Representation of textual content of an auditory event.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for text based descriptors.

Application relevant work and references

[to come]

4.6 Movie scene retrieval by memorable auditory events

Application Description

In our post-modern world, many visual events are referred to by memorable spoken words. This is no more evident than when referring to comedic movie or television scenes ("This parrot is bleedin' demised," and "land shark,") or movies by auteurs ("I'm sorry, did I ruin your concentration?" "thirty-seven!?" and "there's only trouble and desire.") by key words. One should be able to look up a movie (and rent a viewing of a particular scene, for example) by quoting such catch phrases. It is not hard to imagine a new market growing up around such micro-views and micro-payments, based on impulse viewing.

In a similar vein, auditory events in soundtracks may be just as accessible as spoken lines in certain circumstances. A key example is the screeching violins in the "Psycho" soundtrack at the point of the infamous shower scene. Those repeated harsh notes ("Scree-ee-ee-ee!") are iconic to a movie-going public, and a key feature of an important movie.

Application requirements

- Search by example audio.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for audio descriptors.

Application relevant work and references
[to come]

4.7 Registration and retrieval of mark databases

Application Description

Registration of marks is to protect the inventor or service provider in the form of exclusive rights of exploitation through legal proceedings from misuse or imitation. A mark is a sign, or a combination of signs, capable of distinguishing the goods or services of one undertaking from those of other undertakings. In general, the sign may be in the form of two-dimensional image that consists of text, drawings or pictures, emblems including colors. Two-dimensional marks can be categorized into the following three types as listed below:

- Word-in mark
Contains only characters or words in the mark. (best described by text annotation)
- Device mark
Contains graphical or figurative elements only. (shape descriptor needed)
- Composite mark
Consists of characters or words and graphical elements. (combination of above descriptors)

If a mark is registered, then no person or enterprise other than its owner may use it for goods or services identical with or similar to those for which the mark is registered. Any unauthorized use of a sign similar to the protected mark is also prohibited, if such use may lead to confusion in the minds of the public. The protection of a mark is generally not limited in time, provided its registration is periodically renewed (typically, every 10 years) and its use continues. Therefore, this number is expected to keep growing rapidly, and it is estimated that the number of registrations and renewals of marks effected worldwide in 1995 was in the order of millions.

In order to register a mark, one has to make sure no identical ones are registered before. For the types of “Word-in mark” and “Composite-mark,” text annotation may be adequate for the retrieval from the database. “Device-mark” type, however, is characterized only by the shape of the object. In addition, this type may not have distinct orientation or scale. When the operator enters a new mark to database for registration, he/she wants to make sure that no identical one is already in the system in disregard of its orientation angle or scale. Furthermore, he/she may want to see how similar shaped ones are already in the system even if there is no identical one. The search process should be robust to noise in image or minor variations in its shape. Any relevant information such as annotation or textual description of the mark should also be accessible if requested.

A mark designer may want the same thing. In addition, to avoid possible inadvertent infringement of the copyright, the designer may wish to see whether some possible variations of the mark under design are already registered.

In this respect, it is desirable for the system capable of returning the retrieved results in terms of the similarity, and displaying the results simultaneously for comparison. So far, the current practice to retrieve similar or the same device-type mark is performed manually by human operator resulting in many duplicated registrations.

Therefore, there is an enormous potential need for an automatic retrieval of marks by the contents-based similarity not only in the international community but also in domicile. The submission of the mark to the system can be done interactively on-line to refine the search-process on a web-based Internet paradigm [5].

Application requirements

Specific requirements for those applications are:

- Efficient interactive response time.
- Support for a mechanism by which a mark image may be submitted for similarity based retrieval.
- Support for visual based descriptors by which modifications can be made to any of the retrieved results for fine-tuning the search-process (relevance feedback).

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for shape-based and content-based queries;
- Support for precise, shape-oriented similarity queries;
- Support for scale and orientation independence of marks;
- Support for media summaries for fast browsing;
- Support for linking relevant information
- Support for D's and DS's invariant under transformations irrelevant to the intended features

Application relevant work and references

- Andrews, B. (1990). U. S. patent and trademark office ORBIT trademark retrieval system. *T-term user guide, examining attorney's version*, : October, 1990.
- Cortelazzo, G., & Mian, G. A., & Vezzi, G., & Zamperoni, P. (1994). Trademark shapes description by string-matching techniques. *Pattern Recognition*, 27(8), 1005-1018.
- Eakins, J. P. (1994). Retrieval of trademark images by shape feature. *Proc. of Int. Conf. on Electronic Library and Visual Information Research*, 101-109, May, 1994.
- Eakins, J. P., & Shields, K., & Boardman, J. (1996). ARTISAN – a shape retrieval system based on boundary family indexing. *Proc. SPIE, Storage and Retrieval for Image and Video Database IV*, vol. 2670, 17-28, Feb. 1996.
- Lam, C. P., & Wu, J. K., & Mehtre, B. (1995). STAR - a system for trademark archival and retrieval. *Proceedings 2nd Asian Conf. on Computer Vision*, vol. 3, 214-217.
- Kim, Y-S, & Kim, W-Y (1998). Content-Based Trademark Retrieval System Using Visually Salient Feature, *Journal of Image and Vision Computing*, vol. 16/12-13, August 1998.

WORLD INTELLECTUAL PROPERTY ORGANIZATION: (WIPO)

<http://www.wipo.org/eng/dgtext.htm>

5. “Push” Applications

In contrast with the above “pull” applications, the following “push” applications follow a paradigm more akin to broadcasting, and the emerging webcasting. The paradigm moves from indexing and retrieval, as above, to selection and filtering. Such applications have very distinct requirements, generally dealing with streamed descriptions rather than static descriptions stored on databases.

5.1 User agent driven media selection and filtering

Application description

Filtering is essentially the converse of search. Search involves the pull of information, while filtering implies information ‘push.’ Search requests the inclusion of information, while filtering excludes data. Both pursuits benefit strongly from the same sort of meta-information.

Broadcast media are unlikely to disappear any time soon. In fact, there is a movement to make the World Wide Web, primarily a pull medium, more broadcast-like. If we can enable users to select information more appropriate to their uses and desires from a broadcast stream of 500 channels, using the same meta-information as that used in search, then this is an application for MPEG-7.

This application gives rise to several sub-types, primarily divided among types of users. A consumer-oriented selection gives rise to personalised audio-visual programmes, for example. This can go much farther than typical video-on-demand in collecting personally relevant news programmes, for example. A content-producer oriented selection made on the segment or shot level is a way of collecting raw material from archives.

Application requirements

- Efficient interactive response times, and
- The capability to characterise a media object by a set of concepts that may be dependent on locality or language.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for descriptors and description schemes that allow multiple languages.
- There must exist a mechanism by which concepts may be represented

Application relevant work and references

Lieberman, H. (1997) “Autonomous Interface Agent,” In proceedings of Conference on Computers and Human Interface, CHI-97, Atlanta, Georgia.

Maes, P. (1994a) “Agents that Reduce Work and Information Overload,” Communications of the ACM, vol. 37, no. 7, pp. 30 - 40.

Marx, M., & C. Schmandt (1996) “CLUES: Dynamic Personalized Message Filtering,” In proceedings of Conference on Computer Supported Cooperative Work /CSCW 96, edited by M. S. Ackermann, pp. 113 - 121, Hyatt Regency Hotel, Cambridge, Mass.: ACM.

Nack, F. (1997) Considering the Application of Agent Technology for Collaboration in Media-Networked Environments. IRIS20,

Shardanand, U., & P. Maes (1995) "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," In proceedings of CHI-95 Conference, Denver, CO: ACM Press.

See also Annex A.

5.2 Personalised Television Services

In the broadcast area, the MPEG-7 description can provide the user with assistance in selection of broadcast data, be it for immediate or later viewing, or for recording. In a personalized broadcast scenario, the data offered to the user can be filtered from broadcast streams according to his own profile, the generation of which may be done automatically (e.g. based on location, age, gender or on the previous selection behavior) or semi-automatically (e.g. based on pre-set interests). The broadcast of MPEG-7 description streams will enable providers of Electronic Programme Guides (EPGs) with a variety of capabilities, wherein *presentation* of MPEG-7 data (also along with the original AV data) will also be an important aspect. In combination with NVOD (Near-Video on Demand) services and recording, new functionalities like stepping forward/backward based on keyframe selection and changes in the sequel of scenes for speed-up in presentation are possible. Extended interactivity functionalities, related to specific events in the programmes, are of importance for future broadcast services as well. This can include "offline" interactivity based on a recorded broadcast stream, which can require identification of the associated event during callback. It can be expected that MPEG-7 data will be transmitted along with the AV data streams, or (e.g. for an EPG channel) also as separate streams [6].

Application requirements

- Description of broadcast media objects in terms of, for example, content type, author, cast, parental rating, textual description, temporal relationship, locality- and language-dependent features, service provider, and IP protection.
- Description of specific events in broadcast media objects
- Specification of interactivity capability related to specific events (e.g. polling) and definition of associated interaction channels (e.g. telephone numbers or hyperlinks)
- Presentation of content-related data, also along with the associated media objects, and manipulation of the presentation based on the content description (e.g. by event status)
- Support for APIs that define receiver-side filter function, interaction capability or definition of extended content description
- Support for unique presentation of the media objects and the associated content description, controllable by user interaction and description parameters
- The broadcast metadata to be put into MPEG-7 streams must be flexible, the broadcasters must be able to "spice" the content based on available resources. As little information as possible should be mandatory, such that broadcasters can start the service without a huge investment. Upwards compatibility with existing standards like DVB-SI, ATSC PSIP or metadata definitions originating from the EBU/SMPTE task forces also falls under this aspect.

- Capability to set up content- or service-related links between different broadcast media objects (not necessarily transmitted simultaneously)

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for descriptors and description schemes that allow definition of broadcast media objects and events as listed above in their temporal (absolute and relative time bases, duration of validity) and physical (channel, stream) context
- Support for streaming of MPEG-7 data
- Support for interactivity, including identification of the related event during a callback procedure

5.3 Intelligent multimedia presentation

Application Description

Given the vast and increasing amount of information available, people are seeking new ways of automating and streamlining presentation of that data. That may be accomplished by a system that combines knowledge about the context, user, application, and design principles with knowledge about the information to be displayed. Through clever application of that knowledge, one has an intelligent multimedia presentation system.

Application requirements

- The ability to provide contextual and domain knowledge, and
- The ability to represent events and the temporal relationships between events.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- There must exist a mechanism by which contextual information may be encoded.
- There must exist a mechanism by which temporal relationships may be represented

Application relevant work and references

André, E., & Rist, T. (1995). Generating Coherent Presentations Employing Textual and Visual Material. *Artificial Intelligence Review, Special Volume on the Integration of Natural Language and Vision Processing*, 9(2 - 3), 147 - 165.

Bordegoni, M., et al, "A Standard Reference Model for intelligent Multimedia Presentation Systems," April 1997, pre-print.

<http://www.dfki.uni-sb.de/~rist/csi97/csi97.html>

Davenport, G., & Murtaugh, M. (1995). ConText: Towards the Evolving Documentary. In *ACM Multimedia 95 - Electronic Proceedings*. San Francisco, California: November 5-9, 1995.

<http://ic.www.media.edu/icPeople/murtaugh/acm-context/acm-context.html>

Feiner, S. K., & McKeown, K. R. (1991). Automating the Generation of Coordinated Multimedia Explanations. *IEEE Computer*, 24(10), 33 - 41.

Maybury, M. T. (ed.) (1993) Intelligent Multimedia Interfaces. AAAI Press/ MIT Press, Cambridge, MA.

Maybury, Mark T. (ed) (1997) Intelligent multimedia information retrieval. Menlo Park, Calif.

See also Annex A.

5.4 Information access facilities for people with special needs

Application description

In our increasingly information dependent society we have to facilitate accessibility to information to every individual user. However, some people face serious accessibility problems to information, not because they lack the economic or technical basis but rather because they suffer from one or several disabilities, e.g. visual, auditory, motor, or cognitive disabilities. Providing active information representations might help to overcome the problems. The key issue is to allow multi-modal communication to present information optimised for the abilities of individual users.

Thus, it is important to develop technical aids to facilitate communication for people with special needs. For example, a search agent that does not exclude images as information resource for the blind but rather makes available the MPEG-7 meta-data. Aided by that meta-data, sonification (auditory display), or haptic display is made possible. Similarity of meta-data helps to provide a set of information in different modalities, in case the particular information is not accessible for the user.

Such applications provide full participation in society by removing communication and information access barriers that restrict interactions between people with and without disabilities, and they will lead to improved global commerce opportunities.

Application requirements

[no new ones apparent]

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support description schemes that contain descriptors of visual, audio, and/or other features .

Application relevant work and references

The Yuri Rubinsky Insight Foundation:

<http://www.yuri.org/webable/library.html#guidelinesandstandards>

The Centre of Cognitive Science:

<http://www.cogsci.ed.ac.uk/>

The Human Communication Research Centre:

<http://www.hcrc.ed.ac.uk/>

6. Specialised Professional and Control Applications

The following potential MPEG-7 applications do not limit themselves to traditional, media-oriented, multimedia content, but are functional within the meta-content representation to be developed under MPEG-7. They reach into such diverse, but data-intensive, domains as medicine and remote sensing. Such applications can only serve to increase the usefulness and reach of this proposed international standard.

6.1 Teleshopping

Application description

More and more merchandising is being conducted through catalogue sales. Such catalogues are rarely effective if they are restricted to text. The customer who browses such a catalogue is more likely to retain visual memories than text memories, and the catalogue is frequently designed to cultivate those memories. However, given the sheer size of many of these catalogues, and the fact that most people only have a vague idea of what they want ("I'll know it when I see it"), they will only be effective if it is possible to find items by successively refining and/or redirecting the search. Typically, the customer will spot something that is almost right, but not quite. He or she will then want to fine-tune the search-process by interacting with the system. E.g. "I'm looking for brown shoes, a bit like those over there, but with a slightly higher heel," or "I'm looking for curtains with that sort of pattern, but in a more vivid colour."

Catalogues of items for which stock maintenance is difficult or expensive but for which the search process is essentially visual (e.g. garden design, architecture, interior decorating, oriental carpets) are especially aided by this application. For such items, detailed digital image-databases could be supported and updated centrally and accessed from distributed selling points.

Application requirements

- Support for interactive queries with few predicted constraints,
- Support for precise, product-oriented similarity queries, and
- MPEG-7 should operate "as fast as possible," allowing efficient interactive response times.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for visual based descriptors.

Application relevant work and references

André, E., J. Mueller, & T. Rist (1997) "Adding Animated Presentation Agents to the Interface," To appear in the proceedings of IJCAI 97 - Workshop on Animated Interface Agents: Making them intelligent, Nagoya, Japan.

Chavez, A., & P. Maes (1996) "Kasbah: An Agent Marketplace for Buying and Selling Goods," In proceedings of 1. International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, London, UK.

Rossetto, L., & O. Morton (1997) "Push!" Wired, no. 3.03 UK, March 97, pp. 69 - 81.

At the moment there are very few catalogues that provide content-based image retrieval for teleshopping and the ones that do, understandably focus on applications where images of merchandise are relatively homogeneous and standardised, such as wallpaper, flooring, tiles, etc. A searchable online database of full-colour flooring samples including carpet and vinyl products can be found at <http://www.floorspecs.com>. They promise a colour match system in the near future.

6.2 Bio-medical applications

Application description

Medicine is an area in which visual recognition is often a significant technique for diagnosis. The medical literature abounds with atlases, volumes of photographs that depict normal and pathological conditions in different parts of the body, viewed at different scales. An effective diagnosis may often require the ability to recall that a given condition resembles an image in one of the atlases. The amount of material catalogued in such atlases is already large and continues to grow. Furthermore, it is often very difficult to index using textual descriptions only. Therefore, there is a growing demand for search-engines that can respond to image-driven queries. This will allow physicians to access image-based information in a way that is similar to the current keyword-based search-engines such as MEDLINE. (E.g. in order to make a differential diagnosis, a radiologist might want to compare medical records and case-histories of all patients in a medical database for which radiographs showed similar lesions or pathologies. Furthermore, as 3D-imaging techniques keep gaining importance, such image-driven queries will have to be able to handle both 2- and 3-dimensional data. Cross-modal search will apply when one includes associated clinical auditory descriptions, such as associating a cough with a chest x-ray in order to aid diagnosis.

Biochemical interactions crucially depend on the three-dimensional structure of the participating modules (e.g. the shape-complementarity between signal-molecules and cell-receptors, or the key-in-lock concepts for immunological recognition). Thanks to the sustained effort of a large number of biomedical laboratories, the list of molecules for which the chemical composition and spatial structure are documented is growing continuously. Given the fact that it is still extremely difficult to predict the structure of a biomolecule on the basis of its primary structure (i.e. the string of constituent atoms), searching these databases will only be helpful if it can be done on the basis of shape. It is not difficult to make the leap from the ability to search on 3-D models, as proposed by MPEG-7. Such applications would be extremely helpful in drug-design, as one could e.g. search for biomolecules with shapes similar to a candidate-drug to get an idea of possible side effects.

Application requirements

- The ability to link libraries of correlated and relevant information (e.g. images, patient history, clinical findings, medication regime),
- The ability to perform on-line annotations and mark regions-of-interest with any shape or form of connectivity (from small and compact to large and diffuse),
- The ability to search images that contain similar regions-of-interest, ignoring the visual characteristics of the rest of the image,
- The ability to handle both n -dimensional data (and their time-evolution).

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for n-dimensional data descriptors.
- Support for descriptors containing temporal information.
- There must exist a mechanism by which a specific segmentation or hierarchical level may be chosen, to the exclusion of all other segments.

Application relevant work and references

One development that might be relevant is the Picture Archiving and Communication System (PACS) which allows physicians to annotate images with text, references and questions. This greatly facilitates interaction and consultation with colleagues and specialists.

Furthermore, there is the STARE-project (STructured Analysis of the REtina) at the University of California, San Diego, which is an information system for the storage and content-based retrieval of ocular fundus images.

<<http://oni.ucsd.edu/stare>>

6.3 Remote Sensing Applications

Application description

In remote sensing applications, the requirements of satellite image databases, namely several millions of images acquired according to various modalities (panchromatic, multispectral, hyperspectral, and hexagonal sampling, etc.), the diversity of potential users (scientists, military, geologists, etc.), and improvements in telecommunication techniques make it necessary to define a highly efficient description standard. Until now, information search in image libraries is based on *textual* information such as scene name, geographic, spectral, and temporal information. Based on this, information exchange is achieved by means of tapes and photographs.

A challenging aspect is to provide capabilities of exploiting such complex databases from on-line systems supporting the following functionalities:

- textual query,
- image query based on either whole or part of a reference image (one or several spectral bands),
- content-based retrieval,
- browsing, and
- confidentiality and data protection.

MPEG-7 should be an appropriate framework for solving such requests.

Application requirements

- support various description schemes,
- support for different data (e.g. multispectral, hyperspectral, and SAR) associated with various sensors (ground resolution, wavelength),
- ability to include multiple descriptions of the same documents,
- ability to link correlated information, similar region of interest, and
- ability to take into account time evolution for 2D and 3D data.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for multiple descriptors and description schemes for the same data.
- Support for descriptors for unique data types.
- Support for descriptors embodying temporal change.
- There must be a mechanism by which descriptors and description schemes within a description may be linked to other D's and DS's within the same description.

Application relevant work and references

[to come]

6.4 Semi-automated multimedia editing

Application Description

Given sufficient information about its contents, what could a multimedia object do? With sufficient information about its own structure combined with methods on how to manipulate that structure, a 'smart' multimedia clip could start to edit itself in a manner appropriate to its neighbouring multimedia. For example, a piece of music and a video clip, from different sources, could be combined in a way such that the music stretches and contracts to synchronise with specific 'hit' points in the video, and thus create an appropriate and customised soundtrack.

This could be a new paradigm for multimedia, adding a 'method' layer on top of MPEG-7's 'representation' layer. By making multimedia 'aware,' to an extent, one opens access to beginning users and increases productivity for experts. Such hidden intelligence on the part of the data itself shifts multimedia editing from direct manipulation to loose management of data.

Semi-automated multimedia editing is a broad category of applications. It can facilitate video editing for home users as well as experts in studios through varying amounts of guidance or assistance through the process. In its simpler version, assisted editing can consist of an MPEG-7-enabled browser for the selection of video shots, using a suitable *shot description* language. In an intermediate version, assisted editing can include planning, i.e. proposing shot selections and edit points, satisfying a scenario expressed in a *sequence description* language.

Application requirements

- Pointers as 'handles' that refer to the data directly, to allow manipulation of the multimedia.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Ability to link descriptors to the data that they describe.

Application relevant work and references

Bloch, G.R (1986) Elements d'une Machine de Montage Pour l'Audio-Visuel. Ph.D., Ecole Nationale Supérieure Des Telecommunications.

- Nack, F. (1996) AUTEUR: The Application of Video Semantics and Theme Representation for Automated Video Editing. Ph.D. Thesis, Lancaster University.
- Parkes, A.P. (1989) An Artificial Intelligence Approach to the Conceptual Description of Videodisc Images. Ph.D. Thesis, Lancaster University.
- Sack, W., & Davis, M. (1994). IDIC: Assembling Video Sequences from Story Plans and Content Annotations. IEEE International Conference on Multimedia Computing Systems, Boston, MA: May 14 - 19, 1994.
- Sack, W., & Don, A. (1993) Splicer: An Intelligent Video Editor (Unpublished Working Paper, MIT).

6.5 Educational applications

Application description

The challenge of using multimedia in educational software is to make as much use of the intrinsic information as possible to support different pedagogical approaches such as summarisation, question answering, or detection of and reaction to misunderstanding or non-understanding.

By providing direct access to short video sequences within a large database, MPEG-7 can promote the use of audio, video and film archive material in higher education in many areas:

- **History:** Radio, television and film provide detailed accounts of many contemporary events, useful for class-room presentations, provided that a sufficiently precise (MPEG-7) description can be queried based on dates, places, personalities, etc. (*see also 4.5-Historical Speech Databases*)
- **Performing arts (music, theatre):** Fine-grained, standardised descriptions can be used to bring a selection of relevant documents into the classroom for special classes, using on line video archives as opposed to costly local tape libraries. For instance, several productions of a theatrical scene, or musical work, can thus be consulted for comparison and illustration. Because classic and contemporary theatre are widely available in translation, this application can target worldwide audiences.
- **Film Music:** A tool can be developed for improving the knowledge and skills of users in the domain of film theory/practice and film music (music for film genres). Depending on the user's background the system should provide enough material to not only improve the user's ability in understanding the complexity of each single medium but also to handle the complex relationships between the two media film and music. To achieve this, the system should offer an environment in which the student can perform guided/supported experiments, e.g. on editing film, mixing sound, or combining both, which requires that the system can analyse and criticise the results achieved by the user.

Thus, this system must be able to automatically generate film/sound sequences and their synchronisation based on stereotypical music/film pattern for film genres, and perhaps ways to creatively break the established generating rules.

Application requirements

- Linking mechanisms to synchronise between MPEG-7 descriptors and other sources of information (e.g. HTML, SGML, World Wide Web services, etc.)

- Mechanisms for allowing specialised vocabularies.
- The ability to allow real time operation in conjunction with a database.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for interoperation with description schemes.
- Support for descriptors and description schemes that allow specialised languages or vocabularies.
- There must exist a mechanism by which descriptions may link to external information.

Application relevant work and references

Margaret Boden (1991) *The Creative Mind: Myths and Mechanisms*, Basic Books, New York

Schank, R. C. (1994). Active Learning through Multimedia. *IEEE MultiMedia*, 1(1), 69 - 78.

Sharp, D., Kinzer, C., Risko, V. & the Cognition and Technology Group at Vanderbilt University. (1994). The Young Children's Video Project: Video and software tools for accelerating literacy in at-risk children. Paper presented at the National Reading Conference, San Diego, CA
http://www.edc.org/FSC/NCIP/ASL_VidSoft.html

Tagg, Philip (1980, ed.). *Film Music, Mood Music and Popular Music Research. Interviews, Conversations, entretiens.* 1980, SPGUMD 8002

Tagg, Philip (1987). Musicology and the Semiotics of Popular Music. *Semiotica*, 66-1/3: 279-298. (This and other texts accessible on-line via
<http://www.liv.ac.uk/ipm/tagg/taggwbtx.htm>

See also the reference section of 6.3-Semi-automated multimedia editing and Annex A.

6.6 Surveillance applications

Application description

There are a number of surveillance applications, in which a camera monitors sensitive areas and where the system must trigger an action if some event occurs. The system may build its database from no information or limited information, and accumulate a video database and meta-data as time elapses. Meta-content extraction (at an “encoder” site) and meta-data exploitation (at a “decoder” site) should exploit the same database.

As time elapses and the database is sufficiently large, the system, at both sides, should have the ability to support operations on the database, such as:

1. Search on the audio/video database for a specific event (synthetic or current data). Event is a sequence of audio/video data.
2. Find similar events in the past.
3. Make decisions on the current data related to the accumulated database, and/or to a-priori known data.

A related application is in security and forensics, in the matching of faces or fingerprints.

Application requirements

- Real time operation in conjunction with a database, and
- Domain-specific features.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for descriptors and description schemes that allow specialised languages or vocabularies.
- Support for descriptors for unique data types.

Application relevant work and references

Courtney, J. D. (1997). Automatic Video Indexing by Object Motion Analysis. *Pattern Recognition*, vol. 30, no. 4, 607-626.

6.7 Visually-based control

Application description

In the field of control, there have been several developments in the area of visually based control. Instead of using text-based approaches for control programming, images, visual objects, and image sequences are used to specify the control behaviour and are an integral part of the control loop (e.g. visual servoing).

One aspect in the description of control information between (video) objects is that objects are not necessarily associated via temporal spatial relationships. Accumulation of the control and video information allows visually based functions such as redo, undo, search-by-task, or object relationship changes [7].

Application requirements

- Description of relationships between arbitrary object nodes in addition to spatio-temporal relationships, as in e.g. BIFS
- Allow searches based on the arbitrary (control) associations.
- MPEG-7 should operate “as fast as possible,” allowing efficient interactive response times.

MPEG-7 Specific Requirements

Requirements specific to MPEG-7 are:

- Support for descriptors and description schemes containing spatio-temporal relationships.
- Support for descriptors containing relationships between arbitrary objects.

Application relevant work and references

Palm, S.R., Mori, T., Sato, T. (1998) Bilateral Behavior Media: Visually Based Teleoperation Control with Accumulation and Support, Submitted to Robotics & Automation Magazine special issue on Visual Servoing.

7. References

- [1_ MPEG Requirements Group, “MPEG-7: Context and Objectives”, Doc. ISO/MPEG N2460, MPEG Atlantic City Meeting, October 1998.
- [2_ MPEG Requirements Group, “MPEG-7 Requirements”, Doc. ISO/MPEG N2461, MPEG Atlantic City Meeting, October 1998.
- [3] AHG on MPEG-7, “MPEG-7 Applications Document,” Doc. ISO/MPEG M4013, MPEG Atlantic City Meeting, October 1998.[4] Rémi Ronfard, “MPEG-7 Applications in Radio, Film, and TV archives,” Doc. ISO/MPEG M2791, MPEG Fribourg Meeting, October 1997.
- [5] Whoi-Yul Kim et al, “MPEG-7 Applications Document,” Doc. ISO/MPEG M3955, MPEG Atlantic City Meeting, October 1997.
- [6] Jens-Rainer Ohm et al, “Broadcast Application and Requirements for MPEG-7,” Doc. ISO/MPEG M4107, MPEG Atlantic City Meeting, October 1997.
- [7] Stephen Palm, “Visually Based Control: another Application for MPEG-7,” Doc. ISO/MPEG M3399, MPEG Tokyo Meeting, March 1998.

Annex A: Supplementary references, by application

4.1 & 4.2 Storage and retrieval of video databases & Delivery of pictures and video for professional media production

- Aguierre Smith, T. G., & Davenport, G. (1992). The Stratification System. A Design Environment for Random Access Video. In *ACM workshop on Networking and Operating System Support for Digital Audio and Video*, San Diego, California
- Aguierre Smith, T. G., & Pincever, N. C. (1991). Parsing Movies In Context. In *Proceedings of the Summer 1991 Usenix Conference*, (pp. 157-168). Nashville, Tennessee.
- Aigrain, P., & Joly, P. (1994). The automatic real-time analysis of film editing and transformation effects and its applications. *Computer & Graphics*, 18(1), 93 - 103.
- American Library Association's ALCTS/LITA/RUSA. Machine-Readable Bibliographic Information Committee. (1996). The USMARC Formats: Background and Principles. *MARC Standards Office, Library of Congress, Washington, D.C. November 1996*
- Bateman, J. A., Magnini, B., & Rinaldi, F. (1994). The Generalized Italian, German, English Upper Model. In *Proceedings of the ECAI94 Workshop: Comparison of Implemented Ontologies*, Amsterdam.
- Bloch, G. R. (1986) *Elements d'une Machine de Montage Pour l'Audio-Visuel*. Ph.D., Ecole Nationale Superieure Des Telecommunications.
- Bobrow, D. G., & Winograd, T. (1985). An Overview of KRL: A Knowledge Representation Language. In R. J. Brachman & H. J. Levesque (Eds.), *Readings in Knowledge Representation* (pp. 263 - 285). San Mateo, California: Morgan Kaufmann Publishers.
- Butler, S., & Parkes, A. (1996). Film Sequence Generation Strategies for generic Automatic Intelligent Video Editing. *Applied Artificial Intelligence (AAI)* [Ed: Hiroaki Kitano], Vol. 11, No. 4, pp. 367-388.
- Butz, A. (1995). *BETTY - Ein System zur Planung und Generierung informativer Animationssequenzen* (Document No. DFKI-D-95-02). Deutsches Forschungszentrum fur Kunstliche Intelligenz GmbH.
- Chakravarthy, A., Haase, K. B., & Weitzman, L. (1992). A uniform Memory-based Representation for Visual Languages. In B. Neumann (Ed.), *ECAI 92 Proceedings of the 10th European Conference on Artificial Intelligence*, (pp. 769 - 773). Wiley, Chichester: Springer Verlag.
- Chakravarthy, A. S. (1994). Toward Semantic Retrieval of Pictures and Video. In C. Baudin, M. Davis, S. Kedar, & D. M. Russell (Ed.), *AAAI-94 Workshop Program on Indexing and Reuse in Multimedia Systems*, (pp. 12 - 18). Seattle, Washington: AAAI Press.
- Davenport, G., Aguiere Smith, T., & Pincever, N. (1991). Cinematic Primitives for Multimedia. *IEEE Computer Graphics & Applications* (7), 67-74.

- Davenport, G., & Murtaugh, M. (1995). ConText: Towards the Evolving Documentary. In *ACM Multimedia 95 - Electronic Proceedings*. San Francisco, California: November 5-9, 1995.
<http://ic.www.media.edu/icPublications/gdlist.html>
- Davis, M. (1995) *Media Streams: Representing Video for Retrieval and Repurposing*. Ph.D., MIT.
- Del Bimbo, A., Vicario, E., & Zingoni, D. (1992). A Spatio-Temporal Logic for Sequence Coding and Retrieval. In *IEEE Workshop on Visual Languages*, (pp. 228 - 231). Seattle, Washington: IEEE Computer Society Press.
- Del Bimbo, A., Vicario, E., & Zingoni, D. (1993). Sequence Retrieval by Contents through Spatio Temporal Indexing. In *IEEE Symposium on Visual Languages*, (pp. 88 - 92). Bergen, Norway: IEEE Computer Society Press.
- Domeshek, E. A., & Gordon, A. S. (1995). Structuring Indexing for Video. In J. Lee (Ed.), *First International Workshop on Intelligence and Multimodality in Multimedia Interfaces: Research and Applications*.. Edinburgh University: July 13 - 14, 1995.
- Gregory, J. R. (1961) *Some Psychological Aspects of Motion Picture Montage*. Ph.D. Thesis, University of Illinois.
- Haase, K. (1994). FRAMER: A Persistent Portable Representation Library. In *ECAI 94 European Conference on Artificial Intelligence*, (pp. 732- 736). Amsterdam, The Netherlands.
- Hampapur, A., Jain, R., & Weymouth, T. E. (1995a). Indexing in Video Databases. In *Storage and Retrieval for Image and Video Databases II*, (pp. 292 - 306). San Jose, California, 9 - 10 February 1995: SPIE.
- Hampapur, A., Jain, R., & Weymouth, T. E. (1995b). Production Model Based Digital Video Segmentation. *Multimedia Tools and Applications*, 1, 9 - 46.
- International Standard Z39.50: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification". <http://lcweb.loc.gov/z3950/agency/>
- Isenhour, J. P. (1975). The Effects of Context and Order in Film Editing. *AV Communication Review*, 23(1), 69 - 80.
- Lenat, D. B., & Guha, R. V. (1990). *Building Large Knowledge-Based Systems - Representation and Inference in the Cyc Project*. Reading, MA.: Addison-Wesley.
- Lenat, D. B., & Guha, R. V. (1994). Strongly Semantic Information Retrieval. In C. Baudin, M. Davis, S. Kedar, & D. M. Russell (Ed.), *AAAI-94 Workshop Program on Indexing and Reuse in Multimedia Systems*, (pp. 58 - 68). Seattle, Washington: AAAI Press.
- Mackay, W. E., & Davenport, G. (1989). Virtual Video Editing in Interactive Multimedia Applications. *Communications of the ACM*, 32(7), 802 - 810.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993). *Introduction to WordNet: An On-line Lexical Database* (<ftp://clarity.princeton.edu/pub/wordnet/5papers.ps>). Cognitive Science Laboratory, Princeton University.

- Nack, F. (August 1996) AUTEUR: The Application of Video Semantics and Theme Representation for Automated Film Editing. Ph.D. Thesis, Lancaster University
- Nack, F. and Parkes, A. (1995). AUTEUR: The Creation of Humorous Scenes Using Automated Video Editing. Proceedings of JCAI-95 Workshop on AI Entertainment and AI/Alife, pp. 82 - 84, Montreal, Canada, August 19, 1995.
- Nack, F. and Parkes, A. (1997) Towards the Automated Editing of Theme-Oriented Video Sequences. Applied Artificial Intelligence (AAI) [Ed: Hiroaki Kitano], Vol. 11, No. 4, pp. 331-366.
- Nagasaka, A., & Tanaka, Y. (1992). Automatic video indexing and full-search for video appearance. In E. Knuth & I. M. Wegener (Eds.), *Visual Database Systems* (pp. 113 - 127). Amsterdam: Elsevier Science Publishers.
- Oomoto, E., & Tanaka, K. (1993). OVID: Design and Implementation of a Video-Object Database System. *IEEE Transactions On Knowledge And Data Engineering*, 5(4), 629-643.
- Parkes, A. P. (1989a) *An Artificial Intelligence Approach to the Conceptual Description of Videodisc Images*. Ph.D. Thesis, Lancaster University.
- Parkes, A. P. (1989c). Settings and the Settings Structure: The Description and Automated Propagation of Networks for Perusing Videodisk Image States. In N. J. Belkin & C. J. van Rijsbergen (Ed.), *SIGIR '89*, (pp. 229 - 238). Cambridge, MA:
- Parkes, A. P. (1992). Computer-controlled video for intelligent interactive use: a description methodology. In A. D. N. Edwards & S. Holland (Eds.), *Multimedia Interface Design in Education* (pp. 97 - 116). New York: Springer-Verlag.
- Parkes, A., Nack, F. and Butler, S. (1994) Artificial intelligence techniques and film structure knowledge for the representation and manipulation of video. Proceedings of RIAO '94, Intelligent Multimedia Information Retrieval Systems and Management, Vol. 2, Rockefeller University, New York, October 11-13, 1994.
- Pentland, A., Picard, R., Davenport, G., & Welsh, B. (1993). *The BT/MIT Project on Advanced Tools for Telecommunications: An Overview* (Perceptual Computing Technical Report No. 212). MIT.
- Sack, W., & Davis, M. (1994). IDIC: Assembling Video Sequences from Story Plans and Content Annotations. In *IEEE International Conference on Multimedia Computing and Systems*. Boston, Ma: May 14 - 19, 1994.
- Sack, W., & Don, A. (1993). *Splicer: An Intelligent Video Editor* (Unpublished Working Paper).
- Tonomura, Y., Akutsu, A., Taniguchi, Y., & Suzuki, G. (1994). Structured Video Computing. *IEEE MultiMedia*, 1(3), 34 - 43.
- Ueda, H., Miyatake, T., Sumino, S., & Nagasaka, A. (1993). Automatic Structure Visualization for Video Editing. In *ACM & IFIP INTERCHI '93*, (pp. 137 - 141).

Ueda, H., Miyatake, T., & Yoshizawa, S. (1991). IMPACT: An Interactive Natural-Motion-Picture Dedicated Multimedia Authoring System. In *Proc ACM CHI '91 Conference on Human Factors In Computing Systems*, (pp. 343-450).

Yeung, M. M., Yeo, B., Wolf, W. & Liu, B. (1995). Video Browsing using Clustering and Scene Transitions on Compressed Sequences. In *Proceedings IS&T/SPIE '95 Multimedia Computing and Networking, San Jose*. SPIE (2417), 399 - 413.

Zhang, H., Kankanhalli, A., & Smoliar, S. W. (1993). Automatic Partitioning of Full-Motion Video. *Multimedia Systems*, 1, 10 - 28.

We are still seeking references regarding ANSI guidelines for Multi-lingual Thesaurus and International radio and television typology.

5.1 User agent driven media selection and filtering

Maes, P. (1994b) "Modeling Adaptive Autonomous Agents," *Journal of Artificial Life*, vol. 1, no. 1/2, pp. 135 - 162.

Hanke Fjordhotel, Norway, August 9-12, 1997. Parise, S., S. Kiesler, L. Sproull, & K. Waters (1996) "My Partner is a Real Dog: Cooperation with Social Agents," In proceedings of Conference on Computer Supported Cooperative Work /CSCW 96, edited by M. S. Ackermann, pp. 399 - 408, Hyatt Regency Hotel, Cambridge, Mass.: ACM.

Rossetto, L., & O. Morton (1997) "Push!," *Wired*, no. 3.03 UK, March 97 , pp. 69 - 81.

5.2 Intelligent multimedia presentation

Andre, E. (1995). *Ein planbasierter Ansatz zur Generierung multimedialer Präsentationen.*, Ph.D., Sankt Augustin: INFIX, Dr. Ekkerhard Hundt.

Andre, E., & Rist, T. (1994). Multimedia Presentations: The Support of Passive and Active Viewing. In *AAAI Spring Symposium on Intelligent Multi-Media Multi-Modal Systems*, (pp. 22 - 29). Stanford University: AAAI.

Maybury, M. T. (1991) "Planning Multimedia Explanations using Communicative Acts," In proceedings of Ninth National Conference on Artificial Intelligence, AAAI-91, pp. 61 - 66, Anaheim, CA: AAAI/MIT Press.

Riesbeck, C. K., & Schank, R. C. (1989). *Inside case-based reasoning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

6.5 Educational Applications

Tagg, Philip (1981). On the Specificity of Musical Communication. Guidelines for Non-Musicologists. SPGUMD 8115 (23 pp.)

Tagg, Philip (1984). Understanding 'Time Sense': concepts, sketches, consequences. *Tvorspel*: 21-43. [Forthcoming on-line, see Tagg 1987].

Tagg, Philip (1990). Music in Mass Media Studies. *Reading Sounds for Example*. *PMR* 1990: 103-114

Annex B: An example architecture for MPEG-7 Pull applications

A search engine could freely access any complete or partial description associated with any AV object in any set of data, perform a ranking and retrieve the data for display by using the link information. An example architecture is illustrated in fig.1.

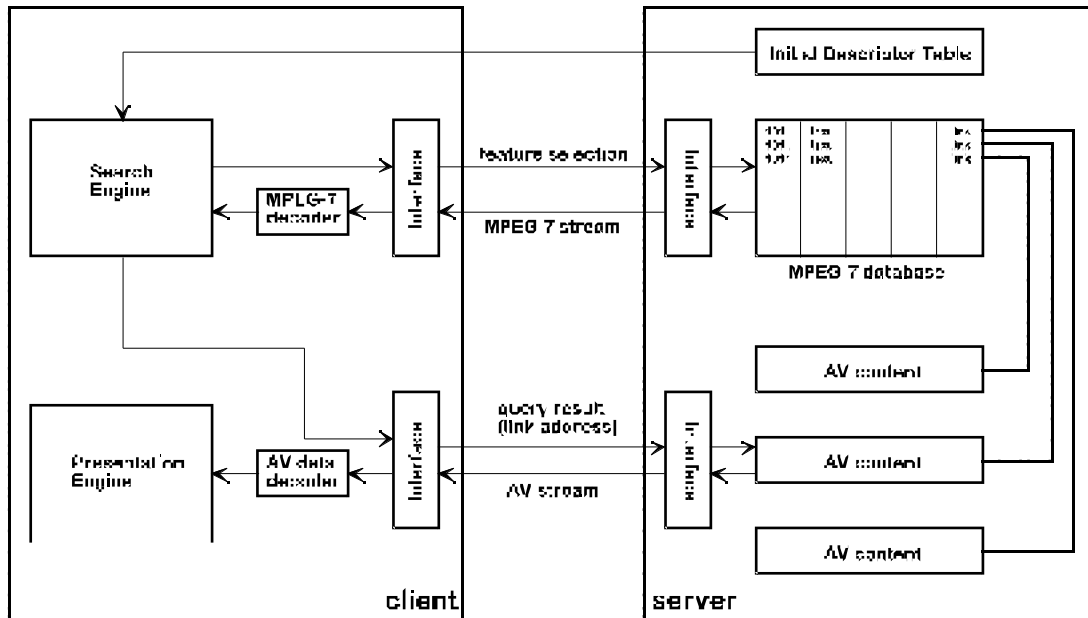


Fig. 1. Example of a client-server architecture in a MPEG-7 based data search.