MPEG-4: An Audio-Visual Representation Standard.

MPEG-4 is an ISO/IEC (International Standardization Organization / International Electro-technical Commission) standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2. These standards made interactive video on CD-ROM and Digital Television possible. MPEG-4 is the result of another international effort involving hundreds of researchers and engineers from all over the world.

It will provide the standardized technological elements enabling the integration of the production, distribution and content access paradigms of the three fields:

- Digital television
- Interactive graphics applications (synthetic content)
- Interactive multimedia (World Wide Web, distribution of and access to content).

The Main Features:

Coded Representation And Composition Of Objects:

MPEG-4 audiovisual scenes are composed of several media objects, organized in a hierarchical fashion such as:

- still images (e.g. as a fixed background)
- video objects (e.g. a talking person - without the background)
- audio objects (e.g. the voice associated with that person)

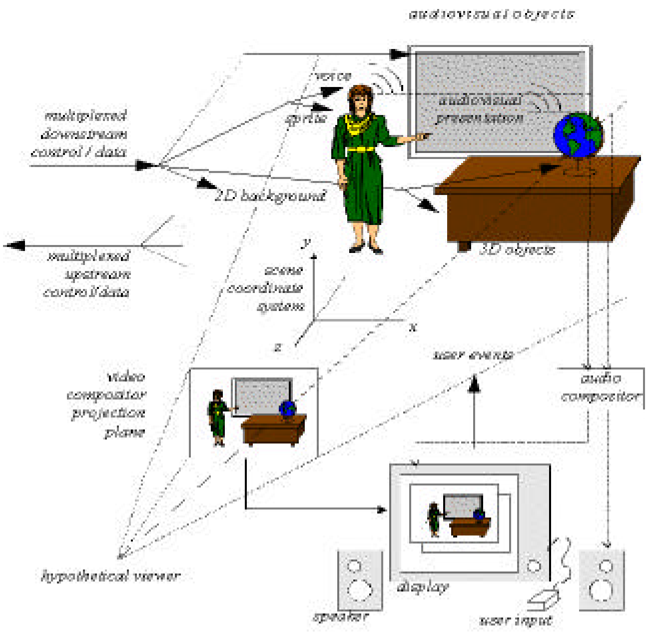

Figure 1-An example of an MPEG-4 Scene

MPEG-4 standardizes 2D or 3D media objects. In addition, MPEG-4 defines the coded representation of objects as:

- text and graphics
- talking synthetic heads and associated text used to synthesize the speech and animate the head
- synthetic sound

A media object in its coded form consists of descriptive elements that allow handling the object, in an audiovisual scene as well as associated streaming data, if needed. Each media object can be represented independent of its surroundings or background. The coded representation of media objects is efficient while taking into account the desired functionalities such as error robustness, easy extraction and editing of an object, or having an object available in a scaleable form. The figure contains compound media objects that group primitive media objects together. With reference to the above figure, the visual object corresponding to the talking person and the corresponding voice are tied together to form a new compound media object, containing both the aural and visual components of that talking person. The various options that a user has are enlisted below:

- place media objects anywhere in a given coordinate system
- apply transforms to change the appearance of a media object
- group primitive media objects in order to form compound media objects
- apply streamed data to media objects, in order to modify their attributes
- change the user's viewing and listening points anywhere in the scene

Synchronization and delivery of streaming data & interaction with media objects: :

Synchronization of elementary streams is achieved through time stamping of individual access units within elementary streams. The synchronization layer manages the identification of such access units and the time stamping. Independent of the media type, this layer allows identification of the type of access unit in elementary streams, recovery of the media object's or scene description's time base, and it enables synchronization among them.

In general, the user observes a scene that is composed following the design of the scene's author. Depending on the degree of freedom allowed by the author, however, the user has the possibility to interact with the scene. Operations a user may be allowed to perform include :

- change the viewing/listening point of the scene
- drag objects in the scene to a different position;
- trigger a cascade of events by clicking on a specific object
- select the desired language when multiple language tracks are available

More complex kinds of behavior can also be triggered, e.g. a virtual phone rings, the user answers and a communication link is established.

Identification of intellectual property:

It is important to have the possibility to identify intellectual property in MPEG-4 media objects. Therefore, MPEG has worked with representatives of different creative industries in the definition of syntax and tools to support this. MPEG-4 incorporates identification of the intellectual property by storing unique identifiers that are issued by international numbering systems. These numbers can be applied to identify a current rights holder of a media object. Since not all content is identified by such a number, MPEG-4 Version 1 offers the possibility to identify intellectual property by a key-value pair (e.g.:»composer«/»Mr. Gates«).

Streams coming from the network (or a storage device), as TransMux Streams, are de-multiplexed into FlexMux Streams and passed to appropriate FlexMux de-multiplexers that retrieve Elementary Streams. The Elementary Streams (ESs) are parsed and passed to the appropriate decoders. Decoding recovers the data in an AV object from its encoded form and performs the necessary operations to reconstruct the original AV object ready for rendering on the appropriate device. Audio and visual objects are represented in their coded form. The reconstructed AV object is made available to the composition layer for potential use during scene rendering. Decoded AVOs, along with scene description information, are used to compose the scene as described by the author. The user can, to the extent allowed by the author, interact with the scene, which is eventually rendered and presented.

Coding of audio and video objects:

MPEG-4 coding of audio objects provides tools for both representing natural sounds and for synthesizing sounds based on structured descriptions. The representation for synthesized sound can be derived from text data or so-called instrument descriptions and by coding parameters to provide effects, such as reverberation and spatialization. The representations provide compression and other functionalities, such as scalability and effects processing. With the intent of identifying a suitable digital audio broadcast format to provide improvements over the existing AM modulation services, several coder configurations involving the MPEG-4 CELP, TwinVQ, and AAC tools have been compared to a reference AM system. It was found that higher quality can be achieved in the same bandwidth with digital techniques and that scalable coder configurations offered performance superior to a simulcast alternative. Additional verification tests were carried out by MPEG, in which the tools for speech and general audio coding were compared to existing standards.

MPEG-4 Video offers technology that covers a large range of existing applications as well as new ones. The low-bit rate and error resilient coding allows for robust communication over limited rate wireless channels, useful for e.g. mobile videophones and space communication. There may also be roles in surveillance data compression since it is possible to have a very low or variable frame rate. At high bit-rates, tools are available to allow the transmission and storage of high-quality video suitable for the studio and other very demanding content creation applications. It is likely that the standard will eventually support data-rates well beyond those of MPEG-2. A major application area is interactive web-based video. Software that provides live MPEG-4 video on a web page has already been demonstrated. There is much room for applications to make use of MPEG-4's object-based characteristics. The binary and grayscale shape-coding tools allow arbitrary-shaped video objects to be composed together with text and graphics. Doing so, a rich interactive experience for web-based presentations and advertising can be provided; this same scenario also applies to set-top box applications. Additionally, it is possible to make use of scalability tools to allow for a smooth control of user experience with terminal and data link capabilities. MPEG-4 video has already been used to encode video captures with a hand-held camera. This form of application is likely to grow in popularity with its fast and easy transfer to a web page, and may also make use of MPEG-4 still-texture mode for still frame capture. The games market is another area where the application of MPEG-4 video, still-texture, interactivity and SNHC shows much promise, with 3-D texture mapping of still images, live video, or extended pre-recorded video sequences enhancing the player experience. Adding live video of users adds to the user experience multi-player 3-D games, as does use of arbitrary-shaped video, where transparency could be combined artistically with 3-D video texture mapping.

Coming to a close the introduction of this new standard shall have a salutary effect in the world of technology. For authors, MPEG-4 will enable the production of content that has far greater reusability and flexibility than is possible today with individual technologies like digital television, animated graphics, Web pages and their extensions. Also, it will be possible to better manage and protect content owner rights. For network service providers MPEG-4 will offer transparent information that will be interpreted and translated into the appropriate native signaling messages of each network. The transmitted information will be directly associated with Quality of Service requests. For end users, MPEG-4 will enable a variety of new functionalities and higher levels of interaction with content. MPEG will thus prevent the emergence of a multitude of proprietary, non-interworking formats and players to the benefit of all parties involved.

Habibullah Pagarkar & Chinmay Shah [S.E. IT]