MPEG-7
Daring to Describe Multimedia Content

[WRITTEN BY NEIL DAY]

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 similar-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-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 that addresses how humans expect to interact with computer systems, since it develops rich descriptions that reflect those expectations.

What Are the MPEG Standards?

The Moving Picture Coding Experts Group (MPEG) is a working group of the Geneva-based ISO/IEC (International Standards Organization/International Electrotechnical Committee, www.itscj.ipsj.or.jp/isc29/) 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 find the content you need.

The Key Role of MPEG-7

MPEG-7, formally named "Multimedia Content Description Interface," 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.

Qualifying MPEG-7

MPEG-7 provides the world's most comprehensive set of audiovisual descriptions. These descriptions are based on the catalog (title, creator, rights), semantic (the who, what, when, where information), and structural (the color histogram – measurement of the amount of color – associated with an image or the timbre of a recorded instrument) features of the AV content, and are leveraged on the AV data representation defined by MPEG-1, -2, and -4. MPEG-7 also uses XML Schema as the language of choice for content description.

However, MPEG-7 won't 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's left up to the creativity and innovation of search engine companies to manipulate and massage the MPEG-7–described content into search indexes that can be used by their browser and retrieval tools.

Typical query examples enabled by MPEG-7 include:
- **Audio:** I want to search for songs by humming or whistling a tune or, using an excerpt of Pavarotti's voice, get a list of Pavarotti's records and video clips in which Pavarotti sings or simply makes an appearance.
- **Graphics**: Sketch a few lines on a screen and get a set of images containing similar graphics, logos, and ideograms.
- **Image:** Define objects, including color patches or textures, and get examples from which you select items to compose your image. Or check if your company logo was advertised on a TV channel as contracted.
- **Video:** Allow mobile phone access to video clips of goals scored in a soccer game, or automatically search and retrieve any unusual movements from surveillance videos.

MPEG-7 Technical Activities

It's important to note that MPEG-7 addresses many different applications in many different environments. This means that it needs to provide a flexible and extensible framework for describing audiovisual data. MPEG-7 defines a library of methods and tools for many types of multimedia applications. It standardizes:
- **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 (and possibly descriptors), 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 one that's been encoded to fulfill relevant requirements such as compression efficiency, error resilience, and random access.

Organization of MPEG-7 Descriptions

Over 100 MPEG-7 descriptions are currently being developed and refined. Here I'll explain the overall relationship between the description schemes and outline an example of one such description, the MovieRegion Description Scheme. The relationships between descriptions in MPEG-7 are organized as outlined in Figure 1.

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.
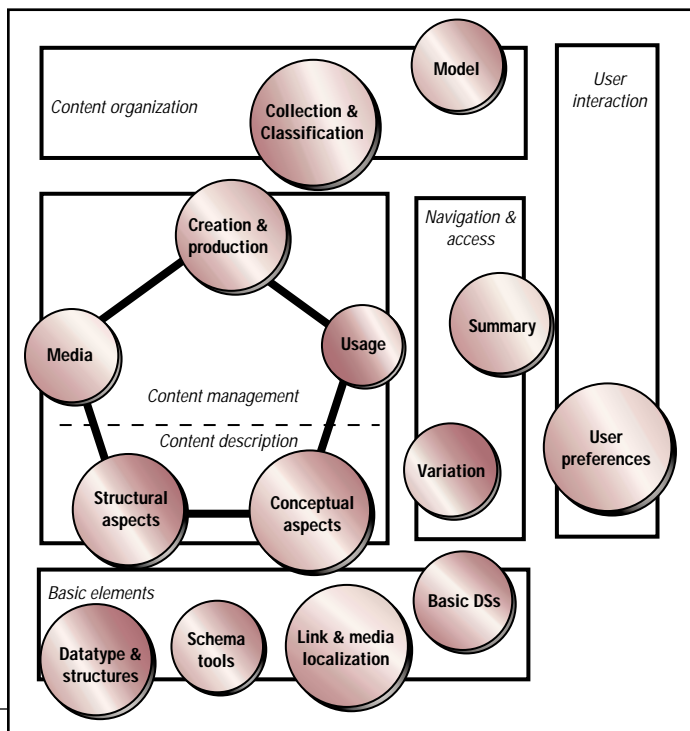

FIGURE 1   Overview of MPEG-7 multimedia description schemes

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).
- **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 spatiotemporal components of the AV content. Each segment may be described by signal-based features (color, texture, shape, motion, audio) and some elementary semantic information.
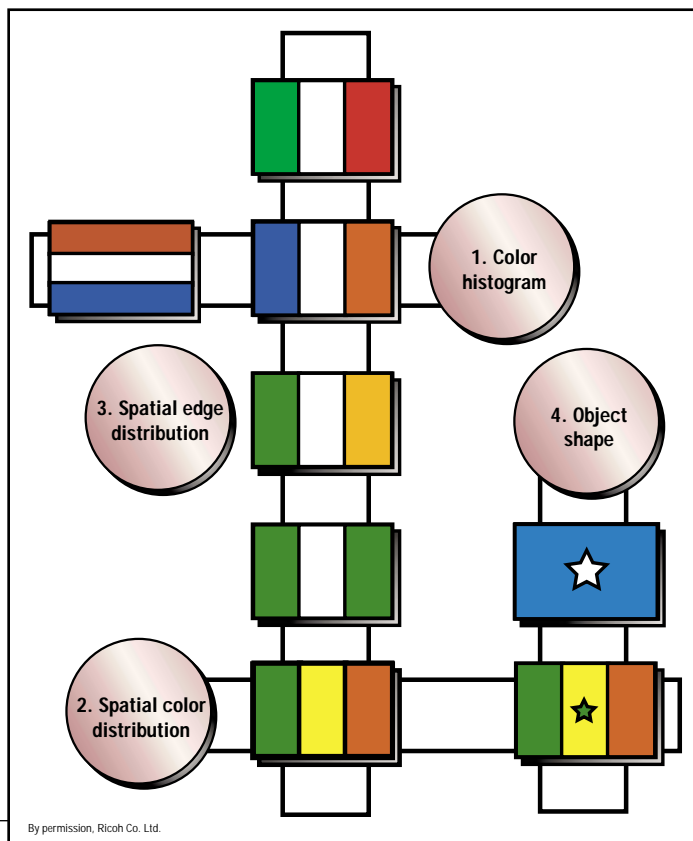
FIGURE 2   Search using image features

By permission, Ricoh Co. Ltd.

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

The five sets of elements are presented as separate entities; however, they're interrelated and may be partially included in each other.

Searching for Content Using Image Features

Figure 2 shows several ways to search for visual content using the inherent structural features of an image. Four image features are detailed. The color histogram feature (1) of an image allows me to search for images that have the same color. *Note:* The position of the colors isn't important, but the amount of similar color in the image is. The next feature, spatial color distribution (2), lets me search for images when the location of the same color is important. You can see that the added object in the right-bottom flag doesn't affect this type of search. In addi-

FIGURE 3   Retrieval example by spatial edge distribution

tion, you can search for images that have a similar edge or contour profile as the spatial edge distribution (3) search technique. *Note:* Color doesn't make a difference in this type of search. In searching by object shape (4), the color and edge profiles are not important.

Figure 3 shows a typical image search application. Here, you select a query image to find all images that look like the query image (see the search box in the figure). Then choose the image feature you wish to use in searching for similar images, for example, the edge feature. (*Note:* The search could also be done using a combination of image features.) After you make your selections, press the search button and the results are displayed in the order of the most similar images. In Figure 3 you can see that images with similar edge profiles have been retrieved irrespective of color content and distribution.

Structural MPEG-7 Description Tools

In MPEG-7 the main tools for describing the structure of AV content are segment entities, segment features, and segment relations. In Figure 4 you can see an example of a structural description of a video sequence that's described using such tools. The description is made up of three video segments, two moving regions, one segment relation, and several two-segment decompositions. The root video segment corresponds to the entire video sequence. The remaining two video segments are formed from a segment decomposition of the root segment. The last two video segments are then related by the segment relation "Before". With a further decomposition of one of these segments, you obtain two moving regions that correspond to the two objects: "Interviewer" and "Coach". What does MPEG-7 DDL syntax look like? Let's look at segment entities and the MPEG-7 syntax for the MovingRegion Description Scheme in Listing 1 to get an idea.
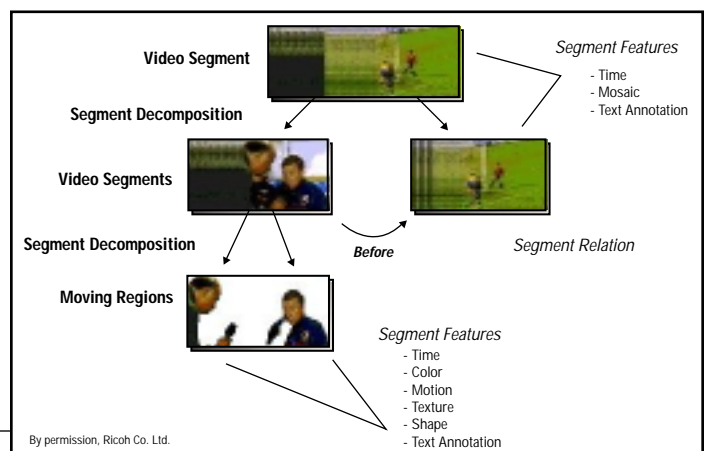

By permission, Ricoh Co. Ltd.

FIGURE 4   Video segment decomposition into its constituent lower-level descriptions

Segment Entities

Segment entities are tools that represent physical spatial, temporal, or spatiotemporal components of AV content. The decomposition of segments into other segments is described by the SegmentDecomposition DS.

There are six specialized types of segment DSs: VideoSegment, MovingRegion, StillRegion, VideoText, AudioSegment, and AudioVisualSegment. A temporal segment may be a set of samples in an audio sequence represented by an AudioSegment DS, or a set of frames in a video sequence represented by a VideoSegment DS. When you have a spatiotemporal segment that includes both audio samples and frames of AV content, you use an AudioVisualSegment DS. *Note:* A spatial segment may be a region in an image, or a frame in a video sequence represented by a StillRegion DS. Finally, you can have a spatiotemporal segment that may correspond to a moving region in a video sequence represented by a MovingRegion DS. The VideoText DS, a specialized MovingRegion DS, represents a text region in a still image or in a set of video frames. The following is a list outlining the MPEG-7 syntax for the MovingRegion DS.

Outstanding Issues

IPMP (INTELLECTUAL PROPERTY MANAGEMENT AND PROTECTION)

MPEG-7 and representatives of Copyright Industries are working together to protect the creations of artists. This is an ongoing concern in which the availability of rich multimedia descriptive data can lead to the abuse of content. for example, a hacker could change the ownership of a content description.

Hopefully by September 2001, when MPEG-7 officially becomes an International Standard, most of the concerns of the copyright organizations will have been addressed.

An Invitation to Participate

All those organizations that are interested in the search and retrieval of AV content or that wish to describe AV content databases in a standard way are welcome to participate in the development of the MPEG-7 standard. Please get in touch with me at neil@garage.co.jp.

For more information about MPEG-7 and these issues, see www.cselt.it/mpeg/ and MPEG-7 Industry Focus Group Web site (under development) at www.mpeg-7.com. ✖

■ AUTHOR BIO

Neil Day recently moved to a pioneering Web solutions company in Tokyo called Digital Garage (www.garage.co.jp), where he heads the R&D department. He's currently exploring the uses of MPEG-7 and XML-related technologies in Internet entertainment applications for the Japanese and international markets. Neil holds an engineering degree from Trinity College, Dublin.

NEIL@GARAGE.CO.JP

LISTING 1

```
MPEG-7 MovingRegion DS Syntax
<!-- ############################################# -->
<!-- Definition of "MovingRegion DS"              -->
<!-- ############################################# -->

<element name="MovingRegion" type="mpeg7:MovingRegionType"
     equivClass="Segment"/>
<complexType name="MovingRegionType" base="mpeg7:Segment-
Type"
       derivedBy="extension">
<element ref="mpeg7:MediaTime" minOccurs="1" maxOccurs="1"/>
<element ref="mpeg7:MediaTimeMask" minOccurs="1" maxOc-
curs="1"/>
<element ref="mpeg7:ColorSpace" minOccurs="0" maxOc-
curs="1"/>
<element ref="mpeg7:ColorQuantization" minOccurs="0" maxOc-
curs="1"/>
<element ref="mpeg7:GofGopColorHistogram" minOccurs="0" max-
Occurs="1"/>
<element ref="mpeg7:ColorLayout" minOccurs="0" maxOc-
curs="1"/>
<element ref="mpeg7:MotionTrajectory" minOccurs="0" maxOc-
curs="1"/>
<element ref="mpeg7:ParametricMotion" minOccurs="0" maxOc-
curs="1
<element ref="mpeg7:SpatioTemporalLocator" minOccurs="0"
maxOccurs="1"/>
<attribute name="spatialConnectivity" type="boolean"
use="required"/>
<!-- Restriction of refEltName to MovingRegion DS -->
<attribute name="idref" type="IDREF" refEltName="MovingRe-
gion"
use="optional"/>
</complexType>

"/>
```

The code listing for this article can also be located at www.xml-journal.com

DOWNLOAD THE CODE!

Insight
www.xml
boutique.com