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

ISO/IEC JTC1/SC29/WG11 **N4032**

Singapore, March 2001

Title:	**Introduction to MPEG-7 (v3.0)**
Source:	**Requirements**
Status:	**approved**
Editors:	**Neil Day (Digital Garage Inc, JP), José M. Martínez (UPM-GTI, ES)**

1. Introduction to MPEG-7

How many times have you seen science fiction movies such as 2001: A Space Odyssey and thought, "Wow, we're so far away from having any of the fancy gadgets depicted in these movies!" In 2001, Hal, the talking computer intelligently navigates and retrieves information or runs complex operations instigated by spoken input. Or how about using an image-based query, say an image of the motorbike used by Arnold Schwartzenegger in the movie T2, to find images of similarly looking motorbikes. Dreams or reality?

As more and more audiovisual information becomes available from many sources around the world, many people would like to use this information for various purposes. This challenging situation led to the need for a solution that quickly and efficiently searches for and/or filters various types of multimedia material that's interesting to the user. For example, finding information by rich-spoken queries, hand-drawn images, and humming improves the user-friendliness of computer systems and finally addresses what most people have been expecting from computers. For professionals, a new generation of applications will enable high-quality information search and retrieval. For example, TV program producers can search with "laser-like precision" for occurrences of famous events or references to certain people, stored in thousands of hours of audiovisual records, in order to collect material for a program. This will reduce program production time and increase the quality of its content.

MPEG-7 is a multimedia content description standard, (to be defined by September 2001), that addresses how humans expect to interact with computer systems, since it develops rich descriptions that reflect those expectations. This document gives an introductory overview of the MPEG-7 standard. More information about MPEG-7 can be found at the MPEG-7 website **http://drogo.cselt.it/mpeg/** and the MPEG-7 Industry Focus Group website **http://www.mpeg-7.com**. These web pages contain links to a wealth of information about MPEG, including many publicly available documents, several lists of 'Frequently Asked Questions' and links to other MPEG-7 web pages.

1.1 What Are the MPEG Standards?

The Moving Picture Coding Experts Group (MPEG) is a working group of the Geneva-based ISO/IEC standards organization, (International Standards Organization/International Electro-technical Committee) in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and a combination of the two. MPEG-7 then is an ISO/IEC standard being developed by MPEG, the committee that also developed the Emmy Award-winning standards known as MPEG-1 and MPEG-2, and the 1999 MPEG-4 standard.

- **MPEG-1:** For the storage and retrieval of moving pictures and audio on storage media.
- **MPEG-2:** For digital television, it's the timely response for the satellite broadcasting and cable television industries in their transition from analog to digital formats.
- **MPEG-4:** Codes content as objects and enables those objects to be manipulated individually or collectively on an audiovisual scene.

MPEG-1, -2, and -4 make content available. MPEG-7 lets you to find the content you need.

Besides these standards, MPEG is currently also working in MPEG-21 a Technical Report about Multimedia Framework.

1.2 Defining MPEG-7

MPEG-7 is a standard for describing features of multimedia content.

1.2.1 Qualifying MPEG-7

MPEG-7 provides the world's richest set of audio-visual descriptions.

These descriptions are based on catalogue (e.g., title, creator, rights), semantic (e.g., the who, what, when, where information about objects and events) and structural (e.g., the colour histogram - measurement of the amount of colour associated with an image or the timbre of an recorded instrument) features of the AV content and leverages on AV data representation defined by MPEG-1, 2 and 4.

Comprehensive Scope of Data Interoperability.

MPEG-7 uses XML Schema as the language of choice for content description MPEG-7 will be interoperable with other leading standards such as, SMPTE Metadata Dictionary, Dublin Core, EBU P/Meta, and TV Anytime.

1.3 The Key Role of MPEG-7

MPEG-7, formally named "Multimedia Content Description Inter-face," is the standard that describes multimedia content so users can search, browse, and retrieve that content more efficiently and effectively than they could using today's mainly text-based search engines. It's a standard for describing the features of multimedia content. *However…*

MPEG-7 will not standardize the (automatic) extraction of AV descriptions/features. Nor will it specify the search engine (or any other program) that can make use of the description. It will be left to the creativity and innovation of search engine companies, for example, to manipulate and massage the MPEG-7-described content into search indices that can be used by their browser and retrieval tools, (*see figure 1*).

2. MPEG-7 Technical Activities

It is important to note that MPEG-7 addresses many different applications in many different environments, which

means that it needs to provide a flexible and extensible framework for describing audio-visual data. Therefore, MPEG-7 will define a multimedia library of methods and tools. It will standardize:

• *A set of descriptors:* A descriptor (D) is a representation of a feature that defines the syntax and semantics of the feature representation.

• *A set of description schemes:* A description scheme (DS) specifies the structure and semantics of the relationships between its components, which may be both descriptors and description schemes.

• *A language that specifies description schemes, the Description Definition Language (DDL):* It also allows for the extension and modification of existing description schemes. MPEG-7 adopted XML Schema Language as the MPEG-7 DDL. However, the DDL requires some specific extensions to XML Schema Language to satisfy all the requirements of MPEG-7. These extensions are currently being discussed through liaison activities between MPEG and W3C, the group standardizing XML.

• *One or more ways (textual, binary) to encode descriptions:* A coded description is a description that's been encoded to fulfill relevant requirements such as compression efficiency, error resilience, and random access.
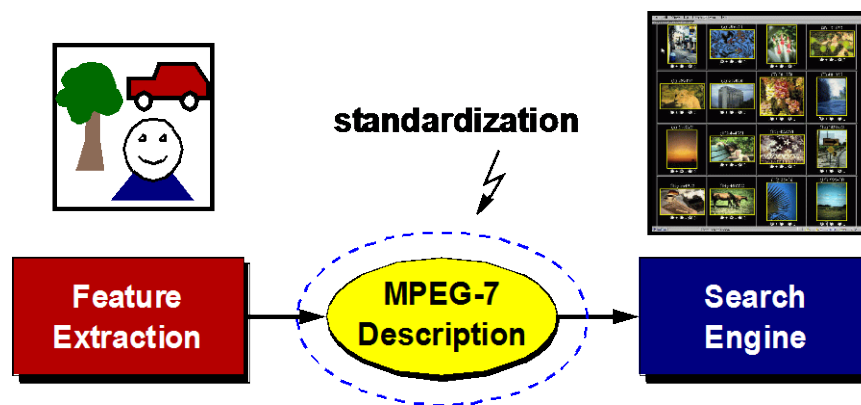
| Feature Extraction:	MPEG-7 Scope:	Search Engine:
Content analysis (D, DS)	Description Schemes (DSs)	Searching & filtering
Feature extraction (D, DS)	Descriptors (Ds)	Classification
Annotation tools (DS)	Language (DDL)	Manipulation
Authoring (DS)	Ref: MPEG-7	Summarization Indexing

Figure 1: The Scope of MPEG-7

2.1 Organization of MPEG-7 Description Tools

Over 100 MPEG-7 Description Tools are currently being developed and refined. The relationships between the MPEG-7 Description Tools are outlined in Figure 2.

The basic elements, at the lower level, deal with basic data types, mathematical structures, schema tools, linking and media localization tools, as well as basic DSs, which are elementary components of more complex DSs. The Schema tools section specifies elements for creating valid MPEG-7 schema instance documents and description fragments.

In addition, this section specifies tools for managing and organizing the elements and datatypes of the schema. Based on this lower level, content description and management elements can be defined. These elements describe

the content from several viewpoints. Currently five viewpoints are defined: creation and production, media, usage, structural aspects, and conceptual aspects. The first three elements primarily address information that's related to the management of the content (content management), whereas the last two are mainly devoted to the description of perceivable information (content description).
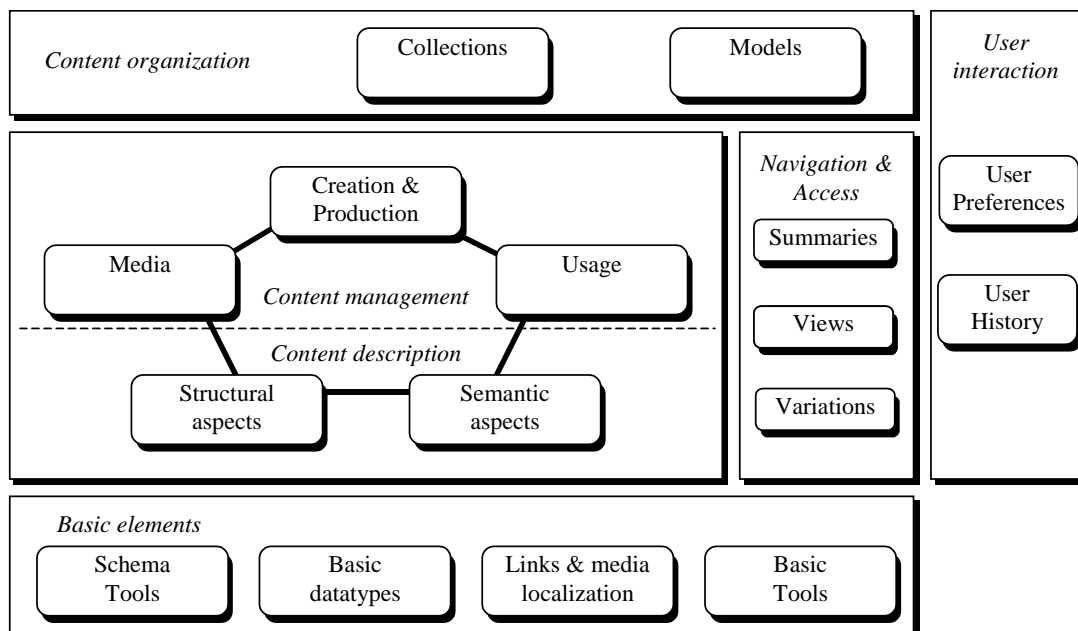
Figure 2: Overview of MPEG-7 Multimedia Description Schemes (DSs)

• *Creation and Production:* Contains meta information that describes the creation and production of the content; typical features include title, creator, classification, and purpose of the creation. Most of the time this information is author-generated since it can't be extracted from the content.

• *Usage:* Contains meta information that's related to the usage of the content; typical features involve rights holders, access rights, publication, and financial information. This information may be subject to change during the lifetime of the AV content.

• *Media:* Contains the description of the storage media; typical features include the storage format, the encoding of the AV content, and elements for the identification of the media. *Note:* Several instances of storage media for the same AV content can be described.

• *Structural aspects:* Contains the description of the AV content from the viewpoint of its structure. The description is structured around segments that represent physical, spatial, temporal, or spatio-temporal components of the AV content. Each segment may be described by signal-based features (color, texture, shape, motion, audio) and some elementary semantic information.

• *Conceptual Aspects:* Contains a description of the AV content from the viewpoint of its conceptual notions.

The five sets of Description Tools are presented here as separate entities, however, they are interrelated and may be partially included in each other. For example, Media, Usage or Creation & Production elements can be attached to individual segments involved in the structural description of the content. Tools are also defined for *navigation and access* and there is another set of tools for *Content organization* which addresses the organization of content by

classification, by the definition of collections and by modeling. Finally, the last set of tools is *User Interaction* which describes user's preferences for the consumption of multimedia content and usage history.

2.2 MPEG-7 Working Groups

Currently MPEG-7 concentrates on the specification of description tools (Descriptors and Description Schemes), together with the development of the MPEG-7 reference software, known as XM (eXperimentation Model). The XML Schema Language was chosen as the base for the Description Definition Language (DDL).

The **MPEG-7 Audio group** develops a range of Description Tools, from generic audio descriptors (e.g., waveform and spectrum envelopes, fundamental frequency) to more sophisticated description tools like Spoken Content and Timbre. Generic Audio Description tools will allow the search for similar voices, by searching similar envelopes and fundamental frequencies of a voice sample against a database of voices. The Spoken Content Description Scheme (DS) is designed to represent the output of a great number of state of the art Automatic Speech Recognition systems, containing both words and phonemes representations and most likely transitions. This alleviates the problem of out-of-vocabulary words, allowing retrieval even when the original word was wrongly decoded. The Timbre descriptors (Ds) describe the perceptual features of instrument sound, that make two sounds having the same pitch and loudness appear different to the human ear. These descriptors allow searching for melodies independently of the instruments.

The **MPEG-7 Visual group** is developing four groups of description tools: Color, Texture, Shape and Motion. Color and Texture Description Tools will allow the search and filtering of visual content (images, graphics, video) by dominant color or textures in some (arbitrarily shaped) regions or the whole image. Shape Description Tools will facilitate "query by sketch" or by contour similarity in image databases, or, for example, searching trademarks in registration databases. Motion Description Tools will allow searching of videos with similar motion patterns that can be applicable to news (e.g. similar movements in a soccer or football game) or to surveillance applications (e.g., detect intrusion as a movement towards the safe zone).

The **MPEG-7 Multimedia Description Schemes group** is developing the description tools dealing with generic and audiovisual and archival features. Its central tools deal with content management and content description as outlined in section 2.1.

The **MPEG-7 Implementation Studies group** is designing and implementing the MPEG-7 Reference Software known as XM.

The **MPEG-7 Systems group** is developing the DDL and the binary format (known as BiM), besides working in the definition of the terminal architecture and access units.

3. MPEG-7 Application Domains

The elements that MPEG-7 standardizes will support a broad a range of applications (for example, multimedia digital libraries, broadcast media selection, multimedia editing, home entertainment devices, etc.). MPEG-7 will also make the web as searchable for multimedia content as it is searchable for text today. This would apply especially to large content archives, which are being made accessible to the public, as well as to multimedia catalogues enabling people to identify content for purchase. The information used for content retrieval may also be used by *agents*, for the selection and filtering of broadcasted "push" material or for personalized advertising. Additionally, MPEG-7 descriptions will allow fast and cost-effective usage of the underlying data, by enabling semi-automatic multimedia presentation and editing. All domains making use of multimedia will benefit from

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

fastcase®
Smarter legal research.