

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

ISO/IEC JTC1/SC29/WG11

M3039

**MPEG 97/
January 1998**

Title: contribution to MPEG-7 Proposal Package Description (PPD)

Authors: Pascal Faudemay, Yong Rui

Status: Proposal for discussion on second Draft

1. Introduction

Notice

This document was written based on “very first draft” of MPEG-7 PPD (document N1923), and on some of the discussions in Fribourg and in the AHG. However, it is a personal contribution. For time reasons, we did not include the modifications to N1923 presented by Ibrahim Sezan, which synchronize it to other input documents. Modifications vs. N1923 are indicated through the Word versioning mechanisms.

More and more audiovisual information is available in digital form, in various places around the world. Along with the information, people appear that want to use it. Before one can use any information, however, it will have to be located first. At the same time, the increasing availability of potentially interesting material makes this search harder.

Currently, solutions exist that allow searching for textual information. Many text-based search engines are available on the World Wide Web, and they are among the most visited sites - indicating they foresee a real demand. Identifying information is, however, not possible for audiovisual content, as no generally recognized description of this material exists.

In general, it is not possible to efficiently search the web for, say, a picture of ‘the Motorbike from Terminator II’, and still less, to search a sequence where “King Lear congratulates its assistants on the night after the battle”, or “twenty minutes of video according to my preferences of today”. In specific cases, solutions do exist. Multimedia databases on the market today allow searching for pictures using characteristics like color, texture and information about the shape of objects in the picture. Solutions exist to retrieve the text content of speech and concepts within this text. One could also envisage a search example for audio, in which one can whistle a melody to find a song.

The question of finding content is not restricted to database retrieval applications; also in other areas similar questions exist. For instance, there is an increasing amount of (digital) broadcast channels available, and this makes it harder to select the broadcast channel (radio or TV) that is potentially interesting, and moreover, which programs to store on an intelligent VCR

About a year ago, MPEG started a new work item to provide a solution to the questions described above. The new member of the MPEG family, called “Multimedia Content Description Interface” (in short ‘MPEG-7’), will extend the limited capabilities of proprietary solutions in identifying content that exist today, notably by including more data types. It will enable easy creation of new data types adapted to various user communities and application domains (e.g. surgery,

science-fiction movies, ethnology of rural areas, etc.), and to avoid the inconvenience of proprietary solutions, such as more limited markets.

In other words: MPEG-7 will specify a way to create a new description type and/ or description model of multimedia documents, and basic description schemes that can be used to describe various types of multimedia information. This description shall be associated with the content itself, to allow fast and efficient searching for material of a user's interest. It will be possible either to include these descriptions with the AV material, or to have separate descriptions and some mapping mechanism with the AV material.

AV material that has MPEG-7 data associated with it, can be indexed and searched for. This 'material' may include: still pictures, graphics, 3D models, animation, audio, speech, video, and information about how these elements are combined in a multimedia presentation ('scenarios', composition information). Special cases of these general data types may include facial expressions and personal characteristics.

2. Description of MPEG-7

The four concepts that are needed for describing the scope of MPEG-7 are

- Data
- Description Scheme
- Descriptors
- Coded Descriptors

They can be defined and explained as follows. (These definitions will be improved. They have to be harmonized with the Requirements and other related documents.)

Data: It refers to information that will be described using MPEG7.

Description Scheme : A data representation scheme that is used to define the associated descriptors. A description scheme can be seen as a model of the multimedia document and/ or of the represented real (or imaginary) world.

A description scheme can be seen as a n-uple with the following components :

$O = O(D, S, F, R, A, W)$

where O is the multimedia object or document, which is a function of D, S, F, R, A, and W.

D is the data, e.g. an MPEG movie, an analog movie, a book on some bookshelf in a library;

$S = \{s_i\}$ is the object structure, in terms of the set of "subobjects", which can be described in the object: e.g. a time interval, a shot, a region in a frame, a sound segment.

$F = \{f_i\}$ is a set of features associated with D, e.g. color, texture, or shape features of a structure element, or the qualification of this part as "superb" or "surrealistic" ;

$R = \{r_j\}$ is a set of representations for a given feature f_i .

$A = \{a_k\}$ is a set of associations between data and features or feature types, or between several features or feature types, or between several data. E.g. the association between a subobject and a “production date” is a 1:1 mapping of some type, the association between a subobject and the “superb” feature is an 1:M mapping, subobjects A and B are representations of the same character C.

$W = \{w_l\}$ is a set of “views” on the data and their content. Each view is defined by a description author or description supplier, or some characteristics of the description author.

Descriptors: An instantiation of the representation structure defining the description scheme.

Descriptions: The entity describing the data and consisting of description schemes and descriptors.

Coded Descriptions: Compact versions of descriptions.

As examples consider the following. Digital still images or MPEG-2 compressed video are examples of *data*. Color of still images is a *description scheme* and an image color histogram is a *descriptor*. Description scheme may be textual and the descriptors can be keywords. As an other example, a description scheme can be the shape of a visual object where the descriptor can be the set of vertices defining the contour of the object. Coded descriptors are compact descriptions.

2.1 Scope of MPEG-7

(The following text is taken from the current Context and Objectives Document. The text is expected to be changed as we further develop MPEG-4 requirements.)

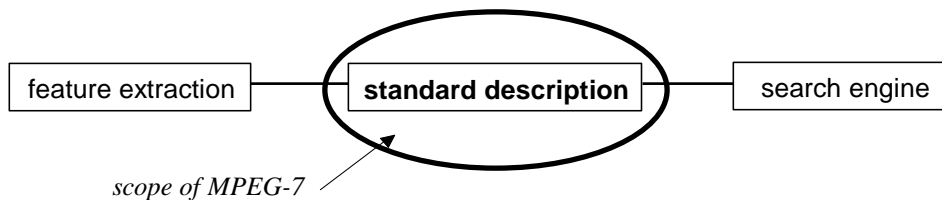
MPEG-7 will address applications that can be stored (on-line or off-line) or streamed (e.g. broadcast, push models on the Internet), and can operate in both real-time and non real-time environments. A ‘real-time environment’ means that information is associated with the content while it is being captured.

Figure 1 below shows a highly abstract block diagram of a possible MPEG-7 processing chain, included here to explain the scope of the MPEG-7 standard. This chain includes feature extraction (analysis), the description itself, and the search engine (application). To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features (or ‘descriptors’) will be extremely useful. It is also clear that automatic extraction is not always possible, however. As was noted above, the higher the level of abstraction, the more difficult automatic extraction is, and interactive extraction tools will be of good use.

However useful they are, neither automatic nor semi-automatic feature extraction algorithms will be inside the scope of the standard. The main reason is that their standardization is not required to allow interoperability, while leaving space for industry competition. Another reason not to standardize analysis is to allow making good use of the expected improvements in these technical areas.

Also the search engines will not be specified within the scope of MPEG-7; again this is not necessary, and here too, competition will produce the best results.

Nevertheless, the proposed MPEG-7 standard should be usable through available (automatic or interactive) feature extraction techniques, and should enable effective and efficient data retrieval. Contributors to the MPEG-7 standard should, as much as possible, demonstrate these possibilities.



Besides the descriptors themselves, the database structure plays a crucial role in the final retrieval's performance. To allow the desired fast judgment about whether the material is of interest, the indexing information will have to be structured, e.g. in a hierarchical or associative way.

2.2 Functionalities

- Ability for inter and intra data referencing
- Ability for ranking
- Amenable to scalable databases
- Amenable to efficient database organization for fast and effective retrieval
- Streaming capability
- Ability of identification of events/processes via spatiotemporal extent
- Support for subjective information
- Support for multimodal queries
- Support for query by example

2.3 The Structure of MPEG-7

The MPEG-7 basic structure is composed of :

2.3.1 Publication and registration of description schemes

In order to avoid proprietary description schemes and the corresponding fragmentation of the market for MPEG-7 metadata and encoders / decoders, MPEG-7 description schemes will have to be published and registered according to some standard specification.

Preferably, the publication and registration mechanism should enable automatic acquisition of new description schemes by suppliers of decoders and / or by the final users.

2.3.2 Formal representation of description schemes

Description schemes can be described according to one or several paradigms, e.g. :

- Database models, such as the well known entity-relationship model
- Languages for expression of formal semantics
- Evolved linguistic descriptions, such as dictionaries, ontologies, or other forms of linguistic resources

- Others

Description schemes should express in a separate and distinct way the structure of data from a content description point of view, the syntax of descriptors, their formal or textual semantics, their relationships, and the characterization of description suppliers and possibly, of users.

Operational semantics should describe the consequences for the user of the presence of some content descriptor. Such consequences as the relevance of the object or subobject for the user, its visualization, the storage of the object on the user system, etc., can be considered.

Production rules are formal sentences of the form “If <condition> then <action>”, where <condition> is a logical expression, and <action> can be a procedure in some programming or scripting language.

Possible mechanisms to express operational semantics can include production rules, such as in the PICS and RDF emerging standards. In such rules, actions such as “visualize” or “store” are conditioned by some descriptors values. If production rules are retained, the standard should specify one or several formalisms to express them.

However other mechanisms than production rules can also be relevant to express operational semantics.

2.3.3 Instances of description schemes

The MPEG-7 standard should not only include mechanisms to express some description schemes, but also some “default” or “example” description schemes.

2.3.4 Benchmarks and Verification models

Appropriate benchmarks and verification models should display :

- The feasibility, effectiveness and efficiency of automatic and interactive structure and feature extraction
- The feasibility, effectiveness and efficiency of data retrieval based on some proposed description schemes, or of some generic description mechanisms.
- The feasibility and efficiency of publication and registration mechanisms, and possibly of mechanisms to import new description schemes in an existing decoder

Example Applications

This section describes some example applications of MPEG-7. Other applications are discussed in Document MPEG97/m2666.

2.3.5 Video and Image Databases

(to be completed).

3. Description of MPEG-7 Workplan

3.1 Objectives of the MPEG-7 Workplan

The MPEG-7 Workplan will include the following main steps :

- More complete specification of MPEG-7 requirements and applications

- Call for proposals for benchmarks and evaluation methods
- Call for proposals for the other elements of MPEG-7, according to paragraph 2.3
- Evaluation and choice of relevant proposals
- Merging and refinement of the relevant proposals, and implementation of the corresponding verification models

3.2 Method of Developing the Standard

It should be stressed that MPEG will not necessarily choose a single standard proposal, but will choose the most relevant elements of one or several proposals, and define the draft standard based on a merge and refining of these elements.

Standard proposals should preferably come with preliminary verification models, which will be further improved in order to validate the final proposal.

3.2.1 What is called for?

(Following list is a result of a preliminary discussion.)

- Publication and registration mechanisms
- Description Schemes
- Descriptors
- Descriptions
- Coding methods for compact representation of Descriptions.
- Tools for creating and interpreting Description Schemes and Descriptors
- Syntax/Format for Description Schemes (syntax and semantics), Descriptions and Coded Descriptors
- Benchmarks and verification models

3.2.2 How is evaluation conducted?

(Following list is a result of a preliminary discussion.)

Evaluation Criteria for Description Schemes and Descriptors

- Accomodation of existing Description Schemes
- Expression efficiency
- Effectiveness
- Distinctiveness
- Processing efficiency (amenability for fast processing)

- Storage-space efficiency (i.e., compactness)
 - efficiency/complexity of coding algorithm for Descriptions
- Scalability
- Flexibility

3.2.3 What is expected for benchmark proposals

Benchmarks are classical in many computer science areas, such as databases, transaction processing of numeric processing. They are usually composed mainly of a query mix, and of a corpus mix (see e.g. [5-7])

Proposers are expected to submit

- Criteria to measure the success of a benchmark instance
- a list of generic or specific queries, about 10 queries maximum
- a generic corpus composition, in no more than about 10 corpus components.

Queries should be generic in the way that they should not define a specific query instance, but some query template or property of a compliant query. Corpus composition should be generic in that it should define the corpus in terms of items classes, characterized by some domain and / or properties, and not in terms of a specific multimedia object.

Proposers are encouraged to submit generic queries or query lists, where each query or the whole query list can be of interest for more than one application. The same property should be verified for the corpus composition. The benchmark should also apply to more than a single multimedia document type, and preferably be applicable to several media streams in a same multimedia document (multimodal queries).

Proposers may submit a family of benchmarks, where each benchmark of the family is differentiated by some queries or by some corpus components. However each element of the benchmark family should as much as possible fulfill the above criteria.

An example of a family of benchmarks is presented in document m2622. Proposers can also submit amendments to any existing proposal, such as m2622. These amendments should be proposed as a new version of the initial document.

As for the MPEG-7 standard, MPEG will not necessarily adopt a single benchmark proposal, but will possibly make its own benchmark by combining queries and corpus components from different proposals, and possibly improving and refining them.

3.2.4 How are Verification Models built?

Verification models will be applied when needed, to publication and registration mechanisms, to feature extraction mechanisms, and to the use of description schemes by retrieval systems. Models should be supplied by contributors, either as source code on some specified platform, or as verification servers, accessible on the Internet to any MPEG member. They should come together with effectiveness and efficiency measurements methods, such as elapsed or cpu time measurements, specification of the hardware and software platform, etc.

MPEG will favour the exchange and improvement of verification models, either as bitstream exchanges, or through cooperation between several verification servers on the Internet. For this

[6] D. Britton, D. DeWitt, C. Turbyfill, "Benchmarking database systems : A systematic approach", Proceedings VLDB Conference, Oct. 1983, extended version as Wisconsin Computer Science TR 526

[7] Transaction Processing Performance Council (TPC), "TPC Benchmark A Standard", Shanley Public Relations, 777 N First St. Suite 600, San Jose, California, November 1989